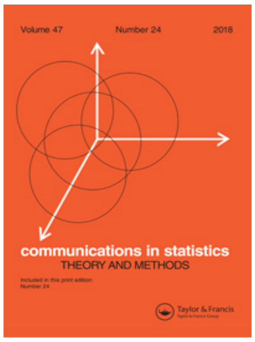

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

Empirical likelihood for panel data models with spatial errors

Yinghua Li, Yuan Li & Yongsong Qin

To cite this article: Yinghua Li, Yuan Li & Yongsong Qin (2020): Empirical likelihood for panel data models with spatial errors, Communications in Statistics - Theory and Methods, DOI: 10.1080/03610926.2020.1780449

To link to this article: https://doi.org/10.1080/03610926.2020.1780449

Published online: 18 Jun 2020.

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
https://doi.org/10.1080/03610926.2020.1780449

Empirical likelihood for panel data models with spatial errors

Yinghua Li[a,b], Yuan Li[a], and Yongsong Qin[b]

[a]College of Economics and Statistics, Guangzhou University, Guangzhou, Guangdong, China; [b]College of Mathematics and Statistics, Guangxi Normal University, Guilin, Guangxi, China

ABSTRACT
For panel data models with spatial errors, the empirical likelihood ratio statistics are constructed for the parameters of the models. 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 of the models. A simulation study is conducted to show the performance of the proposed method.

ARTICLE HISTORY
Received 16 January 2020
Accepted 5 June 2020

KEYWORDS
Panel data; spatial error; empirical likelihood; confidence region

AMS 2010 SUBJECT CLASSIFICATION
Primary 62G05; secondary 62E20

1. Introduction

Linear regression models are the most important statistical models for explaining the relationship between response and explanatory variables. Whenever the variables in a linear regression model refer to attributes of a particular location (height of a plant, population of a country, position in a social network, etc.), one often allows for correlation among the errors (disturbances) by assuming that the errors follow a spatial autoregressive (SAR) correlation (e.g., Dow, Burton, and White 1982; Ord 1975; Krämer and Donninger 1987; Anselin and Bera 1998). These models deal with data in different locations with fixed time point, which are called spatial models. If the data reflect various times and locations, they are called spatial panel data.

In recent years, spatial panel data models studied in Anselin (1988) have drawn more and more attention in empirical economic research, as they offer researchers extended modeling possibilities as compared to the single-equation cross-sectional setting and contain more variation and less collinearity among the variables. Baltagi, Song, and Koh (2003) consider panel regression models with SAR disturbances and focus on the test of spatial correlation for the models. Kapoor, Kelejian, and Prucha (2007) provide a rigorous theoretical framework for analysis of spatial panel models. Lee and Yu (2010a) propose the maximum likelihood (ML) estimation for panel models with both spatial lag and spatial disturbances. Some related recent developments are in Anselin (2001),

CONTACT Yongsong Qin ysqin@mailbox.gxnu.edu.cn College of Mathematics and Statistics, Guangxi Normal University, Guilin, Guangxi 541004, China.

Elhorst (2005, 2010), Anselin, Le Gallo, and Jayet (2008), Yu et al. (2008), Lee and Yu (2010b, 2013), Su and Yang (2015), Qu, Lee, and Yu (2017), among others.

In this article, we study the following special spatial panel data model. Suppose that there are n individual units and T time periods. We consider the following panel data model with spatial error (e.g., Chapter 10 in Anselin (1988)):

$$y_t = X_t\beta_t + \epsilon_t, \epsilon_t = \lambda_t W_n \epsilon_t + \mu_t, t = 1, 2, ..., T \quad (1)$$

where y_t is an n-dimensional column vector of observed dependent variables, X_t is an $n \times K$ matrix of explanatory variables, and β_t is an $K \times 1$ vector of regression coefficients. ϵ_t is an $n \times 1$ vector of errors, W_n is an $n \times n$ spatial weighting matrix of constants, $\mu_t = (\mu_{t1}, ..., \mu_{tn})^\tau$ is an $n \times 1$ column vector, and $\{\mu_{ti}\}$ are i.i.d. across t and i with zero mean and variance σ^2. Model (1) is also called spatial seemingly unrelated regressions (SURs) model, originally suggested by Zellner (1962), and it is designed for empirical situations where a limited degree of simultaneity is present in the form of dependence between the errors in different equations. SUR models are extensions of linear regression models which allow correlated errors between equations, and have been widely used in many research areas, obviously including spatial analysis. Anselin (1988) extends an SUR model to a spatial environment. By incorporating SAR into the error term, the model exhibits spatial autocorrelations across observations. Previously, the development in testing and estimation of SUR models has been summarized in Anselin (1988). When $T = 1$, these models are studied by Cliff and Ord (1973), Ord (1975), Krämer and Donninger (1987), and Kelejian and Prucha (1999), among others. Recently, Wang and Kockelman (2007) derived the ML estimator (under the normality assumption) of an SUR error component panel data model with SAR disturbances. Baltagi and Pirotte (2011) considered various estimators for panel data SUR with spatial error correlation. In terms of testing, Mur, Lòpez, and Herrera (2010) developed a set of Lagrange multipliers to test for the presence of spatial effects in a standard spatial SUR model. Some recent research work on SUR models and their applications can be found in Jiang, Qian, and Sun (2020), Hou and Zhao (2019), Kubáček (2013), Kurata and Matsuura (2016), Sun, Ke, and Tian (2014), Tian (2010), Zhao and Xu (2017), among others.

There are two competing estimation approaches for the parameters in spatial models. One is the ML method (e.g., Anselin 1988). The other is the computationally more efficient method, the generalized method of moments (GMMs) by Kelejian and Prucha (1999). The asymptotic properties of the maximum likelihood estimator (MLE) and the GMM estimator for the spatial models are investigated by Anselin (1988) and Kelejian and Prucha (1999), respectively. These methods may be readily extended to spatial SUR models. However, it may not be easy to use these normal approximation methods to construct confidence region for the parameters in the SUR model as the asymptotic covariance in the asymptotic distribution is unknown. More importantly, the accuracy of the normal approximation-based confidence region of the parameters in the model may be affected by estimating the asymptotic covariance. 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 SUR models. 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. Owen (1991) has used the EL method to construct confidence regions for the vector of regression parameters in a linear model with independent errors. 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 idea in using the EL method for the spatial SUR models is to introduce a martingale sequence to transform the linear quadratic form of the estimating equations (e.g., Equation (5)–(7)) for the spatial SUR models into a linear form. It is interesting to note that the estimation equations for other spatial panel data models may have the linear quadratic forms. Therefore, this approach of transformation also opens a way to use EL methods to more general spatial panel data models.

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

2. Main results

We continue with the Model (1). With $t = 1, 2, \ldots, T$, Model (1) can be written into a matrix form as follows:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} X_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & X_2 & 0 & \cdots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & 0 & X_T \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_T \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix}$$

with

$$\begin{pmatrix} B_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & B_2 & 0 & \cdots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & 0 & B_T \end{pmatrix} \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_T \end{pmatrix}$$

or

$$Y = X\beta + \epsilon \tag{2}$$

with

$$B\epsilon = \mu \tag{3}$$

where $B_t = (I_n - \lambda_t W_n), t = 1, 2, \ldots, T, B = [I_{nT} - (\Lambda \otimes W_n)]$, $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_T)$ is a $T \times T$ diagonal matrix, and $\otimes$ is the Kronecker product,

$$Y_{(nT)\times 1} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}, X_{(nT)\times(KT)} = \begin{pmatrix} X_1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & X_2 & 0 & \cdots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & 0 & X_T \end{pmatrix}$$

$$\beta_{(KT)\times 1} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_T \end{pmatrix}, \epsilon_{(nT)\times 1} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix}, \mu_{(nT)\times 1} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_T \end{pmatrix}$$

Based on Models (2) and (3), we adopt the quasi-maximum likelihood method (QMLE) to estimate $\theta = (\beta_1^\tau, ..., \beta_T^\tau, \lambda_1, ..., \lambda_T, \sigma^2)^\tau$. Under the assumption of normality, the log-likelihood function (ignoring constants) is as follows:

$$L(\theta) = -\frac{nT}{2}\log\sigma^2 + \sum_{t=1}^{T} \log|B_t| - \frac{1}{2\sigma^2}\mu^\tau\mu \tag{4}$$

In order to derive the EL statistic of θ, one can show that:

$$\begin{aligned} \partial L(\theta)/\partial\beta &= \sigma^{-2}X^\tau B^\tau \mu \\ \partial L(\theta)/\partial\lambda_t &= -tr(W_nB_t^{-1}) + \sigma^{-2}\mu^\tau(E^{tt}\otimes W_n)B^{-1}\mu, \;\; t = 1, ..., T \\ \partial L(\theta)/\partial\sigma^2 &= -\frac{nT}{2\sigma^2} + \frac{1}{2\sigma^4}\mu^\tau\mu \end{aligned}$$

where E^{tt} is a $T \times T$ matrix of zeros, except the (t, t)-element which has the value 1. Letting above derivatives be 0, we obtain the following estimating equations:

$$X^\tau B^\tau \mu = 0 \tag{5}$$

$$-\sigma^2 tr(W_nB_t^{-1}) + \mu^\tau(E^{tt}\otimes W_n)B^{-1}\mu = 0, \;\; t = 1, ..., T \tag{6}$$

$$-nT\sigma^2 + \mu^\tau\mu = 0 \tag{7}$$

We observe that the above estimating equations include the quadratic forms of μ. To use the EL method, we need to change the quadratic forms into the linear forms of well-behaved random variables such as martingale difference arrays. To this end, we let $G_{nt} = (E^{tt}\otimes W_n)B^{-1}$ and $\tilde{G}_{nt} = \frac{1}{2}(G_{nt} + G_{nt}^\tau)$. Use $g_{ij,t}, \tilde{g}_{ij,t}$, and b_i to denote the (i, j) element of the matrix G_{nt}, the (i, j) element of the matrix $\tilde{G}_{nt}$, and the i-th column of the matrix $X^\tau B^\tau$, respectively, and adapt the convention that any sum with an upper index of less than one is zero. To deal with the quadratic form in (6), we follow Kelejian and Prucha (2001) to introduce a martingale difference array. Let

$$e_{(nT)\times 1} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_{nT} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_T \end{pmatrix}$$

and define the σ-fields: $\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_i = \sigma(e_1, e_2, ..., e_i), 1 \le i \le nT$. Let

$$\tilde{M}_{in} = \tilde{g}_{ii,t}(e_i^2 - \sigma^2) + 2e_i\sum_{j=1}^{i-1}\tilde{g}_{ij,t}e_j \tag{8}$$

Then, $\mathcal{F}_{i-1} \subseteq \mathcal{F}_i, \tilde{M}_{in}$ is $\mathcal{F}_i-$ measurable and $E(\tilde{M}_{in}|\mathcal{F}_{i-1}) = 0$. Thus, $\{\tilde{M}_{in}, \mathcal{F}_i, 1 \le i \le nT\}$ form a martingale difference array and

$$\mu^\tau\tilde{G}_{nt}\mu - \sigma^2 tr(\tilde{G}_{nt}) = \sum_{i=1}^{nT}\tilde{M}_{in} \tag{9}$$

Based on (5)–(9), we propose the following EL ratio statistic for $\theta \in R^{(K+1)T+1}$:

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

where $\{p_i\}$ satisfy

$$p_i \ge 0, 1 \le i \le nT, \sum_{i=1}^{nT} p_i = 1$$
$$\sum_{i=1}^{nT} p_i b_i e_i = 0$$
$$\sum_{i=1}^{nT} p_i \left\{ \tilde{g}_{ii,1}(e_i^2 - \sigma^2) + 2e_i \sum_{j=1}^{i-1} \tilde{g}_{ij,1} e_j \right\} = 0$$
$$\vdots$$
$$\sum_{i=1}^{nT} p_i \left\{ \tilde{g}_{ii,T}(e_i^2 - \sigma^2) + 2e_i \sum_{j=1}^{i-1} \tilde{g}_{ij,T} e_j \right\} = 0$$
$$\sum_{i=1}^{nT} p_i (e_i^2 - \sigma^2) = 0$$

Let

$$\omega_i(\theta) = \begin{pmatrix} b_i e_i \\ \tilde{g}_{ii,1}(e_i^2 - \sigma^2) + 2e_i \sum_{j=1}^{i-1} \tilde{g}_{ij,1} e_j \\ \vdots \\ \tilde{g}_{ii,T}(e_i^2 - \sigma^2) + 2e_i \sum_{j=1}^{i-1} \tilde{g}_{ij,T} e_j \\ e_i^2 - \sigma^2 \end{pmatrix}_{\{(K+1)T+1\} \times 1}$$

where e_i is the i-th component of $\mu = B(Y - X\beta)$. Following Owen (1990), one can show that:

$$\ell(\theta) \hat{=} -2 \log L(\theta) = 2 \sum_{i=1}^{nT} \log \{1 + \lambda^\tau(\theta)\omega_i(\theta)\} \tag{10}$$

where $\lambda(\theta) \in R^{(K+1)T+1}$ is the solution of the following equation:

$$\frac{1}{nT} \sum_{i=1}^{nT} \frac{\omega_i(\theta)}{1 + \lambda^\tau(\theta)\omega_i(\theta)} = 0 \tag{11}$$

Let $\nu_j = Ee_1^j, j = 3, 4$. 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. Furthermore, Let 1_n stand for the n-dimensional (column) vector with 1 as its components. To obtain the asymptotical distribution of $\ell_n(\theta)$, we need following assumptions:

A1. $\{\mu_{ti}, t = 1, ..., T, i = 1, ..., n\}$, i.e., $\{e_i, i = 1, ..., nT\}$ are independent and identically distributed random variables with mean 0, variance $\sigma^2 > 0$, and $E|e_1|^{4+\eta_1} < \infty$ for some $\eta_1 > 0$.

A2. Let W_n, $\{B_t^{-1}, t = 1, ..., T\}$ and $\{X_t, t = 1, ..., T\}$ be as described above. They satisfy the following conditions:

i. The row and column sums of W_n and $\{B_t^{-1}, t = 1, ..., T\}$ are uniformly bounded in absolute value;
ii. $\{X_t, t = 1, ..., T\}$ are uniformly bounded.

A3. There are constants $c_j > 0, j = 1, 2,$ such that

$$0 < c_1 \le \lambda_{min}\left((nT)^{-1}\Sigma_{(K+1)T+1}\right) \le \lambda_{max}\left((nT)^{-1}\Sigma_{(K+1)T+1}\right) \le c_2 < \infty$$

where $\lambda_{min}(H)$ and $\lambda_{max}(H)$ denote the minimum and maximum eigenvalues of a matrix H, respectively.

$$\begin{aligned}\Sigma_{(K+1)T+1} &= \Sigma^\tau_{(K+1)T+1} = Cov\left\{\sum_{i=1}^{nT}\omega_i(\theta)\right\} \\ &= \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ * & \Sigma_{22} & \Sigma_{23} \\ * & * & \Sigma_{33}\end{pmatrix}_{\{(K+1)T+1\}\times\{(K+1)T+1\}}\end{aligned} \tag{12}$$

where

$$\begin{aligned}&\Sigma_{11} = \sigma^2 X^\tau B^\tau B X, \Sigma_{12} = \nu_3 X^\tau B^\tau\left(Vec(diag\tilde{G}_{n1}), ..., Vec(diag\tilde{G}_{nT})\right) \\ &\Sigma_{13} = \nu_3 X^\tau B^\tau 1_n, \Sigma_{22} = 2\sigma^4 A + (\nu_4 - 3\sigma^4)\tilde{A} \\ &\Sigma_{23} = (\nu_4 - \sigma^4)\left(tr(\tilde{G}_{n1}), ..., tr(\tilde{G}_{nT})\right)^\tau, \Sigma_{33} = nT(\nu_4 - \sigma^4)\end{aligned}$$

with

$$A_{T\times T} = (\alpha_{t_1,t_2}), \alpha_{t_1,t_2} = \sum_{i=1}^{nT}\sum_{j=1}^{nT}\tilde{g}_{ij,t_1}\tilde{g}_{ij,t_2}, \tilde{A}_{T\times T} = (\tilde{\alpha}_{t_1,t_2}), \tilde{\alpha}_{t_1,t_2} = \sum_{i=1}^{nT}\tilde{g}_{ii,t_1}\tilde{g}_{ii,t_2}$$

A4. $n \to \infty$ but T is fixed.

Remark 1. Conditions A1–A3 are common assumptions for SAR models. For example, A1 and A2 are used in Assumptions 1, 4, 5, and 6 in Lee (2004), and the analog of $0 < c_1 \le \lambda_{min}((nT)^{-1}\Sigma_{(K+1)T+1})$ (e.g., $(nT)^{-1}\sigma^2_{\tilde{Q}} \ge c$ for some constant $c > 0$ in Lemma 2 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 $\lambda_{max}((nT)^{-1}\Sigma_{(K+1)T+1}) \le c_2 < \infty$. For the sake of convince, we list this consequence of A1 and A2 as a part of A3.

We now state the main results.

Theorem 1. *Suppose that Assumptions (A1)–(A4) are satisfied. Then under model (1), as* $n \to \infty$,

$$\ell_n(\theta) \xrightarrow{d} \chi^2_{(K+1)T+1}$$

where $\chi^2_{(K+1)T+1}$ is a chi-squared distributed random variable with $(K+1)T+1$ degrees of freedom.

Let $z_\alpha((K+1)T+1)$ satisfy $P(\chi^2_{(K+1)T+1} \le z_\alpha((K+1)T+1)) = \alpha$ 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 follows:

$$\{\theta : \ell_n(\theta) \le z_\alpha((K+1)T+1)\}$$

3. Simulations

Let $\theta = (\beta_1^\tau, ..., \beta_T^\tau, \lambda_1, ..., \lambda_T, \sigma^2)^\tau$. According to Anselin (1988), when the error term $\{\mu_{ti}\}$ is normally distributed, the likelihood ratio (LR): $LR(\theta_0) = 2(L(\hat{\theta}) - L(\theta_0))$ is asymptotically distributed as $\chi^2_{(K+1)T+1}$ under the null hypothesis: $\theta = \theta_0$, where L is the corresponding log-likelihood and $\hat{\theta}$ is the MLE. It follows that the LR-based confidence region for θ with asymptotically correct coverage probability α can be constructed as follows:

$$\{\theta : LR(\theta) \le z_\alpha((K+1)T+1)\}$$

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. This fact implies that the EL method performs better than the LR method theoretically when the population distribution is not normal. Our following simulation results do confirm this conclusion.

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 $LR(\theta_0) \le z_{0.95}((K+1)T+1)$ and $\ell_n(\theta_0) \le z_{0.95}((K+1)T+1)$ respectively in 1000 replications, where θ_0 is the true value of θ.

In the simulations, we used the following two models:

i. Model 1:
$y_t = X_t\beta_t + \epsilon_t, \epsilon_t = \lambda_t W_n \epsilon_t + \mu_t, t = 1, 2,$ where $X_1 = (x_{11}, x_{12}, ..., x_{1n})^\tau, x_{1i} = \frac{i}{n+1}, X_2 = (x_{21}, x_{22}, ..., x_{2n})^\tau, x_{2i} = \frac{i}{n+5}, 1 \le i \le n, (\beta_1, \beta_2) = (2.5, 3.5),$ (λ_1, λ_2) were taken as $(-0.85, -0.75), (-0.15, -0.1), (0.15, 0.1)$ and $(0.85, 0.75)$, respectively, $\mu_t = (\mu_{t1}, \mu_{t2}, ..., \mu_{tn})^\tau$, $t=1$, 2, and $\mu_{ti}'s$ were i.i.d. from $N(0, 1)$, $t(5)$ and $\chi^2_4 - 4$, respectively;

ii. Model 2:
$y_t = X_t\beta_t + \epsilon_t, \epsilon_t = \lambda_t W_n \epsilon_t + \mu_t, t = 1, 2$ with $X_1 = (x_{11}, x_{12}, ..., x_{1n})^\tau, x_{1i} = (\frac{i}{n+1}, 1 + \sin i)^\tau$, $X_2 = (x_{21}, x_{22}, ..., x_{2n})^\tau, x_{2i} = (\frac{i}{n+5}, 2 + \cos i)^\tau, 1 \le i \le n, \beta_1 = (1.5, 1.0)^\tau, \beta_2 = (2, 1.2)^\tau$, (λ_1, λ_2) were taken as $(-0.85, -0.75), (-0.15, -0.1), (0.15, 0.1)$ and $(0.85, 0.75)$, respectively, $\mu_t = (\mu_{t1}, \mu_{t2}, ..., \mu_{tn})^\tau$, $t=1$, 2, and $\mu_{ti}'s$ were i.i.d. from $N(0, 1)$, $t(5)$ and $\chi^2_4 - 4$, respectively.

The results of simulations under Model 1 are reported in Tables 1–3, and the results of simulations under Model 2 are reported in Tables 4–6.

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

(λ_1, λ_2)	W_n	LR	EL	(λ_1, λ_2)	W_n	LR	EL
(−0.85, −0.75)	$grid_{49}$	0.928	0.863	(−0.15, −0.1)	$grid_{49}$	0.930	0.857
	$grid_{100}$	0.941	0.911		$grid_{100}$	0.930	0.893
	$grid_{169}$	0.956	0.942		$grid_{169}$	0.933	0.909
	$grid_{256}$	0.948	0.937		$grid_{256}$	0.962	0.948
	$grid_{400}$	0.950	0.940		$grid_{400}$	0.955	0.944
	W_{49}	0.934	0.858		W_{49}	0.933	0.861
	$I_5 \otimes W_{49}$	0.942	0.928		$I_5 \otimes W_{49}$	0.944	0.934
(0.85, 0.75)	$grid_{49}$	0.910	0.815	(0.15, 0.1)	$grid_{49}$	0.922	0.847
	$grid_{100}$	0.924	0.895		$grid_{100}$	0.938	0.911
	$grid_{169}$	0.930	0.914		$grid_{169}$	0.950	0.926
	$grid_{256}$	0.947	0.929		$grid_{256}$	0.933	0.924
	$grid_{400}$	0.952	0.941		$grid_{400}$	0.945	0.944
	W_{49}	0.944	0.859		W_{49}	0.919	0.836
	$I_5 \otimes W_{49}$	0.961	0.947		$I_5 \otimes W_{49}$	0.951	0.942

For the contiguity weight matrix $W_n = (W_{ij})$, we took $W_{ij} = 1$ if spatial units i and j are neighbors by queen contiguity rule (namely, they share common border or vertex), $W_{ij} = 0$ otherwise (Anselin 1988, 18). Firstly, we considered three ideal cases of spatial units: $n = m \times m$ regular grid with $m = 7, 10, 13, 16, 20$, denoting W_n as $grid_{49}, grid_{100}, grid_{169}, grid_{256}$, and $grid_{400}$, respectively. Secondly, we used the weight matrix W_{49} related to 49 contiguous planning neighborhoods in Columbus, Ohio, United States, which appeared in Anselin (1988, 187). Thirdly, $W_n = I_5 \otimes W_{49}$ was considered, where $\otimes$ is Kronecker product. This corresponds to the pooling of five separate districts with similar neighboring structures in each district.

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 $W_{ij}/\sum_{j=1}^{n} W_{ij}$.

Simulation results under Model 1 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 and n is large, but not well in other cases. The coverage probabilities of the confidence regions based on LR fall to the range [0.800, 0.854] for t distribution and [0.808, 0.864] for χ^2 distribution, which are far from the nominal level 0.95. Simulation results under Model 2 are similar to those under Model 1.

We can see, from Tables 1–6, that 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 are not sure whether the errors are normally distributed.

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

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

Table 2. Coverage probabilities of the LR and EL confidence regions with $\epsilon_i \sim t(5)$ under Model 1.

(λ_1, λ_2)	W_n	LR	EL	(λ_1, λ_2)	W_n	LR	EL
(−0.85, −0.75)	$grid_{49}$	0.838	0.782	(−0.15, −0.1)	$grid_{49}$	0.818	0.763
	$grid_{100}$	0.817	0.832		$grid_{100}$	0.814	0.843
	$grid_{169}$	0.832	0.870		$grid_{169}$	0.800	0.859
	$grid_{256}$	0.816	0.886		$grid_{256}$	0.825	0.883
	$grid_{400}$	0.830	0.901		$grid_{400}$	0.836	0.911
	W_{49}	0.823	0.767		W_{49}	0.832	0.769
	$I_5 \otimes W_{49}$	0.838	0.903		$I_5 \otimes W_{49}$	0.838	0.905
(0.85, 0.75)	$grid_{49}$	0.803	0.730	(0.15, 0.1)	$grid_{49}$	0.818	0.766
	$grid_{100}$	0.809	0.831		$grid_{100}$	0.843	0.867
	$grid_{169}$	0.806	0.850		$grid_{169}$	0.818	0.882
	$grid_{256}$	0.816	0.868		$grid_{256}$	0.833	0.902
	$grid_{400}$	0.854	0.903		$grid_{400}$	0.824	0.909
	W_{49}	0.834	0.771		W_{49}	0.817	0.761
	$I_5 \otimes W_{49}$	0.803	0.869		$I_5 \otimes W_{49}$	0.840	0.890

Table 3. Coverage probabilities of the LR and EL confidence regions with $\epsilon_i + 4 \sim \chi^2_4$ under Model 1.

(λ_1, λ_2)	W_n	LR	EL	(λ_1, λ_2)	W_n	LR	EL
(−0.85, −0.75)	$grid_{49}$	0.833	0.807	(−0.15, −0.1)	$grid_{49}$	0.821	0.797
	$grid_{100}$	0.833	0.861		$grid_{100}$	0.849	0.866
	$grid_{169}$	0.838	0.884		$grid_{169}$	0.833	0.877
	$grid_{256}$	0.850	0.904		$grid_{256}$	0.842	0.901
	$grid_{400}$	0.854	0.914		$grid_{400}$	0.854	0.905
	W_{49}	0.833	0.789		W_{49}	0.837	0.789
	$I_5 \otimes W_{49}$	0.840	0.898		$I_5 \otimes W_{49}$	0.861	0.892
(0.85, 0.75)	$grid_{49}$	0.808	0.732	(0.15, 0.1)	$grid_{49}$	0.822	0.763
	$grid_{100}$	0.829	0.821		$grid_{100}$	0.853	0.873
	$grid_{169}$	0.818	0.862		$grid_{169}$	0.829	0.883
	$grid_{256}$	0.864	0.901		$grid_{256}$	0.839	0.888
	$grid_{400}$	0.864	0.917		$grid_{400}$	0.862	0.922
	W_{49}	0.841	0.768		W_{49}	0.826	0.785
	$I_5 \otimes W_{49}$	0.830	0.891		$I_5 \otimes W_{49}$	0.841	0.891

Table 4. Coverage probabilities of the LR and EL confidence regions with $\epsilon_i \sim N(0, 1)$ under model 2.

(λ_1, λ_2)	W_n	LR	EL	(λ_1, λ_2)	W_n	LR	EL
(−0.85, −0.75)	$grid_{49}$	0.937	0.830	(−0.15, −0.1)	$grid_{49}$	0.927	0.809
	$grid_{100}$	0.932	0.893		$grid_{100}$	0.934	0.895
	$grid_{169}$	0.950	0.937		$grid_{169}$	0.934	0.906
	$grid_{256}$	0.943	0.924		$grid_{256}$	0.942	0.937
	$grid_{400}$	0.967	0.951		$grid_{400}$	0.946	0.939
	W_{49}	0.935	0.832		W_{49}	0.935	0.819
	$I_5 \otimes W_{49}$	0.958	0.935		$I_5 \otimes W_{49}$	0.951	0.933
(0.85, 0.75)	$grid_{49}$	0.870	0.734	(0.15, 0.1)	$grid_{49}$	0.906	0.789
	$grid_{100}$	0.915	0.875		$grid_{100}$	0.940	0.893
	$grid_{169}$	0.924	0.904		$grid_{169}$	0.925	0.905
	$grid_{256}$	0.940	0.919		$grid_{256}$	0.947	0.941
	$grid_{400}$	0.940	0.926		$grid_{400}$	0.938	0.929
	W_{49}	0.919	0.794		W_{49}	0.931	0.797
	$I_5 \otimes W_{49}$	0.947	0.924		$I_5 \otimes W_{49}$	0.937	0.923

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

(λ_1, λ_2)	W_n	LR	EL	(λ_1, λ_2)	W_n	LR	EL
(−0.85, −0.75)	$grid_{49}$	0.840	0.707	(−0.15, −0.1)	$grid_{49}$	0.812	0.695
	$grid_{100}$	0.847	0.855		$grid_{100}$	0.835	0.829
	$grid_{169}$	0.829	0.851		$grid_{169}$	0.830	0.861
	$grid_{256}$	0.827	0.875		$grid_{256}$	0.849	0.883
	$grid_{400}$	0.818	0.900		$grid_{400}$	0.817	0.899
	W_{49}	0.843	0.693		W_{49}	0.827	0.710
	$I_5 \otimes W_{49}$	0.827	0.871		$I_5 \otimes W_{49}$	0.847	0.885
(0.85, 0.75)	$grid_{49}$	0.769	0.635	(0.15, 0.1)	$grid_{49}$	0.803	0.679
	$grid_{100}$	0.792	0.771		$grid_{100}$	0.821	0.809
	$grid_{169}$	0.784	0.813		$grid_{169}$	0.812	0.853
	$grid_{256}$	0.825	0.867		$grid_{256}$	0.816	0.864
	$grid_{400}$	0.836	0.872		$grid_{400}$	0.808	0.875
	W_{49}	0.820	0.660		W_{49}	0.812	0.696
	$I_5 \otimes W_{49}$	0.846	0.887		$I_5 \otimes W_{49}$	0.845	0.896

Table 6. Coverage probabilities of the LR and EL confidence regions with $\epsilon_i + 4 \sim \chi^2_4$ under model 2.

(λ_1, λ_2)	W_n	LR	EL	(λ_1, λ_2)	W_n	LR	EL
(−0.85, −0.75)	$grid_{49}$	0.796	0.723	(−0.15, −0.1)	$grid_{49}$	0.805	0.691
	$grid_{100}$	0.841	0.832		$grid_{100}$	0.839	0.839
	$grid_{169}$	0.837	0.856		$grid_{169}$	0.837	0.862
	$grid_{256}$	0.864	0.904		$grid_{256}$	0.851	0.889
	$grid_{400}$	0.853	0.900		$grid_{400}$	0.857	0.906
	W_{49}	0.812	0.721		W_{49}	0.831	0.735
	$I_5 \otimes W_{49}$	0.852	0.885		$I_5 \otimes W_{49}$	0.860	0.899
(0.85, 0.75)	$grid_{49}$	0.774	0.638	(0.15, 0.1)	$grid_{49}$	0.800	0.679
	$grid_{100}$	0.811	0.807		$grid_{100}$	0.812	0.810
	$grid_{169}$	0.802	0.818		$grid_{169}$	0.832	0.846
	$grid_{256}$	0.854	0.880		$grid_{256}$	0.857	0.889
	$grid_{400}$	0.855	0.898		$grid_{400}$	0.858	0.895
	W_{49}	0.808	0.688		W_{49}	0.801	0.684
	$I_5 \otimes W_{49}$	0.831	0.859		$I_5 \otimes W_{49}$	0.842	0.868

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 2.

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

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

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

$$
\begin{aligned}
\nu_{\tilde{Q}} &= \sum_{i=1}^n a_{nii}\sigma_{ni}^2 \\
\sigma_{\tilde{Q}}^2 &= 2\sum_{i=1}^n\sum_{j=1}^n a_{nij}^2\sigma_{ni}^2\sigma_{nj}^2 + \sum_{i=1}^n b_{ni}^2\sigma_{ni}^2 \\
&\quad + \sum_{i=1}^n \{a_{nii}^2(\mu_{ni}^{(4)} - 3\sigma_{ni}^4) + 2b_{ni}a_{nii}\mu_{ni}^{(3)}\}
\end{aligned}
\tag{13}
$$

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 $n^{-1}\sigma^2_{\tilde{Q}} \geq c$ for some constant $c > 0$. Then

$$\frac{\tilde{Q}_n - \nu_{\tilde{Q}}}{\sigma_{\tilde{Q}}} \xrightarrow{d} N(0,1)$$

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

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

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

Proof. Using Borel–Cantelli lemma and following the proof of (2.3) in Owen (1990), one can prove Lemma 2, where there is no need to assume that $\eta_1, \eta_2, \ldots, \eta_n$ are independent in using Borel–Cantelli lemma.

Lemma 3. *Suppose that assumptions (A1)–(A4) are satisfied. Then as $n \to \infty$,*

$$Z_n = \max_{1\leq i\leq nT} ||\omega_i(\theta)|| = o_p((nT)^{2/(4+\eta_1)}) \ \ a.s. \tag{14}$$

$$\Sigma^{-1/2}_{(K+1)T+1} \sum_{i=1}^{nT} \omega_i(\theta) \xrightarrow{d} N(0, I_{(K+1)T+1}) \tag{15}$$

$$(nT)^{-1} \sum_{i=1}^{nT} \omega_i(\theta)\omega_i^\tau(\theta) = (nT)^{-1}\Sigma_{(K+1)T+1} + o_p(1) \tag{16}$$

$$\sum_{i=1}^{nT} ||\omega_i(\theta)||^3 = O_p(nT^2) \tag{17}$$

Proof. Note that

$$Z_n \leq \max_{1\leq i\leq nT} \Bigg\{ \max_{1\leq i\leq nT} ||b_i e_i||, \ \max_{1\leq i\leq nT} \left| \tilde{g}_{ii,1}(e_i^2 - \sigma^2) + 2e_i \sum_{j=1}^{i-1} \tilde{g}_{ij,1} e_j \right|, \ldots$$
$$\max_{1\leq i\leq nT} \left| \tilde{g}_{ii,T}(e_i^2 - \sigma^2) + 2e_i \sum_{j=1}^{i-1} \tilde{g}_{ij,T} e_j \right|, \ \max_{1\leq i\leq nT} |e_i^2 - \sigma^2| \Bigg\}$$

By Conditions A1 and A2 and Lemma 2, we have:

$$\max_{1\le i\le nT}||b_ie_i|| = \max_{1\le i\le nT}||b_i||o_p((nT)^{1/(4+\eta_1)}) = o_p((nT)^{1/(4+\eta_1)})$$

$$\max_{1\le i\le nT}|\tilde{g}_{ii,t}(e_i^2-\sigma^2)| = \max_{1\le i\le nT}|\tilde{g}_{ii,t}|o_p((nT)^{2/(4+\eta_1)}) = o_p((nT)^{2/(4+\eta_1)})$$

$$\max_{1\le i\le nT}\left|e_i\sum_{j=1}^{i-1}\tilde{g}_{ij,t}e_j\right| \le (\max_{1\le i\le nT}|e_i|)^2$$

$$\max_{1\le i\le nT}\left(\sum_{j=1}^{i-1}|\tilde{g}_{ij,t}|\right) = o_p((nT)^{2/(4+\eta_1)}), 1\le t\le T$$

$$\max_{1\le i\le nT}|e_i^2-\sigma^2| = o_p((nT)^{2/(4+\eta_1)})$$

Thus, $Z_n = o_p((nT)^{2/(4+\eta_1)})$. (14) is proved.

For any given $l = (l_1^\tau, l_2, \ldots, l_{T+1}, l_{T+2})^\tau \in R^{(K+1)T+1}$ with $||l|| = 1$, where $l_1 \in R^{KT}, l_2, \ldots, l_{T+1}, l_{T+2} \in R$. Then,

$$\begin{aligned} l^\tau\omega_i(\theta) &= l_1^\tau b_ie_i + \sum_{t=1}^{T} l_{t+1}\{\tilde{g}_{ii,t}(e_i^2-\sigma^2) + 2e_i\sum_{j=1}^{i-1}\tilde{g}_{ij,t}e_j\} + l_{T+2}(e_i^2-\sigma^2) \\ &= (\sum_{t=1}^{T} l_{t+1}\tilde{g}_{ii,t} + l_{T+2})(e_i^2-\sigma^2) + 2e_i\sum_{j=1}^{i-1}(\sum_{t=1}^{T} l_{t+1}\tilde{g}_{ij,t})e_j + l_1^\tau b_ie_i. \end{aligned}$$

Thus,

$$\begin{aligned} \sum_{i=1}^{nT} l^\tau\omega_i(\theta) &= \sum_{i=1}^{nT}\left(\sum_{t=1}^{T} l_{t+1}\tilde{g}_{ii,t} + l_{T+2}\right)(e_i^2-\sigma^2) \\ &\quad + 2\sum_{i=1}^{nT}\sum_{j=1}^{i-1}\left(\sum_{t=1}^{T} l_{t+1}\tilde{g}_{ij,t}\right)e_ie_j + \sum_{i=1}^{nT} l_1^\tau b_ie_i \end{aligned}$$

Let

$$Q_n = \sum_{i=1}^{nT}\sum_{j=1}^{nT} u_{ij}e_ie_j + \sum_{i=1}^{nT} v_ie_i$$

where

$$u_{ii} = \sum_{t=1}^{T} l_{t+1}\tilde{g}_{ii,t} + l_{T+2}, u_{ij} = \sum_{t=1}^{T} l_{t+1}\tilde{g}_{ij,t}, v_i = l_1^\tau b_i$$

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

$$\sum_{i=1}^{nT}|u_{ij}| \le \sum_{i=1}^{nT}\sum_{t=1}^{T}|l_{t+1}\tilde{g}_{ij,t}| + |l_{T+2}| \le C \tag{18}$$

Further,

$$\begin{aligned}(nT)^{-1}\sum_{i=1}^{nT}|v_i|^3 &= (nT)^{-1}\sum_{i=1}^{nT}|l_1^\tau b_i|^3 \\ &\le C\max_{1\le i\le nT}||x_i||^3\max_{1\le i\le nT}(\sum_{k=1}^{nT}|a_{ik}|)^3 \le C\end{aligned} \tag{19}$$

where a_{ki} is the $(k,\ i)$-element of B. From (18) and (19), it follows that $(nT)^{-1}\sum_{i=1}^{nT}|v_i|^3 \le C$. Therefore, Condition C2 is satisfied.

We now derive the variance of Q_n. Let d_i be the unit vector in the i-th coordinate direction. It can be shown that:

$$\begin{aligned}\sum_{i=1}^{nT}\sum_{j=1}^{nT}u_{ij}^2 &= \sum_{i=1}^{nT}\left\{\left(\sum_{t=1}^{T}l_{t+1}\tilde{g}_{ii,t}+l_{T+2}\right)^2+\sum_{i\ne j}\left(\sum_{t=1}^{T}l_{t+1}\tilde{g}_{ij,t}\right)^2\right\} \\ &= \sum_{i=1}^{nT}\left\{\left(\sum_{t=1}^{T}l_{t+1}\tilde{g}_{ii,t}\right)^2+2\left(\sum_{t=1}^{T}l_{t+1}\tilde{g}_{ii,t}\right)l_{T+2}+l_{T+2}^2+\sum_{i\ne j}\left(\sum_{t=1}^{T}l_{t+1}\tilde{g}_{ij,t}\right)^2\right\} \\ &= 2\sum_{t=1}^{T}l_{t+1}l_{T+2}\left(\sum_{i=1}^{nT}\tilde{g}_{ii,t}\right)+nTl_{T+2}^2+\sum_{i=1}^{nT}\sum_{j=1}^{nT}\sum_{t=1}^{T}l_{t+1}^2\tilde{g}_{ij,t}^2 \\ &\quad+2\sum_{i=1}^{nT}\sum_{j=1}^{nT}\sum_{t_1\ne t_2}l_{t_1+1}l_{t_2+1}\tilde{g}_{ij,t_1}\tilde{g}_{ij,t_2} \\ &= 2\sum_{t=1}^{T}l_{t+1}l_{T+2}tr(\tilde{G}_{nt})+nTl_{T+2}^2+\tilde{l}^\tau A\tilde{l}\end{aligned}$$

where $\tilde{l}=(l_2,\ldots,l_{T+1})^\tau$, and $\sum_{i,j}\tilde{g}_{ij,t_1}\tilde{g}_{ij,t_2}$ is the $(t_1,\ t_2)$-element of A.

$$\begin{aligned}\sum_{i=1}^{nT}u_{ii}^2 &= \sum_{i=1}^{nT}\left(\sum_{t=1}^{T}l_{t+1}\tilde{g}_{ii,t}+l_{T+2}\right)^2 \\ &= \sum_{t=1}^{T}l_{t+1}^2\left(\sum_{i=1}^{nT}\tilde{g}_{ii,t}^2\right)+2\sum_{t_1\ne t_2}l_{t_1+1}l_{t_2+1}\left(\sum_{i=1}^{nT}\tilde{g}_{ii,t_1}\tilde{g}_{ii,t_2}\right) \\ &\quad+2\sum_{t=1}^{T}l_{t+1}l_{T+2}\left(\sum_{i=1}^{nT}\tilde{g}_{ii,t}\right)+nTl_{T+2}^2 \\ &= \tilde{l}^\tau\tilde{A}\tilde{l}+2l_{T+2}\sum_{t=1}^{T}l_{t+1}tr(\tilde{G}_{nt})+nTl_{T+2}^2\end{aligned}$$

where $\sum_{i=1}^{nT}\tilde{g}_{ii,t_1}\tilde{g}_{ii,t_2}$ is the $(t_1,\ t_2)$-element of $\tilde{A}$.

$$\sum_{i=1}^{nT} v_i^2 = \sum_{i=1}^{nT} (l_1^\tau b_i)^2 = l_1^\tau \left(\sum_{i=1}^{nT} b_i b_i^\tau \right) l_1 = l_1^\tau \left(\sum_{i=1}^{nT} X^\tau B^\tau d_i d_i^\tau BX \right) l_1$$
$$= l_1^\tau X^\tau B^\tau \left(\sum_{i=1}^{nT} d_i d_i^\tau \right) BXl_1 = l_1^\tau X^\tau B^\tau BXl_1$$

and that

$$\sum_{i=1}^{nT} u_{ii} v_i = \sum_{i=1}^{nT} \left(\sum_{t=1}^{T} l_{t+1} \tilde{g}_{ii,\,t} + l_{T+2} \right) l_1^\tau b_i$$
$$= \sum_{t=1}^{T} l_{t+1} (\sum_{i=1}^{nT} \tilde{g}_{ii,\,t} l_1^\tau b_i) + l_{T+2} \sum_{i=1}^{nT} (l_1^\tau b_i)$$
$$= \sum_{t=1}^{T} l_{t+1} l_1^\tau X^\tau B^\tau Vec(diag\tilde{G}_{nt}) + l_1^\tau X^\tau B^\tau 1_n l_{T+2}$$

where 1_n is the n-dimensional vector with 1 as its components. It follows from (13) that the variance of Q_n is as follows:

$$\sigma_Q^2 = 2\sum_{i=1}^{nT}\sum_{j=1}^{nT} u_{ij}^2 \sigma^4 + \sum_{i=1}^{nT} v_i^2 \sigma^2 + \sum_{i=1}^{nT} \{u_{ii}^2 (\nu_4 - 3\sigma^4) + 2u_{ii} v_i \nu_3\}$$
$$= 2\sigma^4 \{2\sum_{t=1}^{T} l_{t+1} l_{T+2} tr(\tilde{G}_{nt}) + nTl_{T+2}^2 + \tilde{l}^\tau A\tilde{l}\} + \sigma^2 l_1^\tau X^\tau B^\tau BXl_1$$
$$+ (\nu_4 - 3\sigma^4)\{\tilde{l}^\tau \tilde{A}\tilde{l} + 2l_{T+2} \sum_{t=1}^{T} l_{t+1} tr(\tilde{G}_{nt}) + nTl_{T+2}^2\}$$
$$+ 2\nu_3 \{\sum_{t=1}^{T} l_{t+1} l_1^\tau X^\tau B^\tau Vec(diag\tilde{G}_{nt}) + l_1^\tau X^\tau B^\tau 1_n l_{T+2}\}$$
$$= l^\tau \Sigma_{(K+1)T+1} l$$

where $\Sigma_{(K+1)T+1}$ is given in (12). From Condition A3, one can see that $(nT)^{-1}\sigma_Q^2 \geq c_1 > 0$. From Lemma 1, we have the following:

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

Noting that $Q_n - E(Q_n) = \sum_{i=1}^{nT} l^\tau \omega_i(\theta)$, we thus have (15).

Next we will prove (16), that is,

$$(nT)^{-1} \sum_{i=1}^{nT} (l^\tau \omega_i(\theta))^2 = (nT)^{-1} \sigma_Q^2 + o_p(1) \tag{20}$$

Let

$$M_{in} = l^\tau \omega_i(\theta)$$
$$= u_{ii}(e_i^2 - \sigma^2) + 2\sum_{j=1}^{i-1} u_{ij} e_i e_j + v_i e_i \tag{21}$$
$$= u_{ii}(e_i^2 - \sigma^2) + R_i e_i,$$

where $R_i = 2\sum_{j=1}^{i-1} u_{ij}e_j + v_i$. Let $\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_i = \sigma(e_1, e_2, ..., e_i), 1 \le i \le nT$. Then $\{M_{in}, \mathcal{F}_i, 1 \le i \le nT\}$ form a martingale difference array. Note that

$$\begin{aligned}
&(nT)^{-1}\sum_{i=1}^{nT}\{l^\tau \omega_i(\theta)\}^2 - (nT)^{-1}\sigma_Q^2 = (nT)^{-1}\sum_{i=1}^{nT}(M_{in}^2 - EM_{in}^2) \\
&= (nT)^{-1}\sum_{i=1}^{nT}\{M_{in}^2 - E(M_{in}^2|\mathcal{F}_{i-1}) + E(M_{in}^2|\mathcal{F}_{i-1}) - EM_{in}^2\} \\
&= (nT)^{-1}S_{n1} + (nT)^{-1}S_{n2}
\end{aligned} \tag{22}$$

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

$$(nT)^{-1}S_{n1}=o_p(1) \tag{23}$$

and

$$(nT)^{-1}S_{n2}=o_p(1) \tag{24}$$

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

$$M_{in}^2 = u_{ii}^2(e_i^2 - \sigma^2)^2 + R_i^2e_i^2 + 2u_{ii}R_i(e_i^2 - \sigma^2)e_i$$

Thus

$$E(M_{in}^2|\mathcal{F}_{i-1}) = u_{ii}^2E(e_i^2 - \sigma^2)^2 + R_i^2\sigma^2 + 2u_{ii}R_i\nu_3$$

It follows that:

$$\begin{aligned}
&(nT)^{-2}E(S_{n1}^2) = (nT)^{-2}\sum_{i=1}^{nT}E\{M_{in}^2 - E(M_{in}^2|\mathcal{F}_{i-1})\}^2 \\
&= (nT)^{-2}\sum_{i=1}^{nT}E[u_{ii}^2\{(e_i^2 - \sigma^2)^2 - E(e_i^2 - \sigma^2)^2\} + R_i^2(e_i^2 - \sigma^2) \\
&+ 2u_{ii}R_i(e_i^3 - \sigma^2e_i - \nu_3)]^2 \\
&\le C(nT)^{-2}\sum_{i=1}^{nT}E[u_{ii}^4\{(e_i^2 - \sigma^2)^2 - E(e_i^2 - \sigma^2)^2\}^2] \\
&+ C(nT)^{-2}\sum_{i=1}^{nT}E\{R_i^4(e_i^2 - \sigma^2)^2\} \\
&+ C(nT)^{-2}\sum_{i=1}^{nT}E\{u_{ii}^2R_i^2(e_i^3 - \sigma^2e_i - \nu_3)^2\}
\end{aligned} \tag{25}$$

By Conditions A1 and A2, we have the following:

$$\begin{aligned}
&(nT)^{-2}\sum_{i=1}^{nT}E\left[u_{ii}^4\{(e_i^2 - \sigma^2)^2 - E(e_i^2 - \sigma^2)^2\}^2\right] \\
&\le C(nT)^{-2}\sum_{i=1}^{nT}u_{ii}^4 \le C(nT)^{-2}\sum_{i=1}^{nT}|\sum_{t=1}^{T}l_{t+1}\tilde{g}_{ii,t} + l_{T+2}|^4 \\
&\le C(nT)^{-2}\sum_{i=1}^{nT}\sum_{t=1}^{T}|l_{t+1}\tilde{g}_{ii,t} + l_{T+2}|^4 \le Cn^{-1} \to 0
\end{aligned} \tag{26}$$

and

$$
\begin{aligned}
&(nT)^{-2}\sum_{i=1}^{nT}E\{R_i^4(e_i^2-\sigma^2)^2\}\le C(nT)^{-2}\sum_{i=1}^{nT}E(\sum_{j=1}^{i-1}u_{ij}e_j+v_i)^4\\
&\le C(nT)^{-2}\sum_{i=1}^{nT}E(\sum_{j=1}^{i-1}u_{ij}e_j)^4+C(nT)^{-2}\sum_{i=1}^{nT}v_i^4\\
&\le C(nT)^{-2}\sum_{i=1}^{nT}\sum_{j=1}^{i-1}u_{ij}^4\nu_4+C(nT)^{-2}\sum_{i=1}^{nT}(\sum_{j=1}^{i-1}u_{ij}^2\sigma^2)^2+C(nT)^{-2}\sum_{i=1}^{nT}(l_1^\tau b_i)^4\\
&\le C(nT)^{-2}\sum_{i=1}^{nT}\sum_{j=1}^{i-1}\sum_{t=1}^{T}|l_{t+1}\tilde{g}_{ij,t}|^4+C(nT)^{-2}\sum_{i=1}^{nT}\left(\sum_{j=1}^{i-1}\sum_{t=1}^{T}|l_{t+1}\tilde{g}_{ij,t}|^2\right)^2\\
&+C(nT)^{-2}\sum_{i=1}^{nT}(l_1^\tau b_i)^4\le Cn^{-1}\to 0
\end{aligned}
\tag{27}
$$

Similarly,

$$
(nT)^{-2}\sum_{i=1}^{nT}E\{u_{ii}^2R_i^2(e_i^3-\sigma^2e_i-\nu_3)^2\}\to 0 \tag{28}
$$

From (25)–(28), we have $(nT)^{-2}E(S_{n1}^2)\to 0$. Furthermore,

$$
\begin{aligned}
E(M_{in}^2)&=E\{E(M_{in}^2|\mathcal{F}_{i-1})\}=u_{ii}^2E(e_i^2-\sigma^2)^2+\sigma^2E(R_i^2)+2u_{ii}\nu_3E(R_i)\\
&=u_{ii}^2E(e_i^2-\sigma^2)^2+\sigma^2(4\sum_{j=1}^{i-1}u_{ij}^2\sigma^2+v_i^2)+2u_{ii}\nu_3v_i
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&(nT)^{-2}E(S_{n2}^2)=(nT)^{-2}E[\sum_{i=1}^{nT}\{E(M_{in}^2|\mathcal{F}_{i-1})-EM_{in}^2\}]^2\\
&=(nT)^{-2}E[\sum_{i=1}^{nT}\{R_i^2\sigma^2-\sigma^2(4\sum_{j=1}^{i-1}u_{ij}^2\sigma^2+v_i^2)+2u_{ii}\nu_3(R_i-v_i)\}]^2\\
&=(nT)^{-2}\sum_{i=1}^{nT}E[\sigma^2\{(2\sum_{j=1}^{i-1}u_{ij}e_j)^2-4\sum_{j=1}^{i-1}u_{ij}^2\sigma^2\}+4(\sum_{j=1}^{i-1}u_{ij}e_j)v_i\sigma^2\\
&+2u_{ii}\nu_3(2\sum_{j=1}^{i-1}u_{ij}e_j)]^2\\
&\le C(nT)^{-2}\sum_{i=1}^{nT}E\{\sigma^2(\sum_{j=1}^{i-1}u_{ij}e_j)^2-\sum_{j=1}^{i-1}u_{ij}^2\sigma^2\}^2\\
&+C(nT)^{-2}\sum_{i=1}^{nT}E\{(\sum_{j=1}^{i-1}u_{ij}e_j)v_i\sigma^2\}^2\\
&+C(nT)^{-2}\sum_{i=1}^{nT}E\{2u_{ii}\nu_3(\sum_{j=1}^{i-1}u_{ij}e_j)\}^2
\end{aligned}
\tag{29}
$$

Note that

$$
\begin{aligned}
&(nT)^{-2}\sum_{i=1}^{nT}E\left[\sigma^2\{(\sum_{j=1}^{i-1}u_{ij}e_j)^2-\sum_{j=1}^{i-1}u_{ij}^2\sigma^2\}\right]^2 \\
&\le (nT)^{-2}\sigma^4\sum_{i=1}^{nT}E(\sum_{j=1}^{i-1}u_{ij}e_j)^4 \\
&\le C(nT)^{-2}\sum_{i=1}^{nT}\sum_{j=1}^{i-1}u_{ij}^4\nu_4+C(nT)^{-2}\sum_{i=1}^{nT}(\sum_{j=1}^{i-1}u_{ij}^2\sigma^2)^2 \\
&\le Cn^{-1}\to 0
\end{aligned}
\tag{30}
$$

$$
\begin{aligned}
&(nT)^{-2}\sum_{i=1}^{nT}E\{(\sum_{j=1}^{i-1}u_{ij}e_j)\nu_i\sigma^2\}^2=(nT)^{-2}\sigma^6\sum_{i=1}^{nT}\nu_i^2\sum_{j=1}^{i-1}u_{ij}^2 \\
&\le C(nT)^{-2}\to 0
\end{aligned}
\tag{31}
$$

and

$$
\begin{aligned}
(nT)^{-2}\sum_{i=1}^{nT}E\{2u_{ii}\nu_3(\sum_{j=1}^{i-1}u_{ij}e_j)\}^2&=4\nu_3^2\sigma^2(nT)^{-2}\sum_{i=1}^{nT}u_{ii}^2\sum_{j=1}^{i-1}u_{ij}^2 \\
&\le C(nT)^{-1}\to 0
\end{aligned}
\tag{32}
$$

where we have used Conditions A1 and A2. From (29)–(32), we have $(nT)^{-2}ES_{n2}^2\to 0$. The proof of (20) is thus complete.

Finally, we will prove (17). Note that

$$
\begin{aligned}
\sum_{i=1}^{nT}E||\omega_i(\theta)||^3\le\sum_{i=1}^{nT}E||b_ie_i||^3&+\sum_{t=1}^{T}\sum_{i=1}^{nT}E|\tilde{g}_{ii,t}(e_i^2-\sigma^2)+2e_i\sum_{j=1}^{i-1}\tilde{g}_{ij,t}e_j|^3 \\
&+\sum_{i=1}^{nT}E|e_i^2-\sigma^2|^3.
\end{aligned}
\tag{33}
$$

By Conditions A1 and A2,

$$
\sum_{i=1}^{nT}E||b_ie_i||^3\le CnT(\max_{1\le i\le nT}||x_i||)^3E|e_1|^3=O(nT)
\tag{34}
$$

$$
\begin{aligned}
&\sum_{i=1}^{nT}E\left|\tilde{g}_{ii,t}(e_i^2-\sigma^2)+2e_i\sum_{j=1}^{i-1}\tilde{g}_{ij,t}e_j\right|^3 \\
&\le C\sum_{i=1}^{nT}E|\tilde{g}_{ii,t}(e_i^2-\sigma^2)|^3+C\sum_{i=1}^{nT}E\left|2e_i\sum_{j=1}^{i-1}\tilde{g}_{ij,t}e_j\right|^3 \\
&\le C\sum_{i=1}^{nT}E|\tilde{g}_{ii,t}(e_i^2-\sigma^2)|^3+C\sum_{i=1}^{nT}E|e_i|^3\sum_{j=1}^{i-1}E|\tilde{g}_{ij,t}e_j|^3 \\
&+C\sum_{i=1}^{nT}E|e_i|^3\left\{\sum_{j=1}^{i-1}E(\tilde{g}_{ij,t}e_j)^2\right\}^{3/2}=O(nT)
\end{aligned}
\tag{35}
$$

$$\sum_{i=1}^{nT} E|e_i^2 - \sigma^2|^3 = O(nT) \tag{36}$$

From (33)–(36), we have

$$\sum_{i=1}^{nT} E||\omega_i(\theta)||^3 = O(nT) \tag{37}$$

Further, using (37) and Markov inequality, we obtain $\sum_{i=1}^{nT} ||\omega_i(\theta)||^3 = O_p(nT^2)$. Thus (17) is proved.

Proof of Theorem 1. Using Lemma 3 and following the proof of Theorem 1 in Qin (2019), one can easily show that Theorem 1 holds true.

Funding

This work was partially supported by the National Natural Science Foundation of China (11671102, 11731015, 61662007), 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*. The Netherlands: Kluwer Academic Publishers.

Anselin, L., and A. K. Bera. 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In *Handbook of applied economics statistics*, ed. A. Ullah and D. E. A. Giles, 237–89. New York: Marcel Dekker.

Anselin, L. 2001. Spatial econometrics. In *A companion to theoretical econometrics*, ed. B.H. Baltagi, 310–30. Massachusetts: Blackwell Publishers Ltd..

Anselin, L., J. Le Gallo, and J. Jayet. 2008. Spatial panel econometrics. In *The econometrics of panel data: Fundamentals and recent developments in theory and practice*, ed. L. Mátyás and P. Sevestre, 625–60. Berlin Heidelberg: Springer-Verlag.

Baltagi, B. H., S. H. Song, and W. Koh. 2003. Testing panel data regression models with spatial error correlation. *Journal of Econometrics* 117 (1):123–50. doi:10.1016/S0304-4076(03)00120-9.

Baltagi, B. H., and A. Pirotte. 2011. Seemingly unrelated regressions with spatial error components. *Empirical Economics* 40 (1):5–49. doi:10.1007/s00181-010-0373-8.

Chen, J., and J. Qin. 1993. Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika* 80 (1):107–16. doi:10.1093/biomet/80.1.107.

Chen, S. X., and I. V. Keilegom. 2009. A review on empirical likelihood for regressions (with discussions). *Test* 18 (3):415–47. doi:10.1007/s11749-009-0159-5.

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

Dow, M. M., M. L. Burton, and D. R. White. 1982. Network autocorrelation: A simulation study of a foundational problem in regression and survey research. *Social Networks* 4 (2):169–200. doi:10.1016/0378-8733(82)90031-4.

Elhorst, J. P. 2005. Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geographical Analysis* 37 (1):85–106. doi:10.1111/j.1538-4632.2005.00577.x.

Elhorst, J. P. 2010. Dynamic panels with endogenous interaction effects when T is small. *Regional Science and Urban Economics* 40 (5):272–82. doi:10.1016/j.regsciurbeco.2010.03.003.
Hou, J., and Y. Zhao. 2019. Some remarks on a pair of seemingly unrelated regression models. *Open Mathematics* 17 (1):979–89. doi:10.1515/math-2019-0077.
Jiang, H., J. Qian, and Y. Sun. 2020. Best linear unbiased predictors and estimators under a pair of constrained seemingly unrelated regression models. *Statistics and Probability Letters* 158: 108669. doi:10.1016/j.spl.2019.108669.
Kapoor, M., H. H. Kelejian, and I. Prucha. 2007. Panel data models with spatially correlated error components. *Journal of Econometrics* 140 (1):97–130. doi:10.1016/j.jeconom.2006.09.004.
Kelejian, H. H., and I. R. Prucha. 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40 (2):509–33. doi:10.1111/1468-2354.00027.
Kelejian, H. H., and I. R. Prucha. 2001. On the asymptotic distribution of the Moran *I* test statistic with applications. *Journal of Econometrics* 104 (2):219–57. doi:10.1016/S0304-4076(01)00064-1.
Krämer, W., and C. Donninger. 1987. Spatial autocorrelation among errors and the relative efficiency of OLS in the linear regression model. *Journal of the American Statistical Association* 82 (398):577–9. doi:10.2307/2289467.
Kubáček, L. 2013. Seemingly unrelated regression models. *Applications of Mathematics* 58 (1): 111–23. doi:10.1007/s10492-013-0005-7.
Kurata, L., and S. Matsuura. 2016. Best equivariant estimator of regression coefficients in a seemingly unrelated regression model with known correlation matrix. *Annals of the Institute of Statistical Mathematics* 68 (4):705–23. doi:10.1007/s10463-015-0512-2.
Lee, L. F. 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72 (6):1899–925. doi:10.1111/j.1468-0262.2004.00558.x.
Lee, L. F., and J. Yu. 2010a. Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154 (2):165–85. doi:10.1016/j.jeconom.2009.08.001.
Lee, L. F., and J. Yu. 2010b. Some recent developments in spatial panel data models. *Regional Science and Urban Economics* 40 (5):255–71. doi:10.1016/j.regsciurbeco.2009.09.002.
Lee, L. F., and J. Yu. 2013. Spatial panel data models. In *Oxford handbook of panel data*, ed. B. Baltagi, 363–401. New York: Oxford University Press.
Mur, J., F. Lòpez, and M. Herrera. 2010. Testing for spatial effects in seemingly unrelated regressions. *Spatial Economic Analysis* 5 (4):399–440. doi:10.1080/17421772.2010.516443.
Ord, K. 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70 (349):120–6. doi:10.1080/01621459.1975.10480272.
Owen, A. B. 1988. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75 (2):237–49. doi:10.1093/biomet/75.2.237.
Owen, A. B. 1990. Empirical likelihood ratio confidence regions. *The Annals of Statistics* 18 (1): 90–120. doi:10.1214/aos/1176347494.
Owen, A. B. 1991. Empirical likelihood for linear models. *The Annals of Statistics* 19 (4):1725–47. doi:10.1214/aos/1176348368.
Owen, A. B. 2001. *Empirical likelihood.* London: Chapman & Hall.
Qin, J., and J. Lawless. 1994. Empirical likelihood and general estimating equations. *The Annals of Statistics* 22 (1):300–25. doi:10.1214/aos/1176325370.
Qin, Y. 2019. Empirical likelihood for spatial autoregressive models with spatial autoregressive disturbances. *Sankhyā A*: *The Indian Journal of Statistics*. Advance online publication. doi:10.1007/s13171-019-00166-3.
Qu, X., L. F. Lee, and J. H. Yu. 2017. QML estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices. *Journal of Econometrics* 197 (2):173–201. doi:10.1016/j.jeconom.2016.11.004.
Su, L. J., and Z. L. Yang. 2015. QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics* 185 (1):230–58. doi:10.1016/j.jeconom.2014.11.002.
Sun, Y., R. Ke, and Y. Tian. 2014. Some overall properties of seemingly unrelated regression models. *AStA Advances in Statistical Analysis* 98:103–20.

Tian, Y. 2010. Estimations of parametric functions under a system of linear regression equations with correlated errors. *Acta Mathematica Sinica, English Series* 26 (10):1927–42. doi:10.1007/s10114-010-8434-7.

Wang, X., and K. Kockelman. 2007. Specification and estimation of a spatially and temporally autocorrelated SUR model: Application to crash rates in China. *Transportation* 34 (3):281–300. doi:10.1007/s11116-007-9117-9.

Wu, C. B. 2004. Weighted empirical likelihood inference. *Statistics & Probability Letters* 66 (1): 67–79. doi:10.1016/j.spl.2003.10.007.

Yu, J. H., R. De Jong, and L. F. Lee. 2008. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* 146 (1):118–34. doi:10.1016/j.jeconom.2008.08.002.

Zellner, A. 1962. An efficient method of estimating seemingly unrelated regressions and test of aggregation bias. *Journal of the American Statistical Association* 57 (298):348–68. doi:10.1080/01621459.1962.10480664.

Zhao, L., and X. Xu. 2017. Generalized canonical correlation variables improved estimation in high dimensional seemingly unrelated regression models. *Statistics & Probability Letters* 126: 119–26. doi:10.1016/j.spl.2017.02.037.

Zhong, B., and J. N. K. Rao. 2000. Empirical likelihood inference under stratified random sampling using auxiliary population information. *Biometrika* 87 (4):929–38. doi:10.1093/biomet/87.4.929.