Empirical Likelihood for Spatial Autoregressive Models with Spatial Autoregressive Disturbances

Yongsong Qin

Sankhya A
The Indian Journal of Statistics - Official Journal of Indian Statistical Institute

ISSN 0976-836X
Volume 83
Number 1

Sankhya A (2021) 83:1-25
DOI 10.1007/s13171-019-00166-3

Springer

Sankhyā : The Indian Journal of Statistics
2021, Volume 83-A, Part 1, pp. 1-25

Empirical Likelihood for Spatial Autoregressive Models with Spatial Autoregressive Disturbances

Yongsong Qin
Guangxi Normal University, Guilin, China

Abstract

The empirical likelihood ratio statistics are constructed for the parameters in spatial autoregressive models with spatial autoregressive disturbances. It is shown that the limiting distributions of the empirical likelihood ratio statistics are chi-squared distributions, which are used to construct confidence regions for the parameters in the models.

AMS (2010) subject classification. Primary 62G05; Secondary 62E20.
Keywords and phrases. Spatial ARAR model, Empirical likelihood, Confidence region

1 Introduction

Spatial data are common in ecology, environmental health and epidemiology, where sampling units are geographical areas or spatially located individuals (Cressien, 1993). Analysis of spatial data is challenged by the spatial correlation among the observations. In this article, the following spatial autoregressive model with spatial autoregressive disturbances (spatial ARAR model) is investigated:

$$Y_n = \rho_1 W_n Y_n + X_n\beta + u_{(n)}, u_{(n)} = \rho_2 M_n u_{(n)} + \epsilon_{(n)}, \tag{1.1}$$

where n is the number of spatial units, $\rho_j, j = 1, 2$, are the scalar autoregressive parameters with $|\rho_j| < 1, j = 1, 2$, β is the $k \times 1$ vector of regression parameters, $X_n = (x_1, x_2, \cdots, x_n)^\tau$ is the non-random $n \times k$ matrix of observations on the independent variable, $Y_n = (y_1, y_2, \cdots, y_n)^\tau$ is an $n \times 1$ vector of observations on the dependent variable, W_n and M_n are $n \times n$ spatial weighting matrices of constants, $\epsilon_{(n)}$ is an $n \times 1$ vector of model errors which satisfies

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

The $\rho_1 W_n$ term is a spacial lag in the dependent variable and its coefficient represents the spatial influence due to neighbors realized dependent variable. The $\rho_2 M_n$ term is a spacial lag in the disturbances and its coefficient represents the spacial effect of unobservables on neighboring units. This model is introduced by Cliff and Ord (1973). Excellent surveys and developments in testing and estimation of this model can be found in Cliff and Ord (1973), Anselin (1988), Cressien (1993), Anselin and Bera (1998), Kelejian and Prucha (2001), & Liu et al. (2010), among others.

There are two major estimation approaches for the corresponding parameters of the spatial ARAR model. One is the maximum likelihood (ML) method (e.g. Anselin, 1988). The other is the generalized method of moments (GMM) by Liu et al. (2010). The asymptotic properties of the maximum likelihood estimator (MLE) and the GMM estimator for the spatial ARAR model are investigated by Anselin (1988) and Liu et al. (2010), respectively. However, it may not be easy to use these normal approximation results to construct confidence region for the parameters in the spatial ARAR model as the asymptotic covariance in the asymptotic distribution is unknown. In this article, we propose to use the empirical likelihood (EL) method introduced by Owen (1988, 1990) to construct confidence region for the parameters in the spatial ARAR model. The shape and orientation of the EL confidence region are determined by data and the confidence region is obtained without covariance estimation. These features of the EL confidence region are the major motivations for our current proposal. A comprehensive review on EL for regressions can be found in Chen and Keilegom (2009). More references on EL methods can be found in Owen (2001), Qin and Lawless (1994), Chen and Qin (1993), Zhong and Rao (2000) and Wu (2004), among others.

The main challenge in using the EL method to the spatial ARAR model is that the estimating equation for the spatial ARAR model is a linear-quadratic form of $\epsilon_{(n)}$ (e.g. (2.1)–(2.4)). The idea to use the EL method for the spatial ARAR model is to introduce a martingale sequence to transform the linear-quadratic form into a linear form. It is interesting to note that the estimation equations for other spatial models may have the linear-quadratic forms. Therefore this approach of transformation also opens a way to use EL methods to more general spatial models. The applications of EL method to other spatial models are left for our future study. After the completion of this article, we are informed that Jin and Lee (2019) independently investigate the generalized empirical likelihood (GEL) estimation and tests of the spatial ARAR models by exploring an inherent martingale structure. It is noted that the model and method used in Jin and Lee (2019) are more general than those in this article. However, the statistical inference in Jin and Lee (2019) is based on the EL point estimator. This article focuses on the construction of the EL confidence intervals. There is no need to construct the point estimation of parameters by using EL method. Therefore, the regularity conditions in Jin and Lee (2019) are stronger than those in this article. One can also see, based on the results in Jin and Lee (2019), that the EL estimator for the model in this article is efficient in the sense that the asymptotic variance of the EL estimator is the same as the maximum likelihood estimator when the $\epsilon_{(n)}$ is normally distributed.

The article is organized as follows. Section 2 gives the main results. Results from a simulation study are reported in Section 3. All the technical details are presented in Section 4.

2 Main Results

We continue with model (1.1). Let $A_n(\rho_1) = I_n - \rho_1 W_n, B_n(\rho_2) = I_n - \rho_2 M_n$ and suppose that $A_n(\rho_1)$ and $B_n(\rho_2)$ are nonsingular. Then

$$Y_n = A_n^{-1}(\rho_1)X_n\beta + A_n^{-1}(\rho_1)B_n^{-1}(\rho_2)\epsilon_{(n)}.$$

At this moment, suppose that $\epsilon_{(n)}$ is normally distributed, which is used to derive the EL statistic only and not employed in our main results. Then the log-likelihood function based on the response vector Y_n is

$$L = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 + \log|A_n(\rho_1)| + \log|B_n(\rho_2)| - \frac{1}{2\sigma^2}\epsilon_{(n)}^\tau\epsilon_{(n)},$$

where $\epsilon_{(n)} = B_n(\rho_2)\{A_n(\rho_1)Y_n - X_n\beta\}$. Let $G_n = B_n(\rho_2)W_nA_n^{-1}(\rho_1)B_n^{-1}(\rho_2)$, $H_n = M_nB_n^{-1}(\rho_2)$, $\tilde{G}_n = \frac{1}{2}(G_n + G_n^\tau)$ and $\tilde{H}_n = \frac{1}{2}(H_n + H_n^\tau)$. It can be shown that (e.g. Anselin, 1988, pp. 74-75)

$$\partial L/\partial\beta = \frac{1}{\sigma^2}X_n^\tau B_n^\tau(\rho_2)\epsilon_{(n)},$$

$$\begin{aligned}\partial L/\partial\rho_1 &= \frac{1}{\sigma^2}\{B_n(\rho_2)W_nA_n^{-1}(\rho_1)X_n\beta\}^\tau\epsilon_{(n)} \\ &\quad + \frac{1}{\sigma^2}\{\epsilon_{(n)}^\tau B_n(\rho_2)W_nA_n^{-1}(\rho_1)B_n^{-1}(\rho_2)\epsilon_{(n)} - \sigma^2 tr(A_n^{-1}(\rho_1)W_n)\} \\ &= \frac{1}{\sigma^2}\{B_n(\rho_2)W_nA_n^{-1}(\rho_1)X_n\beta\}^\tau\epsilon_{(n)} + \frac{1}{\sigma^2}\{\epsilon_{(n)}^\tau\tilde{G}_n\epsilon_{(n)} - \sigma^2 tr(\tilde{G}_n)\},\end{aligned}$$

$$\partial L/\partial\rho_2 = \frac{1}{\sigma^2}\{\epsilon_{(n)}^\tau\tilde{H}_n\epsilon_{(n)} - \sigma^2 tr(\tilde{H}_n)\},$$

$$\partial L/\partial\sigma^2 = \frac{1}{2\sigma^4}(\epsilon_{(n)}^\tau\epsilon_{(n)} - n\sigma^2).$$

Letting above derivatives be 0, we obtain the following estimating equations:

$$X_n^\tau B_n^\tau(\rho_2)\epsilon_{(n)} = 0, \tag{2.1}$$

$$\{B_n(\rho_2)W_nA_n^{-1}(\rho_1)X_n\beta\}^\tau\epsilon_{(n)} + \{\epsilon_{(n)}^\tau\tilde{G}_n\epsilon_{(n)} - \sigma^2 tr(\tilde{G}_n)\} = 0, \tag{2.2}$$

$$\epsilon_{(n)}^\tau\tilde{H}_n\epsilon_{(n)} - \sigma^2 tr(\tilde{H}_n) = 0, \tag{2.3}$$

$$\epsilon_{(n)}^\tau\epsilon_{(n)} - n\sigma^2 = 0. \tag{2.4}$$

We use $\tilde{g}_{ij}$, $\tilde{h}_{ij}$, b_i and s_i to denote the (i,j) element of the matrix $\tilde{G}_n$, the (i,j) element of the matrix $\tilde{H}_n$, the i-th column of the matrix $X_n^\tau B_n^\tau(\rho_2)$ and the i-th component of the vector $B_n(\rho_2)W_nA_n^{-1}(\rho_1)X_n\beta$, respectively, and adapt the convention that any sum with an upper index of less than one is zero. To deal with the quadratic forms in (2.2) and (2.3), we follow Kelejian and Prucha (2001) to introduce a martingale difference array. Define the σ-fields: $\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_i = \sigma(\epsilon_1, \epsilon_2, \cdots, \epsilon_i), 1 \le i \le n$. Let

$$\tilde{Y}_{in} = \tilde{g}_{ii}(\epsilon_i^2 - \sigma^2) + 2\epsilon_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\epsilon_j, \tilde{Z}_{in} = \tilde{h}_{ii}(\epsilon_i^2 - \sigma^2) + 2\epsilon_i \sum_{j=1}^{i-1} \tilde{h}_{ij}\epsilon_j. \quad (2.5)$$

Then $\mathcal{F}_{i-1} \subseteq \mathcal{F}_i, \tilde{Y}_{in}$ is $\mathcal{F}_i-$ measurable and $E(\tilde{Y}_{in}|\mathcal{F}_{i-1}) = 0$. Thus $\{\tilde{Y}_{in}, \mathcal{F}_i, 1 \le i \le n\}$ and $\{\tilde{Z}_{in}, \mathcal{F}_i, 1 \le i \le n\}$ form two martingale difference arrays and

$$\epsilon_{(n)}^\tau \tilde{G}_n \epsilon_{(n)} - \sigma^2 tr(\tilde{G}_n) = \sum_{i=1}^n \tilde{Y}_{in}, \quad \epsilon_{(n)}^\tau \tilde{H}_n \epsilon_{(n)} - \sigma^2 tr(\tilde{H}_n) = \sum_{i=1}^n \tilde{Z}_{in}. \quad (2.6)$$

Based on (2.1) to (2.6), we propose the following EL ratio statistic for $\theta \hat{=} (\beta^\tau, \rho_1, \rho_2, \sigma^2)^\tau \in R^{k+3}$:

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

where $\{p_i\}$ satisfy

$$\begin{aligned}
& p_i \ge 0, 1 \le i \le n, \sum_{i=1}^n p_i = 1, \\
& \sum_{i=1}^n p_i b_i \epsilon_i = 0, \\
& \sum_{i=1}^n p_i \left\{ \tilde{g}_{ii}(\epsilon_i^2 - \sigma^2) + 2\epsilon_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\epsilon_j + s_i\epsilon_i \right\} = 0, \\
& \sum_{i=1}^n p_i \left\{ \tilde{h}_{ii}(\epsilon_i^2 - \sigma^2) + 2\epsilon_i \sum_{j=1}^{i-1} \tilde{h}_{ij}\epsilon_j \right\} = 0, \\
& \sum_{i=1}^n p_i(\epsilon_i^2 - \sigma^2) = 0.
\end{aligned}$$

Let

$$\omega_i(\theta) = \begin{pmatrix} b_i \epsilon_i \\ \tilde{g}_{ii}(\epsilon_i^2 - \sigma^2) + 2\epsilon_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\epsilon_j + s_i \epsilon_i \\ \tilde{h}_{ii}(\epsilon_i^2 - \sigma^2) + 2\epsilon_i \sum_{j=1}^{i-1} \tilde{h}_{ij}\epsilon_j \\ \epsilon_i^2 - \sigma^2 \end{pmatrix}_{(k+3)\times 1},$$

where ϵ_i is the i-th component of $\epsilon_{(n)} = B_n(\rho_2)\{A_n(\rho_1)Y_n - X_n\beta\}$. Following Owen (1990), one can show that

$$p_i = \frac{1}{n} \cdot \frac{1}{1 + \lambda^\tau(\theta)\omega_i(\theta)}, 1 \le i \le n,$$

and

$$\ell_n(\theta) \hat{=} -2\log L_n(\theta) = 2\sum_{i=1}^{n} \log\{1 + \lambda^\tau(\theta)\omega_i(\theta)\}, \tag{2.7}$$

where $\lambda(\theta) \in R^{k+3}$ is the solution of the following equation:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\omega_i(\theta)}{1 + \lambda^\tau(\theta)\omega_i(\theta)} = 0. \tag{2.8}$$

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 use $||a||$ to denote the L_2-norm of a vector a. Furthermore, Let $\mathbf{1}_n$ present the n-dimensional (column) vector with 1 as its components. To obtain the asymptotical distribution of $\ell_n(\theta)$, we need following assumptions.

A1. $\{\epsilon_i, 1 \le i \le n\}$ are independent and identically distributed random variables with mean 0, variance $\sigma^2 > 0$ and $E|\epsilon_1|^{4+\eta_1} < \infty$ for some $\eta_1 > 0$.

A2. Let W_n, M_n, $A_n^{-1}(\rho_1), B_n^{-1}(\rho_2)$ and $\{x_i\}$ be as described above. They satisfy the following conditions:

(i) The row and column sums of W_n, M_n, $A_n^{-1}(\rho_1)$ and $B_n^{-1}(\rho_2)$ are uniformly bounded in absolute value;

(ii) $\{x_i\}$ are uniformly bounded.

A3. There is a constants $c_j > 0, j = 1, 2$, such that $0 < c_1 \leq \lambda_{min}\left(n^{-1}\Sigma_{k+3}\right) \leq \lambda_{max}\left(n^{-1}\Sigma_{k+3}\right) \leq c_2 < \infty$, where $\lambda_{min}(A)$ and $\lambda_{max}(A)$ denote the minimum and maximum eigenvalues of a matrix A, respectively,

$$\Sigma_{k+3} = \Sigma_{k+3}^{\tau} = Cov\left\{\sum_{i=1}^{n} \omega_i(\theta)\right\} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} & \Sigma_{24} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} & \Sigma_{34} \\ \Sigma_{41} & \Sigma_{42} & \Sigma_{43} & \Sigma_{44} \end{pmatrix}, \quad (2.9)$$

where

$$\Sigma_{11} = \sigma^2\{B_n(\rho_2)X_n\}^{\tau} B_n(\rho_2)X_n,$$

$$\begin{aligned}\Sigma_{12} &= \sigma^2\{B_n(\rho_2)X_n\}^{\tau} B_n(\rho_2)W_n A_n^{-1}(\rho_1)X_n\beta \\ &\quad + \mu_3\{B_n(\rho_2)X_n\}^{\tau} Vec(diag\tilde{G}_n),\end{aligned}$$

$$\Sigma_{13} = \mu_3\{B_n(\rho_2)X_n\}^{\tau} Vec(diag\tilde{H}_n), \Sigma_{14} = \mu_3\{B_n(\rho_2)X_n\}^{\tau}\mathbf{1}_n$$

$$\begin{aligned}\Sigma_{22} &= 2\sigma^4 tr(\tilde{G}_n{}^2) \\ &\quad + \sigma^2\{B_n(\rho_2)W_n A_n^{-1}(\rho_1)X_n\beta\}^{\tau} B_n(\rho_2)W_n A_n^{-1}(\rho_1)X_n\beta \\ &\quad + (\mu_4 - 3\sigma^4)||Vec(diag\tilde{G}_n)||^2 \\ &\quad + 2\mu_3\{B_n(\rho_2)W_n A_n^{-1}(\rho_1)X_n\beta\}^{\tau} Vec(diag\tilde{G}_n),\end{aligned}$$

$$\begin{aligned}\Sigma_{23} &= 2\sigma^4 tr(\tilde{G}_n\tilde{H}_n) + (\mu_4 - 3\sigma^4)Vec^{\tau}(diag(\tilde{G}_n))Vec(diag(\tilde{H}_n)) \\ &\quad + \mu_3\{B_n(\rho_2)W_n A_n^{-1}(\rho_1)X_n\beta\}^{\tau} Vec(diag\tilde{H}_n),\end{aligned}$$

$$\Sigma_{24} = (\mu_4 - \sigma^4)tr(\tilde{G}_n) + \mu_3\{B_n(\rho_2)W_n A_n^{-1}(\rho_1)X_n\beta\}^{\tau}\mathbf{1}_n,$$

$$\Sigma_{33} = 2\sigma^4 tr(\tilde{H}_n{}^2) + (\mu_4 - 3\sigma^4)||Vec(diag(\tilde{H}_n))||^2,$$

$$\Sigma_{34} = (\mu_4 - \sigma^4)tr(\tilde{H}_n), \Sigma_{44} = n(\mu_4 - \sigma^4).$$

Remark 1. (2.9) *is verified in the proof of Lemma 3. Conditions A1 to A3 are common assumptions for spatial models. For example, A1 and A2 are used in Assumptions 1, 4, 5 and 6 in Lee (2004), the analog of* $0 < c_1 \le \lambda_{min}\left(n^{-1}\Sigma_{k+3}\right)$ *(e.g.* $n^{-1}\sigma^2_{\widetilde{Q}} \ge c$ *for some constant* $c > 0$ *in Lemma 1 in this article) is employed in the assumption of Theorem 1 in Kelejian and Prucha (2001). From Conditions A1 and A2, one can see that* $\lambda_{max}\left(n^{-1}\Sigma_{k+3}\right) \le c_2 < \infty$*. For the sake of argument, we list this consequence of A1 and A2 as a condition here.*

We now state the main results.

Theorem 1. *Suppose that Assumptions (A1) to (A3) are satisfied. Then under model* (1.1)*, as* $n \to \infty$,

$$\ell_n(\theta) \xrightarrow{d} \chi^2_{k+3},$$

where χ^2_{k+3} *is a chi-squared distributed random variable with* $k+3$ *degrees of freedom.*

Let $z_\alpha(k+3)$ satisfy $P(\chi^2_{k+3} \le z_\alpha(k+3)) = \alpha$ 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

$$\{\theta : \ell_n(\theta) \le z_\alpha(k+3)\}.$$

In some occasions, one may want to obtain the confidence region for $\psi = (\beta^\tau, \rho_1, \rho_2)^\tau \in R^{k+2}$. To serve this, we can let $\ell_{n2}(\psi) = 2\log\{L_n(\hat{\theta})\} - 2\log\{L_n(\psi, \hat{\sigma}^2)\}$, where $\hat{\theta}$ and $\hat{\sigma}^2$ are the EL estimators of θ and σ^2 (with ψ fixed for the later estimator), respectively. Then following the proof of Corollary 5 in Qin and Lawless (1994), we have the following result.

Theorem 2. *Suppose that Assumptions (A1) to (A3) are satisfied. Then as* $n \to \infty$,

$$\ell_{n2}(\psi) \xrightarrow{d} \chi^2_{k+2},$$

where χ^2_{k+2} *is a chi-squared distributed random variable with* $k+2$ *degrees of freedom.*

Based on this result, the EL based confidence region for ψ with asymptotically correct coverage probability α can be constructed as

$$\{\psi : \ell_{n2}(\psi) \leq z_\alpha(k+2)\}.$$

3 Simulations

According to Anselin (1988), when the error term $\epsilon_{(n)}$ is normal distributed, the likelihood ratio (LR) $LR(\theta_0) = 2(L(\hat{\theta}) - L(\theta_0))$ is asymptotically distributed as χ^2_{k+3} under the null hypothesis: $\theta = \theta_0$, where L is the corresponding log-likelihood and $\hat{\theta}$ is the maximum likelihood estimator. It follows that the LR based confidence region for θ with asymptotically correct coverage probability α can be constructed as

$$\{\theta : LR(\theta) \leq z_\alpha(k+3)\}.$$

We note that the LR method requires to know the form of the distribution of the population in study, while the EL method does not.

We conducted a small simulation study to compare the finite sample performances of the confidence regions based on EL and LR methods with confidence level $\alpha = 0.95$, and report the proportion of $LR(\theta_0) \leq z_{0.95}(k+3)$ and $\ell_n(\theta_0) \leq z_{0.95}(k+3)$ respectively in our 2,000 simulations, where θ_0 is the true value of θ. The results of simulations are reported in Tables 1, 2 and 3.

In the simulations, we used the model: $Y_n = \rho_1 W_n Y_n + X_n\beta + u_{(n)}, u_{(n)} = \rho_2 M_n u_{(n)} + \epsilon_{(n)}$ with $X_n = (x_1, x_2, \cdots, x_n)^\tau, x_i = \frac{i}{n+1}, 1 \leq i \leq n$, $\beta = 3.5$, (ρ_1, ρ_2) were taken as $(-0.85, -0.15)$, $(-0.85, 0.15)$ $(0.85, -0.15)$ and $(0.85, 0.15)$, respectively, and $\epsilon_i's$ were taken from $N(0,1)$, $t(5)$ and $\chi^2_4 - 4$, respectively.

For the contiguity weight matrix $W_n = (W_{ij})$, we took $W_{ij} = 1$ 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). We first considered 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.

Table 1: Coverage probabilities of the LR and EL confidence regions with $\epsilon_i \sim N(0,1)$

(ρ_1,ρ_2)	$W_n = M_n$	LR	EL	(ρ_1,ρ_2)	$W_n = M_n$	LR	EL
$(-0.85,-0.15)$	$grid_{49}$	0.9405	0.8700	$(-0.85,0.15)$	$grid_{49}$	0.9500	0.8835
	$grid_{100}$	0.9370	0.9220		$grid_{100}$	0.9515	0.9290
	$grid_{169}$	0.9385	0.9290		$grid_{169}$	0.9505	0.9360
	W_{49}	0.9460	0.8705		W_{49}	0.9525	0.8780
	$I_5 \bigotimes W_{49}$	0.9495	0.9380		$I_5 \bigotimes W_{49}$	0.9540	0.9485
	W_{345}	0.9485	0.9335		W_{345}	0.9505	0.9435
$(0.85,-0.15)$	$grid_{49}$	0.9405	0.8830	$(0.85,\ 0.15)$	$grid_{49}$	0.9395	0.8725
	$grid_{100}$	0.9440	0.9185		$grid_{100}$	0.9450	0.9180
	$grid_{169}$	0.9375	0.9220		$grid_{169}$	0.9435	0.9265
	W_{49}	0.9395	0.8815		W_{49}	0.9555	0.8905
	$I_5 \bigotimes W_{49}$	0.9495	0.9395		$I_5 \bigotimes W_{49}$	0.9505	0.9485
	W_{345}	0.9535	0.9465		W_{345}	0.9505	0.9460

Table 2: Coverage probabilities of the LR and EL confidence regions with $\epsilon_i \sim t(5)$

(ρ_1,ρ_2)	$W_n = M_n$	LR	EL	(ρ_1,ρ_2)	$W_n = M_n$	LR	EL
$(-0.85,-0.15)$	$grid_{49}$	0.8430	0.8030	$(-0.85,0.15)$	$grid_{49}$	0.8395	0.7890
	$grid_{100}$	0.8470	0.8725		$grid_{100}$	0.8440	0.8675
	$grid_{169}$	0.8230	0.8845		$grid_{169}$	0.8265	0.8990
	W_{49}	0.8530	0.8070		W_{49}	0.8560	0.8135
	$I_5 \bigotimes W_{49}$	0.8315	0.9105		$I_5 \bigotimes W_{49}$	0.8300	0.9040
	W_{345}	0.8095	0.9065		W_{345}	0.8095	0.9130
$(0.85,-0.15)$	$grid_{49}$	0.8550	0.8085	$(0.85,\ 0.15)$	$grid_{49}$	0.8555	0.8070
	$grid_{100}$	0.8340	0.8705		$grid_{100}$	0.8355	0.8690
	$grid_{169}$	0.8120	0.8870		$grid_{169}$	0.8275	0.8950
	W_{49}	0.8440	0.8175		W_{49}	0.8335	0.8010
	$I_5 \bigotimes W_{49}$	0.8105	0.9025		$I_5 \bigotimes W_{49}$	0.8130	0.9065
	W_{345}	0.8045	0.9100		W_{345}	0.8125	0.9130

Table 3: Coverage probabilities of the LR and EL confidence regions with $\epsilon_i + 4 \sim \chi_4^2$

(ρ_1, ρ_2)	$W_n = M_n$	LR	EL	(ρ_1, ρ_2)	$W_n = M_n$	LR	EL
$(-0.85, -0.15)$	$grid_{49}$	0.8465	0.7990	$(-0.85, 0.15)$	$grid_{49}$	0.8485	0.8010
	$grid_{100}$	0.8355	0.8630		$grid_{100}$	0.8450	0.8735
	$grid_{169}$	0.8495	0.9100		$grid_{169}$	0.8415	0.8985
	W_{49}	0.8570	0.7975		W_{49}	0.8295	0.7710
	$I_5 \bigotimes W_{49}$	0.8420	0.9200		$I_5 \bigotimes W_{49}$	0.8485	0.9140
	W_{345}	0.8615	0.9170		W_{345}	0.8470	0.9205
$(0.85, -0.15)$	$grid_{49}$	0.8420	0.8060	$(0.85, 0.15)$	$grid_{49}$	0.8615	0.8120
	$grid_{100}$	0.8345	0.8720		$grid_{100}$	0.8315	0.8540
	$grid_{169}$	0.8410	0.9025		$grid_{169}$	0.8455	0.8975
	W_{49}	0.8385	0.7895		W_{49}	0.8375	0.7945
	$I_5 \bigotimes W_{49}$	0.8420	0.9050		$I_5 \bigotimes W_{49}$	0.8505	0.9100
	W_{345}	0.8430	0.9240		W_{345}	0.8410	0.9240

Secondly, we used the weight matrix W_{49} related to 49 contiguous planning neighborhoods in Columbus, Ohio, U.S., which appeared in Anselin (1988, p. 188). Thirdly, $W_n = I_5 \bigotimes W_{49}$ was considered, where $\bigotimes$ is kronecker product. This corresponds to the pooling of five separate districts with similar neighboring structures in each district. Finally, weight matrix W_{345} was included in the simulations, which is related to 345 major cities (referring to https://www.gadm.org/download_country_v3.html) in mainland, China.

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 simulations, we took $M_n = W_n$.

Simulation results show that the confidence regions based on LR behave well with coverage probabilities very close to the nominal level 0.95 when the error term ϵ_i is normally distributed, but not well in other cases. The coverage probabilities of the confidence regions based on LR fall to the range [0.8045,0.8560] for t distribution and [0.8295, 0.8615] for χ^2 distribution, which are far from the nominal level 0.95.

We can see, from Tables 1, 2 and 3, the confidence regions based on EL method converge to the nominal level 0.95 as the number of spatial units n is large enough, whether the error term ϵ_i is normally distributed or not. Our simulation results recommend EL method when we can not confirm the normal distribution of the error term.

4 Proofs

In the proof of the main results, we need to use Theorem 1 in Kelejian and Prucha (2001). We now state this result. Let

$$\widetilde{Q}_n = \sum_{i=1}^n \sum_{j=1}^n a_{nij}\epsilon_{ni}\epsilon_{nj} + \sum_{i=1}^n b_{ni}\epsilon_{ni},$$

where ϵ_{ni} are real valued random variables, and the a_{nij} and b_{ni} denote the real valued coefficients of the linear-quadratic form. We need the following assumptions in Lemma 1.

(C1) $\{\epsilon_{ni}, 1 \leq i \leq n\}$ are independent random variables with mean 0 and $\sup_{1\leq i\leq n, n\geq 1} E|\epsilon_{ni}|^{4+\eta_1} < \infty$ for some $\eta_1 > 0$;

(C2) For all $1 \leq i, j \leq n, n \geq 1, a_{nij} = a_{nji}$, $\sup_{1\leq j\leq n, n\geq 1} \sum_{i=1}^n |a_{nij}| < \infty$, and $\sup_{n\geq 1} n^{-1} \sum_{i=1}^n |b_{ni}|^{2+\eta_2} < \infty$ for some $\eta_2 > 0$.

Given the above assumptions (C1) and (C2), the mean and variance of $\tilde{Q}_n$ are given as (e.g. Kelejian and Prucha, 2001)

$$\mu_{\widetilde{Q}} = \sum_{i=1}^n a_{nii}\sigma_{ni}^2,$$

$$\begin{aligned}\sigma^2_{\widetilde{Q}} &= 2\sum_{i=1}^n \sum_{j=1}^n a_{nij}^2\sigma_{ni}^2\sigma_{nj}^2 + \sum_{i=1}^n b_{ni}^2\sigma_{ni}^2 \\ &\quad + \sum_{i=1}^n \{a_{nii}^2(\mu_{ni}^{(4)} - 3\sigma_{ni}^4) + 2b_{ni}a_{nii}\mu_{ni}^{(3)}\}, \qquad (4.1)\end{aligned}$$

with $\sigma_{ni}^2 = E(\epsilon_{ni}^2)$ and $\mu_{ni}^{(s)} = E(\epsilon_{ni}^s)$ for $s = 3, 4$.

Lemma 1. *Suppose that Assumptions C1 and C2 hold true and* $n^{-1}\sigma^2_{\widetilde{Q}} \geq c$ *for some constant* $c > 0$*. Then*

$$\frac{\widetilde{Q}_n - \mu_{\widetilde{Q}}}{\sigma_{\widetilde{Q}}} \xrightarrow{d} N(0,1).$$

Proof. See Theorem 1 and Remark 12 in Kelejian and Prucha (2001).

Lemma 2. *Let* $\eta_1, \eta_2, \cdots, \eta_n$ *be a sequence of stationary random variables, with* $E|\eta_1|^s < \infty$ *for some constants* $s > 0$ *and* $C > 0$*. Then*

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

Proof. Using Borel-Cantelli lemma and following the proof of (2.3) in Owen (1990), one can prove Lemma 2.

Lemma 3. *Suppose that Assumptions (A1) to (A3) are satisfied. Then as* $n \to \infty$,

$$Z_n = \max_{1 \leq i \leq n} ||\omega_i(\theta)|| = o_p(n^{1/2}) \quad a.s., \tag{4.2}$$

$$\Sigma_{k+3}^{-1/2} \sum_{i=1}^{n} \omega_i(\theta) \xrightarrow{d} N(0, I_{k+3}), \tag{4.3}$$

$$n^{-1} \sum_{i=1}^{n} \omega_i(\theta)\omega_i^{\tau}(\theta) = n^{-1}\Sigma_{k+3} + o_p(1), \tag{4.4}$$

$$\sum_{i=1}^{n} ||\omega_i(\theta)||^3 = O_p(n), \tag{4.5}$$

where Σ_{k+3} *is given in* (2.9).

Proof. Note that

$$Z_n \leq \max_{1 \leq i \leq n} ||b_i \epsilon_i|| + \max_{1 \leq i \leq n} \left| \tilde{g}_{ii}(\epsilon_i^2 - \sigma^2) + 2\epsilon_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\epsilon_j + s_i\epsilon_i \right|$$

$$
\begin{aligned}
&+\max_{1\le i\le n}\left|\tilde{h}_{ii}(\epsilon_i^2-\sigma^2)+2\epsilon_i\sum_{j=1}^{i-1}\tilde{h}_{ij}\epsilon_j\right|+\max_{1\le i\le n}|\epsilon_i^2-\sigma^2| \\
\le\ & \max_{1\le i\le n}||b_i\epsilon_i||+\max_{1\le i\le n}|\tilde{g}_{ii}(\epsilon_i^2-\sigma^2)|+\max_{1\le i\le n}\left|2\epsilon_i\sum_{j=1}^{i-1}\tilde{g}_{ij}\epsilon_j\right| \\
&+\max_{1\le i\le n}|\tilde{h}_{ii}(\epsilon_i^2-\sigma^2)|+\max_{1\le i\le n}\left|2\epsilon_i\sum_{j=1}^{i-1}\tilde{h}_{ij}\epsilon_j\right| \\
&+\max_{1\le i\le n}|s_i\epsilon_i|+\max_{1\le i\le n}|\epsilon_i^2-\sigma^2|.
\end{aligned}
$$

By Conditions A1 and A2 and Lemma 2, we have

$$\max_{1\le i\le n}||b_i\epsilon_i||=\max_{1\le i\le n}||b_i||o_p(n^{1/4})=o_p(n^{1/4}),$$

$$\max_{1\le i\le n}||s_i\epsilon_i||=\max_{1\le i\le n}||s_i||o_p(n^{1/4})=o_p(n^{1/4}),$$

$$\max_{1\le i\le n}\left|\epsilon_i\sum_{j=1}^{i-1}\tilde{g}_{ij}\epsilon_j\right|\le(\max_{1\le i\le n}|\epsilon_i|)^2\cdot\max_{1\le i\le n}\left(\sum_{j=1}^{i-1}|\tilde{g}_{ij}|\right)=o_p(n^{1/2}),$$

$$\max_{1\le i\le n}|\tilde{g}_{ii}(\epsilon_i^2-\sigma^2)|=\max_{1\le i\le n}|\tilde{g}_{ii}|o_p(n^{1/2})=o_p(n^{1/2}),$$

$$\max_{1\le i\le n}|\epsilon_i^2-\sigma^2|=o_p(n^{1/2}).$$

Similarly,

$$\max_{1\le i\le n}\left|\epsilon_i\sum_{j=1}^{i-1}\tilde{h}_{ij}\epsilon_j\right|=o_p(n^{1/2}),\ \max_{1\le i\le n}|\tilde{h}_{ii}(\epsilon_i^2-\sigma^2)|=o_p(n^{1/2}).$$

Thus $Z_n=o_p(n^{1/2})$. (4.2) is proved.

For any given $l=(l_1^\tau,l_2,l_3,l_4)^\tau\in R^{k+3}$ with $||l||=1$, where $l_1\in R^k, l_j\in R, j=2,3,4$. Then

$$l^\tau\omega_i(\theta)\ =\ l_1^\tau b_i\epsilon_i+l_2\left\{\tilde{g}_{ii}(\epsilon_i^2-\sigma^2)+2\epsilon_i\sum_{j=1}^{i-1}\tilde{g}_{ij}\epsilon_j+s_i\epsilon_i\right\}$$

$$
\begin{aligned}
&+l_3\left\{\tilde{h}_{ii}(\epsilon_i^2-\sigma^2)+2\epsilon_i\sum_{j=1}^{i-1}\tilde{h}_{ij}\epsilon_j\right\}+l_4(\epsilon_i^2-\sigma^2)\\
&=(l_2\tilde{g}_{ii}+l_3\tilde{h}_{ii}+l_4)(\epsilon_i^2-\sigma^2)+2\epsilon_i\sum_{j=1}^{i-1}(l_2\tilde{g}_{ij}+l_3\tilde{h}_{ij})\epsilon_j+(l_1^\tau b_i+l_2 s_i)\epsilon_i.
\end{aligned}
$$

Thus

$$
\begin{aligned}
\sum_{i=1}^n l^\tau\omega_i(\theta) &= \sum_{i=1}^n(l_2\tilde{g}_{ii}+l_3\tilde{h}_{ii}+l_4)(\epsilon_i^2-\sigma^2)+2\sum_{i=1}^n\sum_{j=1}^{i-1}(l_2\tilde{g}_{ij}+l_3\tilde{h}_{ij})\epsilon_i\epsilon_j\\
&\quad+\sum_{i=1}^n(l_1^\tau b_i+l_2 s_i)\epsilon_i.
\end{aligned}
$$

Let

$$
Q_n=\sum_{i=1}^n\sum_{j=1}^n u_{ij}\epsilon_i\epsilon_j+\sum_{i=1}^n v_i\epsilon_i,
$$

where

$$
u_{ii}=l_2\tilde{g}_{ii}+l_3\tilde{h}_{ii}+l_4, u_{ij}=l_2\tilde{g}_{ij}+l_3\tilde{h}_{ij}(i\neq j), v_i=l_1^\tau b_i+l_2 s_i.
$$

Then

$$
Q_n=\sum_{i=1}^n l^\tau\omega_i(\theta)=\sum_{i=1}^n\left\{u_{ii}(\epsilon_i^2-\sigma^2)+\sum_{j=1}^{i-1}u_{ij}\epsilon_i\epsilon_j+v_i\epsilon_i\right\}.
$$

To obtain the asymptotic distribution of Q_n, we need to check Condition C2. From Condition A2(i), it can be shown that

$$
\sum_{i=1}^n|u_{ij}| \;\leq\; |l_2|\sum_{i=1}^n|\tilde{g}_{ij}|+|l_3|\sum_{i=1}^n|\tilde{h}_{ij}|+|l_4|\leq C.
$$

And by C_r-inequality, we have

$$
n^{-1}\sum_{i=1}^n|v_i|^3 \;\leq\; Cn^{-1}\sum_{i=1}^n|l_1^\tau b_i|^3+Cn^{-1}\sum_{i=1}^n|l_2 s_i|^3. \tag{4.6}
$$

Further,

$$
n^{-1}\sum_{i=1}^n|l_1^\tau b_i|^3\leq C\max_{1\leq i\leq n}||x_i||^3\max_{1\leq i\leq n}\left(\sum_{k=1}^n|b_{ik}|\right)^3\leq C, \tag{4.7}
$$

where b_{ik} is the (i,k)-element of $B_n(\rho_2)$. On the other hand, by Condition A2, it can be shown that

$$n^{-1}\sum_{i=1}^{n}|l_2 s_i|^3 \le C. \tag{4.8}$$

From (4.6)–(4.8), it follows that $n^{-1}\sum_{i=1}^{n}|v_i|^3 \le C$. Therefore, Condition C2 is satisfied.

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

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{n}u_{ij}^2 &= \sum_{i=1}^{n}\Big\{(l_2\tilde{g}_{ii}+l_3\tilde{h}_{ii}+l_4)^2+\sum_{i\neq j}(l_2\tilde{g}_{ij}+l_3\tilde{h}_{ij})^2\Big\}\\ &= 2l_2l_4tr(\tilde{G}_n)+2l_3l_4tr(\tilde{H}_n)+2l_2l_3tr(\tilde{G}_n\tilde{H}_n)\\ &\quad +nl_4^2+l_2^2tr(\tilde{G}_n{}^2)+l_3^2tr(\tilde{H}_n{}^2),\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{n}u_{ii}^2 &= \sum_{i=1}^{n}(l_2\tilde{g}_{ii}+l_3\tilde{h}_{ii}+l_4)^2\\ &= l_2^2||Vec(diag(\tilde{G}_n))||^2+l_3^2||Vec(diag(\tilde{H}_n))||^2+nl_4^2\\ &\quad +2l_2l_4tr(\tilde{G}_n)+2l_3l_4tr(\tilde{H}_n)+2l_2l_3Vec^{\tau}(diag(\tilde{G}_n))Vec(diag(\tilde{H}_n)),\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{n}v_i^2 &= \sum_{i=1}^{n}(l_1^{\tau}b_i+l_2s_i)^2\\ &= l_1^{\tau}\Big(\sum_{i=1}^{n}b_ib_i^{\tau}\Big)l_1+l_2^2\sum_{i=1}^{n}s_i^2+2l_1^{\tau}\Big(\sum_{i=1}^{n}b_is_i\Big)l_2\\ &= l_1^{\tau}\{B_n(\rho_2)X_n\}^{\tau}B_n(\rho_2)X_nl_1\\ &\quad +l_2^2\{B_n(\rho_2)W_nA_n^{-1}(\rho_1)X_n\beta\}^{\tau}B_n(\rho_2)W_nA_n^{-1}(\rho_1)X_n\beta\\ &\quad +2l_1^{\tau}l_2\{B_n(\rho_2)X_n\}^{\tau}B_n(\rho_2)W_nA_n^{-1}(\rho_1)X_n\beta,\end{aligned}$$

and that

$$\begin{aligned}\sum_{i=1}^{n}u_{ii}v_i &= \sum_{i=1}^{n}(l_2\tilde{g}_{ii}+l_3\tilde{h}_{ii}+l_4)(l_1^{\tau}b_i+l_2s_i)\\ &= l_1^{\tau}l_2\sum_{i=1}^{n}b_i\tilde{g}_{ii}+l_2^2\sum_{i=1}^{n}\tilde{g}_{ii}s_i+l_1^{\tau}l_3\sum_{i=1}^{n}b_i\tilde{h}_{ii}+l_2l_3\sum_{i=1}^{n}s_i\tilde{h}_{ii}\end{aligned}$$

$$
\begin{aligned}
& +l_1^\tau l_4 \sum_{i=1}^n b_i + l_2 l_4 \sum_{i=1}^n s_i \\
= \;& l_1^\tau l_2 \{B_n(\rho_2) X_n\}^\tau Vec(diag\tilde{G}_n) \\
& + l_2^2 \{B_n(\rho_2) W_n A_n^{-1}(\rho_1) X_n \beta\}^\tau Vec(diag\tilde{G}_n) \\
& + l_1^\tau l_3 \{B_n(\rho_2) X_n\}^\tau Vec(diag\tilde{H}_n) \\
& + l_2 l_3 \{B_n(\rho_2) W_n A_n^{-1}(\rho_1) X_n \beta\}^\tau Vec(diag\tilde{H}_n) \\
& + l_1^\tau l_4 \{B_n(\rho_2) X_n\}^\tau \mathbf{1}_n + l_2 l_4 \{B_n(\rho_2) W_n A_n^{-1}(\rho_1) X_n \beta\}^\tau \mathbf{1}_n,
\end{aligned}
$$

where $\mathbf{1}_n$ is the n-dimensional vector with 1 as its components. It follows from (4.1) that the variance of Q_n is

$$
\begin{aligned}
\sigma_Q^2 \;=\;& 2\sum_{i=1}^n \sum_{j=1}^n u_{ij}^2 \sigma^4 + \sum_{i=1}^n v_i^2 \sigma^2 + \sum_{i=1}^n \{u_{ii}^2(\mu_4 - 3\sigma^4) + 2u_{ii} v_i \mu_3\} \\
=\;& 2\sigma^4 \Big\{ 2l_2 l_4 tr(\tilde{G}_n) + 2l_3 l_4 tr(\tilde{H}_n) + 2l_2 l_3 tr(\tilde{G}_n \tilde{H}_n) \\
& + n l_4^2 + l_2^2 tr(\tilde{G}_n{}^2) + l_3^2 tr(\tilde{H}_n{}^2) \Big\} + \sigma^2 \Big[l_1^\tau \{B_n(\rho_2) X_n\}^\tau B_n(\rho_2) X_n l_1 \\
& + l_2^2 \{B_n(\rho_2) W_n A_n^{-1}(\rho_1) X_n \beta\}^\tau B_n(\rho_2) W_n A_n^{-1}(\rho_1) X_n \beta \\
& + 2 l_1^\tau l_2 \{B_n(\rho_2) X_n\}^\tau B_n(\rho_2) W_n A_n^{-1}(\rho_1) X_n \beta \Big] \\
& + (\mu_4 - 3\sigma^4) \Big\{ l_2^2 ||Vec(diag(\tilde{G}_n))||^2 + l_3^2 ||Vec(diag(\tilde{H}_n))||^2 + n l_4^2 \\
& + 2l_2 l_4 tr(\tilde{G}_n) + 2l_3 l_4 tr(\tilde{H}_n) + 2 l_2 l_3 Vec^\tau(diag(\tilde{G}_n)) Vec(diag(\tilde{H}_n)) \Big\} \\
& + 2\mu_3 \Big[l_1^\tau l_2 \{B_n(\rho_2) X_n\}^\tau Vec(diag\tilde{G}_n) \\
& + l_2^2 \{B_n(\rho_2) W_n A_n^{-1}(\rho_1) X_n \beta\}^\tau Vec(diag\tilde{G}_n) \\
& + l_1^\tau l_3 \{B_n(\rho_2) X_n\}^\tau Vec(diag\tilde{H}_n) \\
& + l_2 l_3 \{B_n(\rho_2) W_n A_n^{-1}(\rho_1) X_n \beta\}^\tau Vec(diag\tilde{H}_n) \\
& + l_1^\tau l_4 \{B_n(\rho_2) X_n\}^\tau \mathbf{1}_n + l_2 l_4 \{B_n(\rho_2) W_n A_n^{-1}(\rho_1) X_n \beta\}^\tau \mathbf{1}_n \Big] \\
=\;& l^\tau \Sigma_{k+3} l,
\end{aligned}
$$

where Σ_{k+3} is given in (2.9). From Condition A3, one can see that $n^{-1}\sigma_Q^2 \geq c_1 > 0$. From Lemma 1, we have

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

Noting that $E(Q_n) = 0$, we thus have (4.3).

Next we will prove (4.4), i. e.

$$n^{-1}\sum_{i=1}^{n}\{l^\tau \omega_i(\theta)\}^2 = n^{-1}\sigma_Q^2 + o_p(1). \tag{4.9}$$

Let

$$\begin{aligned} Y_{in} &= l^\tau \omega_i(\theta) \\ &= u_{ii}(\epsilon_i^2 - \sigma^2) + 2\sum_{j=1}^{i-1} u_{ij}\epsilon_i\epsilon_j + v_i\epsilon_i \\ &= u_{ii}(\epsilon_i^2 - \sigma^2) + B_i\epsilon_i, \end{aligned} \tag{4.10}$$

where $B_i = 2\sum_{j=1}^{i-1} u_{ij}\epsilon_j + v_i$. Let $\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_i = \sigma(\epsilon_1, \epsilon_2, \cdots, \epsilon_i), 1 \leq i \leq n$. Then $\{Y_{in}, \mathcal{F}_i, 1 \leq i \leq n\}$ form a martingale difference array. Note that

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}\{l^\tau \omega_i(\theta)\}^2 - n^{-1}\sigma_Q^2 = n^{-1}\sum_{i=1}^{n}(Y_{in}^2 - EY_{in}^2) \\ = \;& n^{-1}\sum_{i=1}^{n}\{Y_{in}^2 - E(Y_{in}^2|\mathcal{F}_{i-1}) + E(Y_{in}^2|\mathcal{F}_{i-1}) - EY_{in}^2\} \\ = \;& n^{-1}S_{n1} + n^{-1}S_{n2}, \end{aligned} \tag{4.11}$$

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

$$n^{-1}S_{n1} = o_p(1), \tag{4.12}$$

and

$$n^{-1}S_{n2} = o_p(1). \tag{4.13}$$

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

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

Thus

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

It follows that

$$\begin{aligned} & n^{-2}ES_{n1}^2 = n^{-2}\sum_{i=1}^n E\{Y_{in}^2 - E(Y_{in}^2|\mathcal{F}_{i-1})\}^2 \\ = \ & n^{-2}\sum_{i=1}^n E[u_{ii}^2\{(\epsilon_i^2 - \sigma^2)^2 - E(\epsilon_i^2 - \sigma^2)^2\} + B_i^2(\epsilon_i^2 - \sigma^2) \\ & +2u_{ii}B_i(\epsilon_i^3 - \sigma^2\epsilon_i - \mu_3)]^2 \\ \le \ & Cn^{-2}\sum_{i=1}^n E[u_{ii}^4\{(\epsilon_i^2 - \sigma^2)^2 - E(\epsilon_i^2 - \sigma^2)^2\}^2] + Cn^{-2}\sum_{i=1}^n E\{B_i^4(\epsilon_i^2 - \sigma^2)^2\} \\ & +Cn^{-2}\sum_{i=1}^n E\{u_{ii}^2B_i^2(\epsilon_i^3 - \sigma^2\epsilon_i - \mu_3)^2\}. \end{aligned} \tag{4.14}$$

By Condition A1, we have

$$n^{-2}\sum_{i=1}^n E[u_{ii}^4\{(\epsilon_i^2 - \sigma^2)^2 - E(\epsilon_i^2 - \sigma^2)^2\}^2] \le Cn^{-1} \to 0, \tag{4.15}$$

and

$$\begin{aligned} & n^{-2}\sum_{i=1}^n E\{B_i^4(\epsilon_i^2 - \sigma^2)^2\} \le Cn^{-2}\sum_{i=1}^n E(\sum_{j=1}^{i-1} u_{ij}\epsilon_j + v_i)^4 \\ \le \ & Cn^{-2}\sum_{i=1}^n E(\sum_{j=1}^{i-1} u_{ij}\epsilon_j)^4 + Cn^{-2}\sum_{i=1}^n v_i^4 \\ \le \ & Cn^{-2}\sum_{i=1}^n\sum_{j=1}^{i-1} u_{ij}^4\mu_4 + Cn^{-2}\sum_{i=1}^n(\sum_{j=1}^{i-1} u_{ij}^2\sigma^2)^2 + Cn^{-2}\sum_{i=1}^n(l_1^\tau b_i + l_2 s_i)^4 \\ \le \ & Cn^{-1} \to 0. \end{aligned} \tag{4.16}$$

Similarly, one can show that

$$n^{-2}\sum_{i=1}^n E\{u_{ii}^2B_i^2(\epsilon_i^3 - \sigma^2\epsilon_i - \mu_3)^2\} \to 0. \tag{4.17}$$

From (4.14)-(4.17), we have $n^{-2}ES_{n1}^2 \to 0$. Furthermore,

$$\begin{aligned} EY_{in}^2 &= E\{E(Y_{in}^2|\mathcal{F}_{i-1})\} = u_{ii}^2 E(\epsilon_i^2-\sigma^2)^2 + \sigma^2 E(B_i^2) + 2u_{ii}\mu_3 E(B_i) \\ &= u_{ii}^2 E(\epsilon_i^2-\sigma^2)^2 + \sigma^2(4\sum_{j=1}^{i-1} u_{ij}^2\sigma^2 + v_i^2) + 2u_{ii}\mu_3 v_i. \end{aligned}$$

Thus,

$$\begin{aligned} &n^{-2}ES_{n2}^2 = n^{-2}E[\sum_{i=1}^n \{E(Y_{in}^2|\mathcal{F}_{i-1}) - EY_{in}^2\}]^2 \\ &= n^{-2}E[\sum_{i=1}^n \{B_i^2\sigma^2 - \sigma^2(4\sum_{j=1}^{i-1} u_{ij}^2\sigma^2 + v_i^2) + 2u_{ii}\mu_3(B_i - v_i)\}]^2 \\ &= n^{-2}\sum_{i=1}^n E[\sigma^2\{(2\sum_{j=1}^{i-1} u_{ij}\epsilon_j)^2 - 4\sum_{j=1}^{i-1} u_{ij}^2\sigma^2\} + 4(\sum_{j=1}^{i-1} u_{ij}\epsilon_j)v_i\sigma^2 \\ &\quad +2u_{ii}\mu_3(2\sum_{j=1}^{i-1} u_{ij}\epsilon_j)]^2 \\ &\le Cn^{-2}\sum_{i=1}^n E\{\sigma^2(\sum_{j=1}^{i-1} u_{ij}\epsilon_j)^2 - \sum_{j=1}^{i-1} u_{ij}^2\sigma^2\}^2 + Cn^{-2}\sum_{i=1}^n E\{(\sum_{j=1}^{i-1} u_{ij}\epsilon_j)v_i\sigma^2\}^2 \\ &\quad +Cn^{-2}\sum_{i=1}^n E\{2u_{ii}\mu_3(\sum_{j=1}^{i-1} u_{ij}\epsilon_j)\}^2. \end{aligned} \tag{4.18}$$

Note that

$$\begin{aligned} &n^{-2}\sum_{i=1}^n E[\sigma^2\{(\sum_{j=1}^{i-1} u_{ij}\epsilon_j)^2 - \sum_{j=1}^{i-1} u_{ij}^2\sigma^2\}]^2 \le n^{-2}\sigma^4\sum_{i=1}^n E(\sum_{j=1}^{i-1} u_{ij}\epsilon_j)^4 \\ &\le Cn^{-2}\sum_{i=1}^n\sum_{j=1}^{i-1} u_{ij}^4\mu_4 + Cn^{-2}\sum_{i=1}^n(\sum_{j=1}^{i-1} u_{ij}^2\sigma^2)^2 \le Cn^{-1} \to 0, \end{aligned} \tag{4.19}$$

$$n^{-2}\sum_{i=1}^n E\{(\sum_{j=1}^{i-1} u_{ij}\epsilon_j)v_i\sigma^2\}^2 = n^{-2}\sigma^6\sum_{i=1}^n v_i^2\sum_{j=1}^{i-1} u_{ij}^2 \le Cn^{-2} \to 0, \tag{4.20}$$

and

$$n^{-2}\sum_{i=1}^n E\{2u_{ii}\mu_3(\sum_{j=1}^{i-1} u_{ij}\epsilon_j)\}^2 = 4\mu_3^2\sigma^2 n^{-2}\sum_{i=1}^n u_{ii}^2\sum_{j=1}^{i-1} u_{ij}^2 \le Cn^{-1} \to 0, \tag{4.21}$$

where we have used Conditions A1 and A2. From (4.18)–(4.21), we have $n^{-2}ES_{n2}^2 \to 0$. The proof of (4.9) is thus complete.

Finally, we will prove (4.5). Note that

$$\begin{aligned}\sum_{i=1}^n E||\omega_i(\theta)||^3 \quad \le \quad & \sum_{i=1}^n E||b_i\epsilon_i||^3 + \sum_{i=1}^n E|\tilde{g}_{ii}(\epsilon_i^2-\sigma^2) + 2\epsilon_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\epsilon_j + s_i\epsilon_i|^3 \\ & + \sum_{i=1}^n E|\tilde{h}_{ii}(\epsilon_i^2-\sigma^2) + 2\epsilon_i \sum_{j=1}^{i-1} \tilde{h}_{ij}\epsilon_j|^3 \\ & + \sum_{i=1}^n E|\epsilon_i^2-\sigma^2|^3. \end{aligned} \tag{4.22}$$

By Conditions A1 and A2,

$$\sum_{i=1}^n E||b_i\epsilon_i||^3 = O(n), \tag{4.23}$$

$$\begin{aligned} & \sum_{i=1}^n E\left|\tilde{g}_{ii}(\epsilon_i^2-\sigma^2) + 2\epsilon_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\epsilon_j + s_i\epsilon_i\right|^3 \\ \le \quad & C\sum_{i=1}^n E|\tilde{g}_{ii}(\epsilon_i^2-\sigma^2)|^3 + C\sum_{i=1}^n E\left|2\epsilon_i \sum_{j=1}^{i-1} \tilde{g}_{ij}\epsilon_j\right|^3 + C\sum_{i=1}^n E|s_i\epsilon_i|^3 \\ \le \quad & C\sum_{i=1}^n E|\tilde{g}_{ii}(\epsilon_i^2-\sigma^2)|^3 + C\sum_{i=1}^n E|\epsilon_i|^3 \sum_{j=1}^{i-1} E|\tilde{g}_{ij}\epsilon_j|^3 \\ & +C\sum_{i=1}^n E|\epsilon_i|^3 \left\{\sum_{j=1}^{i-1} E(\tilde{g}_{ij}\epsilon_j)^2\right\}^{3/2} + C\sum_{i=1}^n E|s_i\epsilon_i|^3 = O(n), \end{aligned} \tag{4.24}$$

$$\sum_{i=1}^n E|\epsilon_i^2-\sigma^2|^3 = O(n). \tag{4.25}$$

From (4.22)–(4.25),we have

$$\sum_{i=1}^n E||\omega_i(\theta)||^3 = O(n). \tag{4.26}$$

Further, using (4.26) and Markov inequality, we obtain $\sum_{i=1}^n ||\omega_i(\theta)||^3 = O_p(n)$. Thus (4.5) is proved.

We now in the position to prove the main results in this article.

Proof of Theorem 1. Let $\lambda = \lambda(\theta), \rho_0 = ||\lambda||, \lambda = \rho_0\eta_0$. From (2.8), we have

$$\frac{\eta_0^\tau}{n}\sum_{j=1}^n \omega_j(\theta) - \frac{\rho_0}{n}\sum_{j=1}^n \frac{(\eta_0^\tau\omega_j(\theta))^2}{1+\lambda^\tau\omega_j(\theta)} = 0.$$

It follows that

$$|\eta_0^\tau\bar{\omega}| \geq \frac{\rho_0}{1+\rho_0 Z_n}\lambda_{min}(S_0),$$

where Z_n is defined in (4.2), $\bar{\omega} = n^{-1}\sum_{i=1}^n \omega_i(\theta), S_0 = n^{-1}\sum_{i=1}^n \omega_i(\theta)\omega_i^\tau(\theta)$. That is

$$|\eta_0^\tau \Sigma_{k+3}^{1/2}\Sigma_{k+3}^{-1/2}\bar{\omega}| \geq \frac{\rho_0}{1+\rho_0 Z_n}\lambda_{min}(S_0),$$

i. e.

$$\lambda_{max}(\Sigma_{k+3}^{1/2})||\eta_0||\cdot||\Sigma_{k+3}^{-1/2}\bar{\omega}|| \geq \frac{\rho_0}{1+\rho_0 Z_n}\lambda_{min}(S_0).$$

Combining with Lemma 3 and Condition A3, we have

$$\frac{\rho_0}{1+\rho_0 Z_n} = O_p(n^{-1/2}).$$

Therefore, from Lemma 3,

$$\rho_0 = O_p(n^{-1/2}).$$

Let $\gamma_i = \lambda^\tau\omega_i(\theta)$. Then

$$\max_{1\leq i\leq n}|\gamma_i| = o_p(1). \tag{4.27}$$

Using (2.8) again, we have

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{j=1}^n \frac{\omega_j(\theta)}{1+\lambda^\tau\omega_j(\theta)} \\ &= \frac{1}{n}\sum_{j=1}^n \omega_j(\theta) - \frac{1}{n}\sum_{j=1}^n \frac{\omega_j(\theta)\{\lambda^\tau\omega_j(\theta)\}}{1+\lambda^\tau\omega_j(\theta)} \\ &= \frac{1}{n}\sum_{j=1}^n \omega_j(\theta) - \{\frac{1}{n}\sum_{j=1}^n \omega_j(\theta)\omega_j(\theta)^\tau\}\lambda + \frac{1}{n}\sum_{j=1}^n \frac{\omega_j(\theta)\{\lambda^\tau\omega_j(\theta)\}^2}{1+\lambda^\tau\omega_j(\theta)} \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{j=1}^{n}\omega_j(\theta) - \{\frac{1}{n}\sum_{j=1}^{n}\omega_j(\theta)\omega_j(\theta)^\tau\}\lambda + \frac{1}{n}\sum_{j=1}^{n}\frac{\omega_j(\theta)\gamma_j^2}{1+\gamma_j} \\
&= \overline{\omega} - S_0\lambda + \frac{1}{n}\sum_{j=1}^{n}\frac{\omega_j(\theta)\gamma_j^2}{1+\gamma_j}.
\end{aligned}
$$

Combining with Lemma 3 and Condition A3, we may write

$$\lambda = S_0^{-1}\overline{\omega} + \varsigma, \tag{4.28}$$

where $||\varsigma||$ is bounded by

$$n^{-1}\sum_{j=1}^{n}||\omega_j(\theta)||^3||\lambda||^2 = O_p(n^{-1}).$$

By (4.27) we may expand $\log(1+\gamma_i) = \gamma_i - \gamma_i^2/2 + \nu_i$ where, for some finite $B > 0$,

$$P(|\nu_i| \leq B|\gamma_i|^3, 1 \leq i \leq n) \to 1, \text{ as } n \to \infty.$$

Therefore, from (2.7), (4.28) and Taylor expansion, we have

$$
\begin{aligned}
\ell_n(\theta) &= 2\sum_{j=1}^{n}\log(1+\gamma_j) = 2\sum_{j=1}^{n}\gamma_j - \sum_{j=1}^{n}\gamma_j^2 + 2\sum_{j=1}^{n}\nu_j \\
&= 2n\lambda^\tau\overline{\omega} - n\lambda^\tau S_0\lambda + 2\sum_{j=1}^{n}\nu_j \\
&= 2n(S_0^{-1}\overline{\omega})^\tau\overline{\omega} + 2n\varsigma^\tau\overline{\omega} - n\overline{\omega}^\tau S_0^{-1}\overline{\omega} - \\
&\quad 2n\varsigma^\tau\overline{\omega} - n\varsigma^\tau S_0\varsigma + 2\sum_{j=1}^{n}\nu_j \\
&= n\overline{\omega}^\tau S_0^{-1}\overline{\omega} - n\varsigma^\tau S_0\varsigma + 2\sum_{j=1}^{n}\nu_j \\
&= \{n\Sigma_{k+2}^{-1/2}\overline{\omega}\}^\tau\{n\Sigma_{k+2}^{-1/2}S_0\Sigma_{k+2}^{-1/2}\}^{-1}\{n\Sigma_{k+2}^{-1/2}\overline{\omega}\} \\
&\quad -n\varsigma^\tau S_0\varsigma + 2\sum_{j=1}^{n}\nu_j.
\end{aligned}
$$

From Lemma 3 and Condition A3, we have

$$\{n\Sigma_{k+3}^{-1/2}\overline{\omega}\}^\tau\{n\Sigma_{k+3}^{-1/2}S_0\Sigma_{k+3}^{-1/2})\}^{-1}\{n\Sigma_{k+3}^{-1/2}\overline{\omega}\} \xrightarrow{d} \chi_{k+3}^2.$$

On the other hand, using Lemma 3 and above derivations, we can see that $n\varsigma^\tau S_0\varsigma = O_p(n^{-1}) = o_p(1)$ and

$$|\sum_{j=1}^{n}\nu_j| \leq B||\lambda||^3 \sum_{j=1}^{n} ||\omega_j(\theta)||^3 = O_p(n^{-1/2}) = o_p(1).$$

The proof of Theorem 1 is thus complete.

Acknowledgements. This work was partially supported by the National Natural Science Foundation of China (11671102), the Natural Science Foundation of Guangxi (2016GXNSFAA3800163, 2017GXNSFAA198349) and the Program on the High Level Innovation Team and Outstanding Scholars in Universities of Guangxi Province.

References

ANSELIN, L. (1988). *Spatial Econometrics: Methods and Models.* Kluwer Academic Publishers, The Netherlands.

ANSELIN, L. and BERA, A.K. (1998). *Spatial dependence in linear regression models with an introduction to spatial econometrics, Handbook of Applied Economics Statistics, ed. by Ullah A. and Giles, D. E A.* Marcel Dekker, New York.

CHEN, J. and QIN, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika* **80**, 107–116.

CHEN, S.X. and KEILEGOM, I.V. (2009). A review on empirical likelihood for regressions (with discussions). *Test* **3**, 415–447.

CLIFF, A.D. and ORD, J.K. (1973). *Spatial Autocorrelation.* Pion Ltd, London.

CRESSIEN, N. (1993). *Statistics for Spatial Data.* Wiley, New York.

JIN, F. and LEE, L.F. (2019). GEL Estimation and tests of spatial autoregressive models. Journal of Econometrics, Forthcoming.

KELEJIAN, H.H. and PRUCHA, I.R. (2001). On the asymptotic distribution of the Moran I test statistic with applications. *J. Econ.* **104**, 219–257.

LEE, L.F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* **72**, 1899–1925.

LIU, X., LEE, L.F. and BOLLINGER, C.R. (2010). An efficient GMM estimator of spatial autoregressive models. *J. Econ.* **159**, 303–319.

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. *Ann. Statist.* **18**, 90–120.

OWEN, A.B. (2001). *Empirical Likelihood.* Chapman & Hall, London.

qin, j. and lawless, j. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300–325.

wu, c.b. (2004). Weighted empirical likelihood inference. *Statist. Probab. Lett.* **66**, 67–79.

zhong, b. and rao, j.n.k. (2000). Empirical likelihood inference under stratified random sampling using auxiliary population information. *Biometrika* **87**, 929–938.

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

Yongsong Qin
Department of Statistics,
Guangxi Normal University,
Guilin, Guangxi 541004, China
E-mail: ysqin@gxnu.edu.cn

Paper received: 6 January 2019.