



# Empirical likelihood for spatial dynamic panel data models

Yinghua Li<sup>1</sup> · Yongsong Qin<sup>1</sup>

Received: 29 September 2020 / Accepted: 13 September 2021 / Published online: 29 September 2021  
© Korean Statistical Society 2021

## Abstract

Spatial dynamic panel data (SDPD) models have received great attention in economics in recent 10 years. Existing approaches for the estimation and test of SDPD models are quasi-maximum likelihood (QML) approach and generalized method of moments (GMM). In this article, we introduce the empirical likelihood (EL) method to the statistical inference for SDPD models. The EL ratio statistics are constructed for the parameters of spatial dynamic panel data 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. Simulation results show that the EL based confidence regions outperform the normal approximation based confidence regions.

**Keywords** Spatial dynamic panel data model · Spatial error · Empirical likelihood · Confidence region

**Mathematics subject classification** Primary 62G05 · secondary 62E20

## 1 Introduction

Real data are often observed at different locations and times, which are called as spatial panel data (SPD). Examples are economic growth rates of major cities in China over last 40 years, monthly unemployment rates of states in USA in the last decade and daily infection rates of COVID-19 in major cities in Hubei province in China over 3 months since December 31, 2019. These data may be modelled by SPD models. The research to various SPD models can be found in Anselin (1988), Elhorst (2003), Baltagi et al. (2003), Baltagi and Li (2006), Chen and Conley (2001), Pesaran (2004), Kapoor et al. (2007), Baltagi et al. (2007), Lee and Yu (2010a), Mutl and Pfaffermayr (2011), Parent and LeSage (2011) and

---

✉ Yongsong Qin  
ysqin@gxnu.edu.cn

<sup>1</sup> College of Mathematics and Statistics, Guangxi Normal University, Guilin 541004, Guangxi, China

Baltagi et al. (2013), among others. By adding a dynamic element into a SPD model Anselin (2001) proposes a spatial dynamic panel data (SDPD) model, which increases the flexibility of a SPD model. Obviously, SPD models belong to SDPD models. There has been a growing interest in the statistical inferences for SDPD models since then. For an overview on the SDPD models, refer to Yu et al. (2008) and Lee and Yu (2010b), Su and Yang (2015), Lee and Yu (2015a, b), Yu and Lee (2010), Elhorst (2010), Elhorst (2005), Yang et al. (2006), Mutl (2006), Su and Yang (2007), and Lee and Yu (2010c), among others. There are two popular methods for the estimation and test of SPD and SDPD models: quasi-maximum likelihood (QML) approach and generalized method of moments (GMM), which can be seen in above references.

In this article, we use the empirical likelihood (EL) method, proposed by Owen (1988, 1990), to construct the confidence regions for the parameters in a SDPD model. It is observed, in the case of independent observations, that the EL method to construct confidence intervals/regions has many advantages over its counterparts like the normal-approximation-based method and the bootstrap method (e.g., Hall and La Scala 1990; Hall 1992). A excellent review on EL for regressions can be found in Chen and Keilegom (2009). There are a lot of references on EL methods for independent samples or in the context of sample surveys. To save space, we list a few of them such as Owen (2001), Qin and Lawless (1994), Chen and Qin (1993), Zhong and Rao (2000) and Wu (2004). The data are dependent and follow certain structures in SDPD models. The study of the EL method for some SDPD models enjoys a certain progress. For example, there are a few articles studying the EL method for the pure spatial data (PSD, the special case of SPD with a fixed time). For instance, Nordman (2008a, b) and Bandyopadhyay et al. (2015) use the blockwise EL (BEL) proposed by Kitamura (1997) to PSD. Recently, by exploring inherent martingale structures, Qin (2021) and Jin and Lee (2019) use the EL method to construct confidence intervals/regions in PSD models. Further, Li and Qin (2020) extends the EL method proposed by Qin (2021) and Jin and Lee (2019) to SPD models. We note that there is no research work on the EL method for SDPD models.

There are many kinds of SDPD models. Lee and Yu (2010c) gives a review on the classification and research development of some SDPD models. Su and Yang (2015) introduces the QML method to SDPD models with spatial errors, where different types of space-specific effects and different ways that initial observations being generated (exogenously or endogenously) are investigated. Since SDPD models are quite complicated, as a starting point, in this article, we study the EL method for the SDPD models in Su and Yang (2015) with the restriction that there is no space-specific effects (or named as zero drift) and initial observations are generated exogenously. The study of the EL method for general SDPD models without above restriction is left for our future study. Our research results show that the EL based confidence regions generally outperform the normal approximation (NA) based confidence regions when the space units are large enough.

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

## 2 Main results

In this article, we suppose that there are  $n$  individual units and  $T$  time periods and the sampling data satisfy the following SDPD model with spatial error:

$$y_t = \rho y_{t-1} + x_t \beta + z \gamma + \epsilon_t, \quad (1)$$

$$\epsilon_t = \lambda W_n \epsilon_t + v_t, t = 1, 2, \dots, T, \quad (2)$$

where  $y_t = (y_{1t}, \dots, y_{nt})'$  is an  $n$ -dimensional column vector of observed dependent variables,  $\rho$  ( $|\rho| < 1$ ) characterizes the dynamic effect,  $x_t = (x_{1t}, \dots, x_{nt})'$  is an  $n \times p$  matrix of time-varying exogenous variables,  $z = (z_1, \dots, z_n)'$  is an  $n \times q$  matrix of time-invariant exogenous variables, and  $\beta$  and  $\gamma$  are  $p \times 1$  and  $q \times 1$  regression coefficients, respectively. The disturbance vector  $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{nt})'$  is an  $n \times 1$  vector of errors. The parameter  $\lambda$  is a spatial autoregressive coefficient and  $W_n$  is an  $n \times n$  spatial weighting matrix of constants,  $v_t = (v_{1t}, \dots, v_{nt})'$  is an  $n \times 1$  column vector, and  $\{v_{it}\}$  are i.i.d. across  $t$  and  $i$  with zero mean and variance  $\sigma_v^2$ . The spatial weighting matrix is also called contiguity matrix, which is determined by the spatial dependence of  $n$  spatial units. There are many ways to define  $W_n$  (e.g. pages 17–19 in Anselin 1988). Let  $W_{ij}$  be the  $(i, j)$  element of  $W_n$ . Commonly used  $W_n$  includes Rook contiguity, Bishop contiguity and Queen contiguity as follows. Rook contiguity: define  $W_{ij} = 1$  if the units  $i$  and  $j$  share a common side and  $W_{ij} = 0$ , otherwise. Bishop contiguity: define  $W_{ij} = 1$  if the units  $i$  and  $j$  share a common vertex and  $W_{ij} = 0$ , otherwise. Queen contiguity: define  $W_{ij} = 1$  if the units  $i$  and  $j$  share a common side or vertex and  $W_{ij} = 0$ , otherwise. The choice of  $W_n$  is important. Our results hold true for all these commonly used  $W_n$ .

The models (2.1)–(2.3) in Su and Yang (2015) are as follows:

$$y_t = \rho y_{t-1} + x_t \beta + z \gamma + \mu + \epsilon_t, \\ \epsilon_t = \lambda W_n \epsilon_t + v_t, t = 1, 2, \dots, T,$$

where  $\mu = (\mu_1, \dots, \mu_n)'$  represent the unobservable individual or space-specific effects and other notations are the same as in model (1) and (2). Compared to models (2.1)–(2.3) in Su and Yang (2015), the only difference is that there is no space-specific effects  $\mu$  in model (1) and (2). Our initial investigation shows that the EL method for SDPD models with space-specific effects may need an adjusted EL method. Further research is needed and left for our future work.

We develop the EL method for the SDPD model when  $y_0$  is exogenous. In this case, we can treat  $y_0$  as a fixed constant vector as it contains no information about the model parameters. For convenience, we use  $\mathbf{1}_k$  to denote a  $k \times 1$  vector of ones,  $\mathbf{0}_k$  to denote a  $k \times 1$  vector of zeros, and  $J_k = \mathbf{1}_k \mathbf{1}'_k$ , where  $\otimes$  is the Kronecker product.

Let  $Y = (y'_1, y'_2, \dots, y'_T)'$ ,  $Y_{-1} = (y'_0, y'_1, \dots, y'_{T-1})'$ ,  $X = (x'_1, x'_2, \dots, x'_T)'$ ,  $v = (v'_1, v'_2, \dots, v'_T)'$ ,  $Z = \mathbf{1}_T \otimes z$ ,  $B = B(\lambda) = I_n - \lambda W_n$  and  $\epsilon = (\epsilon'_1, \epsilon'_2, \dots, \epsilon'_T)'$ . Then model (1) and (2) can be written in a matrix form as:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \rho \begin{pmatrix} y_0 \\ y_1 \\ \vdots \\ y_{T-1} \end{pmatrix} + \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{pmatrix} \beta + \begin{pmatrix} z \\ \vdots \\ z \end{pmatrix} \gamma + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix},$$

with

$$\begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_T \end{pmatrix} = \begin{pmatrix} B^{-1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & B^{-1} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & B^{-1} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_T \end{pmatrix},$$

or

$$Y = \rho Y_{-1} + X\beta + Z\gamma + \epsilon, \tag{3}$$

with

$$\epsilon = (I_T \otimes B^{-1})v, \tag{4}$$

where  $\epsilon \sim (0, \sigma_v^2 \Omega)$ , with

$$\Omega = \Omega(\lambda) = I_T \otimes (B' B)^{-1}. \tag{5}$$

Let  $\theta = (\beta', \gamma', \rho)'$  and  $\psi = (\theta', \sigma_v^2, \lambda)'$ . We adopt the QML method to derive the estimating equations for the EL method. Under the assumption of normality (which is only used at this moment), based on (3) and (4), the log-likelihood function (ignoring constants) is

$$\tilde{L}(\psi) = -\frac{nT}{2} \log \sigma_v^2 - \frac{1}{2} \log |\Omega| - \frac{1}{2\sigma_v^2} \epsilon' \Omega^{-1} \epsilon, \tag{6}$$

where  $\epsilon = Y - \rho Y_{-1} - X\beta - Z\gamma$ . It can be shown that

$$\begin{aligned} \partial \tilde{L}(\psi) / \partial \theta &= \sigma_v^{-2} \tilde{X}' \Omega^{-1} \epsilon, \\ \partial \tilde{L}(\psi) / \partial \sigma_v^2 &= -\frac{nT}{2\sigma_v^2} + \frac{1}{2\sigma_v^4} \epsilon' \Omega^{-1} \epsilon, \\ \partial \tilde{L}(\psi) / \partial \lambda &= -\frac{1}{2} \text{tr}(\Omega^{-1} (I_T \otimes A)) + \frac{1}{2\sigma_v^2} \epsilon' \Omega^{-1} (I_T \otimes A) \Omega^{-1} \epsilon, \end{aligned}$$

where  $\tilde{X} = (X, Z, Y_{-1})$ ,  $A = (B' B)^{-1} (W_n' B + B' W_n) (B' B)^{-1}$ . Letting above derivatives be 0, we obtain the following estimating equations of the QML method:

$$\tilde{X}' \Omega^{-1} \epsilon = 0, \tag{7}$$

$$-nT \sigma_v^2 + \epsilon' \Omega^{-1} \epsilon = 0, \tag{8}$$

$$-\sigma_v^2 \text{tr}(\Omega^{-1}(I_T \otimes A)) + e' \Omega^{-1}(I_T \otimes A) \Omega^{-1} e = 0, \tag{9}$$

Substituting (4) into (7)–(9), we have

$$\tilde{X}'(I_T \otimes B')v = 0, \tag{10}$$

$$-nT\sigma_v^2 + v'v = 0, \tag{11}$$

$$-\sigma_v^2 \text{tr}(I_T \otimes (BAB')) + v'(I_T \otimes (BAB'))v = 0. \tag{12}$$

Noting that  $\tilde{X} = (X, Z, Y_{-1})$  and  $Y_{-1}$  contains  $X, z$  and  $v$ , we need to separate out  $v$  from  $\tilde{X}$ . To this end, denote  $l_\rho = (0, c_{\rho,1}, \dots, c_{\rho,T-1})'$ ,  $c_{\rho,t} = (1 - \rho^t)/(1 - \rho)$ ,  $Y_0 = (Y'_{0,0}, Y'_{0,1}, \dots, Y'_{0,T-1})'$ ,  $Y_{0,t} = \rho^t y_0$ .

$$F_\rho = \begin{pmatrix} 0 & 1 & \rho & \dots & \rho^{T-2} \\ 0 & 0 & 1 & \dots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$

$A_x = F'_\rho \otimes I_n$  and  $A_v = F'_\rho \otimes B^{-1}$ . We use (B.3) in Su and Yang (2015) to obtain that

$$Y_{-1} = A_x X \beta + (l_\rho \otimes I_n) z \gamma + A_v v + Y_0,$$

Let  $\tilde{X}_1 = (X, Z)$  and  $\tilde{X}_2 = A_x X \beta + (l_\rho \otimes I_n) z \gamma + Y_0$ . Then (10) can be decomposed into

$$\tilde{X}'_1(I_T \otimes B')v = 0, \tag{13}$$

$$\tilde{X}'_2(I_T \otimes B')v + v'A'_v(I_T \otimes B')v = 0. \tag{14}$$

For convenience, let  $e = v$ , i.e.

$$e_{(nT) \times 1} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_{nT} \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_T \end{pmatrix}. \tag{15}$$

Then (11)–(14) can be rewritten as

$$\tilde{X}'_1(I_T \otimes B')e = 0, \tag{16}$$

$$\tilde{X}'_2(I_T \otimes B')e + e'A'_v(I_T \otimes B')e = 0, \tag{17}$$

$$-nT\sigma_v^2 + e'e = 0, \tag{18}$$

$$-\sigma_v^2 \text{tr}(I_T \otimes (BAB')) + e'(I_T \otimes (BAB'))e = 0. \tag{19}$$

Observing that the above estimating equations include the quadratic forms of  $e$ , to use the EL method, we need to change the quadratic forms into the linear forms of a well behaved random variables. To this end, we let  $H_1 = \frac{1}{2}(A'_v(I_T \otimes B') + (I_T \otimes B)A_v)$  and  $H_2 = I_T \otimes (BAB')$ . Use  $h_{ij,k}$ ,  $a_{i,1}$  and  $a_{i,2}$  to denote the  $(i, j)$  element of the matrix  $H_k$  ( $k = 1, 2$ ), the  $i$ -th column of the matrix  $\tilde{X}'_1(I_T \otimes B')$  and the  $i$ -th element of the vector  $\tilde{X}'_2(I_T \otimes B')$ , 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 (17) and (19), we follow Kelejian and Prucha (2001) to introduce a martingale difference array. Define the  $\sigma$ -fields:  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ ,  $\mathcal{F}_i = \sigma(e_1, e_2, \dots, e_i)$ ,  $1 \leq i \leq nT$ . Let

$$M_{ik} = h_{ii,k}(e_i^2 - \sigma_v^2) + 2e_i \sum_{j=1}^{i-1} h_{ij,k}e_j, k = 1, 2. \tag{20}$$

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

$$e' H_k e - \sigma_v^2 \text{tr}(H_k) = \sum_{i=1}^{nT} M_{ik}, k = 1, 2. \tag{21}$$

Based on (16)–(21), we propose the following EL ratio statistic for  $\psi \in R^{p+q+3}$ :

$$L(\psi) = \sup_{p_i, 1 \leq i \leq nT} \prod_{i=1}^{nT} ((nT)p_i),$$

where  $\{p_i\}$  satisfy

$$\begin{aligned} p_i &\geq 0, 1 \leq i \leq nT, \sum_{i=1}^{nT} p_i = 1, \\ \sum_{i=1}^{nT} p_i a_{i,1} e_i &= 0, \\ \sum_{i=1}^{nT} p_i \left\{ a_{i,2} e_i + h_{ii,1}(e_i^2 - \sigma_v^2) + 2e_i \sum_{j=1}^{i-1} h_{ij,1} e_j \right\} &= 0, \\ \sum_{i=1}^{nT} p_i (e_i^2 - \sigma_v^2) &= 0, \\ \sum_{i=1}^{nT} p_i \left\{ h_{ii,2}(e_i^2 - \sigma_v^2) + 2e_i \sum_{j=1}^{i-1} h_{ij,2} e_j \right\} &= 0. \end{aligned}$$

Let

$$\omega_i(\psi) = \begin{pmatrix} a_{i,1}e_i \\ e_i^2 - \sigma_v^2 \\ a_{i,2}e_i + h_{ii,1}(e_i^2 - \sigma_v^2) + 2e_i \sum_{j=1}^{i-1} h_{ij,1}e_j \\ h_{ii,2}(e_i^2 - \sigma_v^2) + 2e_i \sum_{j=1}^{i-1} h_{ij,2}e_j \end{pmatrix}_{(p+q+3) \times 1},$$

where  $e_i$  is the  $i$ th component of  $(I_T \otimes B)(Y - \rho Y_{-1} - X\beta - Z\gamma)$ . Following Owen (1990), one can show that

$$\ell(\psi) \hat{=} -2 \log L(\psi) = 2 \sum_{i=1}^{nT} \log \{1 + \tilde{\lambda}'(\psi)\omega_i(\psi)\}, \tag{22}$$

where  $\tilde{\lambda}(\psi) \in R^{p+q+3}$  is the solution of the following equation:

$$\frac{1}{nT} \sum_{i=1}^{nT} \frac{\omega_i(\psi)}{1 + \tilde{\lambda}'(\psi)\omega_i(\psi)} = 0. \tag{23}$$

Let  $\vartheta_j = Ev_{11}^j, j = 3, 4$ . Use  $Vec(diagA)$  to denote the vector formed by the diagonal elements of a matrix  $A$ ,  $\|a\|$  to denote the  $L_2$ -norm of a vector  $a$ , and  $\lambda_{min}(H)$  and  $\lambda_{max}(H)$  to denote the minimum and maximum eigenvalues of a matrix  $H$ , respectively. To obtain the asymptotic distribution of  $\ell(\psi)$ , we need following assumptions.

- A1. (1)  $v_{jt}$  are mutually independent, and they are independent of  $x_{ks}$  and  $z_k$  for all  $j, k, t, s$ ;
- (2) All elements in  $(x_{it}, z_i)$  have  $4 + \eta_1$  moments for some  $\eta_1 > 0$ .
- A2. (1)  $\{v_{it}, t = 1, \dots, T, i = 1, \dots, n\}$  are independent and identically distributed for all  $i$  and  $t$  with mean 0, variance  $\sigma_v^2 > 0$  and  $E|v_{it}|^{4+\eta_1} < \infty$  for some  $\eta_1 > 0$ .
- (2)  $\{x_{it}, t = \dots, -1, 0, 1, \dots\}$  and  $\{z_i\}$  are strictly exogenous and independent across  $i$ .
- (3)  $|\rho| < 1$ .
- A3. Let  $W_n$  and  $\{B^{-1}\}$  be as described above. They satisfy the following conditions:
  - (1) The row and column sums of  $W_n$  are uniformly bounded in absolute value;
  - (2)  $\{B^{-1}\}$  are uniformly bounded in either row or column sums, uniformly in  $\lambda$  in a compact parameter space  $\Lambda$ , and  $\underline{c}_\lambda \leq \inf_{\lambda \in \Lambda} \lambda_{max}(B'B) \leq \sup_{\lambda \in \Lambda} \lambda_{max}(B'B) \leq \bar{c}_\lambda < \infty$ .
- A4. There are constants  $c_j > 0, j = 1, 2$ , such that

$$0 < c_1 \leq \lambda_{min}((nT)^{-1}\Sigma_{p+q+3}) \leq \lambda_{max}((nT)^{-1}\Sigma_{p+q+3}) \leq c_2 < \infty,$$

where

$$\begin{aligned} \Sigma_{p+q+3} &= \Sigma'_{p+q+3} = \text{Cov} \left\{ \sum_{i=1}^{nT} \omega_i(\psi) \right\} \\ &= \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} \\ * & \Sigma_{22} & \Sigma_{23} & \Sigma_{24} \\ * & * & \Sigma_{33} & \Sigma_{34} \\ * & * & * & \Sigma_{44} \end{pmatrix}_{(p+q+3) \times (p+q+3)}, \end{aligned} \tag{24}$$

where

$$\begin{aligned} \Sigma_{11} &= \sigma_v^2 E(\tilde{X}'_1 \Omega^{-1} \tilde{X}_1), \Sigma_{12} = \vartheta_3 E(\tilde{X}'_1)(I_T \otimes B') \mathbf{1}_{nT}, \\ \Sigma_{13} &= \sigma_v^2 E(\tilde{X}'_1 \Omega^{-1} \tilde{X}_2), \Sigma_{14} = \vartheta_3 E(\tilde{X}'_1)(I_T \otimes B') \text{vec}_D(H_2) \\ \Sigma_{22} &= nT(\vartheta_4 - \sigma_v^4), \Sigma_{23} = (\vartheta_4 - 3\sigma_v^4) \mathbf{1}'_{nT} \text{vec}_D(H_1) + 2\sigma_v^4 \text{tr}(H_1) + \vartheta_3 E(\tilde{X}'_2)(I_T \otimes B') \mathbf{1}_{nT}, \\ \Sigma_{24} &= (\vartheta_4 - 3\sigma_v^4) \mathbf{1}'_{nT} \text{vec}_D(H_2) + 2\sigma_v^4 \text{tr}(H_2), \\ \Sigma_{33} &= (\vartheta_4 - 3\sigma_v^4) \|\text{vec}_D(H_1)\|^2 + 2\sigma_v^4 \text{tr}(H_1^2) + \sigma_v^2 E(\tilde{X}'_2 \Omega^{-1} \tilde{X}_2) + 2\vartheta_3 E(\tilde{X}'_2)(I_T \otimes B') \text{vec}_D(H_1), \\ \Sigma_{34} &= (\vartheta_4 - 3\sigma_v^4) \text{vec}'_D(H_1) \text{vec}_D(H_2) + 2\sigma_v^4 \text{tr}(H_1 H_2) + \vartheta_3 E(\tilde{X}'_2)(I_T \otimes B') \text{vec}_D(H_2), \\ \Sigma_{44} &= (\vartheta_4 - 3\sigma_v^4) \|\text{vec}_D(H_2)\|^2 + 2\sigma_v^4 \text{tr}(H_2^2). \end{aligned}$$

A5.  $n \rightarrow \infty$  but  $T$  is fixed.

**Remark 1** Conditions A1–A3 are common assumptions for spatial models, which are used in Su and Yang (2015), and the analog of  $0 < c_1 \leq \lambda_{\min}((nT)^{-1} \Sigma_{p+q+3})$  is employed in the assumption of Theorem 1 in Kelejian and Prucha (2001).

We now state the main results.

**Theorem 1** Suppose that Assumptions A1–A5 are satisfied. Then under model (1)–(2), as  $n \rightarrow \infty$ ,

$$\ell(\psi) \xrightarrow{d} \chi^2_{p+q+3},$$

where  $\chi^2_{p+q+3}$  is a chi-squared distributed random variable with  $p + q + 3$  degrees of freedom.

Let  $z_\alpha(p + q + 3)$  satisfy  $P(\chi^2_{p+q+3} \leq z_\alpha(p + q + 3)) = \alpha$  for  $0 < \alpha < 1$ . It follows from Theorem 1 that an EL based confidence region for  $\psi$  with asymptotically correct coverage probability  $\alpha$  can be constructed as

$$\{\psi : \ell(\psi) \leq z_\alpha(p + q + 3)\}.$$



### 3 Simulations

Recall that  $\theta = (\beta', \gamma', \rho)'$  and  $\psi = (\theta', \sigma_v^2, \lambda)'$ . Denote  $P_\lambda = \Omega^{-1}\Omega_\lambda\Omega^{-1}$ ,  $\Omega_\lambda = I_T \otimes A$  and  $\Omega_{\lambda\lambda} = I_T \otimes \{2(B'B)^{-1}[(W'_n B + B'W_n)A - W'_n W_n(B'B)^{-1}]\}$ . It can be shown that

$$\begin{aligned} \frac{\partial^2 \tilde{L}(\psi)}{\partial \theta \partial \theta'} &= -\frac{1}{\sigma_v^2} \tilde{X}' \Omega^{-1} \tilde{X}, \\ \frac{\partial^2 \tilde{L}(\psi)}{\partial \theta \partial \sigma_v^2} &= -\frac{1}{\sigma_v^4} \tilde{X}' \Omega^{-1} \epsilon, \\ \frac{\partial^2 \tilde{L}(\psi)}{\partial \theta \partial \lambda} &= -\frac{1}{\sigma_v^2} \tilde{X}' P_\lambda \epsilon, \\ \frac{\partial^2 \tilde{L}(\psi)}{\partial \sigma_v^2 \partial \sigma_v^2} &= -\frac{1}{\sigma_v^6} \epsilon' \Omega^{-1} \epsilon + \frac{nT}{2\sigma_v^4}, \\ \frac{\partial^2 \tilde{L}(\psi)}{\partial \sigma_v^2 \partial \lambda} &= -\frac{1}{2\sigma_v^4} \epsilon' P_\lambda \epsilon, \\ \frac{\partial^2 \tilde{L}(\psi)}{\partial \lambda^2} &= \frac{1}{2} \text{tr}(P_\lambda \Omega_\lambda - \Omega^{-1} \Omega_{\lambda\lambda}) \\ &\quad - \frac{1}{2\sigma_v^2} \epsilon' (2P_\lambda \Omega_\lambda - \Omega^{-1} \Omega_{\lambda\lambda}) \Omega^{-1} \epsilon. \end{aligned}$$

According to Su and Yang (2015), the QMLE  $\hat{\psi}$  of  $\psi$  satisfies:

$$\sqrt{nT}(\hat{\psi} - \psi) \xrightarrow{d} N(0, -\Sigma^{-1}),$$

where  $\Sigma = \lim_{n \rightarrow \infty} \frac{1}{nT} E[\Sigma_n(\psi)]$  and  $\Sigma_n(\psi) = \frac{\partial^2}{\partial \psi \partial \psi'} \tilde{L}(\psi)$ .

Based on the above asymptotic result, we can obtain the NA based confidence region for  $\psi$ . However, we note that the NA method depends on the availability of a consistent estimator of the asymptotic covariance matrix in practical applications, while the EL method does not. This can save the implementation time for the EL method and the EL method outperforms the NA method.

We conducted a small simulation study to compare the finite sample performances of the confidence regions based on EL and NA methods with confidence level  $\alpha = 0.95$ , and report the proportion of  $\ell(\psi) \leq z_{0.95}(p + q + 3)$  and  $(\hat{\psi} - \psi)'(-\Sigma)(\hat{\psi} - \psi) \leq z_{0.95}(p + q + 3)$  respectively in 1000 replications.

In the simulations, we used the following two models:

(1) Model 1:

$y_t = \rho y_{t-1} + x_t \beta + z\gamma + \epsilon_t$ ,  $\epsilon_t = \lambda W_n \epsilon_t + v_t$ ,  $t = 1, 2, 3$ , where  $x_t$  were generated from  $N(0, 4)$ , alternatively,  $x_t$  can be randomly generated in a similar fashion as in Hsiao et al. (2002), and the elements of  $z$  were randomly generated from Bernoulli(0.5). We selected  $\beta = 1$ ,  $\gamma = 1$ ,  $\sigma_v^2 = 1$  and  $(\rho, \lambda)$  were taken as  $(-0.8, -0.7)$ ,  $(-0.2, -0.1)$ ,  $(0.2, 0.1)$ ,  $(0.8, 0.7)$ ,  $(-0.8, 0.7)$  and  $(0.2, -0.1)$  respectively, and  $v_{it}$ 's were i.i.d. from  $N(0, 1)$ ,  $t(5)$  and  $\chi_4^2 - 4$ , respectively;

(2) Model 2:

$y_t = \rho y_{t-1} + x_t \beta + z \gamma + \epsilon_t$ ,  $\epsilon_t = \lambda W_n \epsilon_t + v_t$ ,  $t = 1, 2, 3$ , where  $x_t = \left( x_t^{(1)}, x_t^{(2)} \right)$  is an  $n \times 2$  matrix, where  $x_t^{(1)}$  were randomly generated from  $N(0, 1)$  and  $x_t^{(2)}$  were randomly generated from  $N(0, 4)$ . Moreover,  $z = \left( z^{(1)}, z^{(2)} \right)$  is an  $n \times 2$  matrix, the elements of  $z^{(1)}$  were randomly generated from Bernoulli(0.3) and the elements of  $z^{(2)}$  were randomly generated from Bernoulli(0.6). We selected  $\beta = (1.5, 1.0)'$ ,  $\gamma = (2, 1.2)'$ ,  $(\rho, \lambda)$  were taken as  $(-0.8, -0.7)$ ,  $(-0.2, -0.1)$ ,  $(0.2, 0.1)$ ,  $(0.8, 0.7)$ ,  $(-0.8, 0.7)$  and  $(0.2, -0.1)$  respectively, and  $v_{it}$ s were i.i.d. from  $N(0, 1)$ ,  $t(5)$ ,  $\chi_4^2 - 4$ ,  $0.1N(0, 4) + 0.9N(0, 1)$  and  $0.1t(3) + 0.9t(5)$ , respectively.

The results of simulations under model 1 are reported in Tables 1, 2 and 3, and the results of simulations under model 2 are reported in Tables 4, 5, 6, 7 and 8.

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 considered five 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. 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 NA behave well with coverage probabilities being very close to the nominal level 0.95 when the error term  $\epsilon_i$  is normally distributed and  $n$  is large, but not well in other cases. The coverage probabilities of the confidence regions based on NA fall to the

**Table 1** Coverage probabilities of the NA and EL confidence regions with  $v_{it} \sim N(0, 1)$  under model 1

$(\rho, \lambda)$	$W_n$	NA	EL	$(\rho, \lambda)$	$W_n$	NA	EL
(-0.8, -0.7)	$grid_{49}$	0.944	0.906	(-0.2, -0.1)	$grid_{49}$	0.947	0.896
	$grid_{100}$	0.953	0.930		$grid_{100}$	0.951	0.926
	$grid_{169}$	0.942	0.940		$grid_{169}$	0.954	0.936
	$grid_{256}$	0.944	0.946		$grid_{256}$	0.945	0.951
	$grid_{400}$	0.924	0.930		$grid_{400}$	0.952	0.950
(0.8, 0.7)	$grid_{49}$	0.927	0.895	(0.2, 0.1)	$grid_{49}$	0.945	0.904
	$grid_{100}$	0.926	0.930		$grid_{100}$	0.954	0.928
	$grid_{169}$	0.950	0.940		$grid_{169}$	0.951	0.944
	$grid_{256}$	0.940	0.935		$grid_{256}$	0.936	0.946
	$grid_{400}$	0.942	0.944		$grid_{400}$	0.964	0.956
(-0.8, 0.7)	$grid_{49}$	0.911	0.914	(0.2, -0.1)	$grid_{49}$	0.941	0.915
	$grid_{100}$	0.939	0.941		$grid_{100}$	0.949	0.934
	$grid_{169}$	0.948	0.930		$grid_{169}$	0.949	0.948
	$grid_{256}$	0.937	0.939		$grid_{256}$	0.951	0.944
	$grid_{400}$	0.958	0.956		$grid_{400}$	0.960	0.968

**Table 2** Coverage probabilities of the NA and EL confidence regions with  $v_{it} \sim t(5)$  under model 1

$(\rho, \lambda)$	$W_n$	NA	EL	$(\rho, \lambda)$	$W_n$	NA	EL
(-0.8, -0.7)	<i>grid</i> <sub>49</sub>	0.851	0.829	(-0.2, -0.1)	<i>grid</i> <sub>49</sub>	0.847	0.819
	<i>grid</i> <sub>100</sub>	0.830	0.876		<i>grid</i> <sub>100</sub>	0.853	0.888
	<i>grid</i> <sub>169</sub>	0.835	0.902		<i>grid</i> <sub>169</sub>	0.820	0.890
	<i>grid</i> <sub>256</sub>	0.832	0.922		<i>grid</i> <sub>256</sub>	0.816	0.916
	<i>grid</i> <sub>400</sub>	0.834	0.910		<i>grid</i> <sub>400</sub>	0.862	0.924
(0.8, 0.7)	<i>grid</i> <sub>49</sub>	0.829	0.845	(0.2, 0.1)	<i>grid</i> <sub>49</sub>	0.835	0.825
	<i>grid</i> <sub>100</sub>	0.832	0.888		<i>grid</i> <sub>100</sub>	0.847	0.876
	<i>grid</i> <sub>169</sub>	0.836	0.908		<i>grid</i> <sub>169</sub>	0.829	0.896
	<i>grid</i> <sub>256</sub>	0.828	0.902		<i>grid</i> <sub>256</sub>	0.837	0.916
	<i>grid</i> <sub>400</sub>	0.848	0.920		<i>grid</i> <sub>400</sub>	0.846	0.924
(-0.8, 0.7)	<i>grid</i> <sub>49</sub>	0.821	0.838	(0.2, -0.1)	<i>grid</i> <sub>49</sub>	0.855	0.842
	<i>grid</i> <sub>100</sub>	0.837	0.879		<i>grid</i> <sub>100</sub>	0.829	0.886
	<i>grid</i> <sub>169</sub>	0.819	0.898		<i>grid</i> <sub>169</sub>	0.853	0.903
	<i>grid</i> <sub>256</sub>	0.843	0.902		<i>grid</i> <sub>256</sub>	0.840	0.920
	<i>grid</i> <sub>400</sub>	0.828	0.914		<i>grid</i> <sub>400</sub>	0.812	0.916

**Table 3** Coverage probabilities of the NA and EL confidence regions with  $v_{it} + 4 \sim \chi^2_4$  under model 1

$(\rho, \lambda)$	$W_n$	NA	EL	$(\rho, \lambda)$	$W_n$	NA	EL
(-0.8, -0.7)	<i>grid</i> <sub>49</sub>	0.842	0.851	(-0.2, -0.1)	<i>grid</i> <sub>49</sub>	0.838	0.845
	<i>grid</i> <sub>100</sub>	0.842	0.890		<i>grid</i> <sub>100</sub>	0.838	0.874
	<i>grid</i> <sub>169</sub>	0.842	0.891		<i>grid</i> <sub>169</sub>	0.858	0.907
	<i>grid</i> <sub>256</sub>	0.830	0.896		<i>grid</i> <sub>256</sub>	0.862	0.914
	<i>grid</i> <sub>400</sub>	0.858	0.934		<i>grid</i> <sub>400</sub>	0.856	0.902
(0.8, 0.7)	<i>grid</i> <sub>49</sub>	0.809	0.827	(0.2, 0.1)	<i>grid</i> <sub>49</sub>	0.845	0.838
	<i>grid</i> <sub>100</sub>	0.827	0.883		<i>grid</i> <sub>100</sub>	0.850	0.890
	<i>grid</i> <sub>169</sub>	0.862	0.909		<i>grid</i> <sub>169</sub>	0.844	0.910
	<i>grid</i> <sub>256</sub>	0.833	0.907		<i>grid</i> <sub>256</sub>	0.858	0.914
	<i>grid</i> <sub>400</sub>	0.834	0.902		<i>grid</i> <sub>400</sub>	0.840	0.902
(-0.8, 0.7)	<i>grid</i> <sub>49</sub>	0.833	0.846	(0.2, -0.1)	<i>grid</i> <sub>49</sub>	0.843	0.841
	<i>grid</i> <sub>100</sub>	0.853	0.885		<i>grid</i> <sub>100</sub>	0.850	0.877
	<i>grid</i> <sub>169</sub>	0.843	0.907		<i>grid</i> <sub>169</sub>	0.856	0.913
	<i>grid</i> <sub>256</sub>	0.850	0.915		<i>grid</i> <sub>256</sub>	0.858	0.906
	<i>grid</i> <sub>400</sub>	0.868	0.926		<i>grid</i> <sub>400</sub>	0.850	0.936

range [0.812, 0.862] for the  $t$  distribution and [0.809, 0.868] for the  $\chi^2$  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, 2, 3, 4, 5, 6, 7 and 8, that the coverage probabilities of 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  $\epsilon_i$  is normally

**Table 4** Coverage probabilities of the NA and EL confidence regions with  $v_{it} \sim N(0, 1)$  under model 2

$(\rho, \lambda)$	$W_n$	NA	EL	$(\rho, \lambda)$	$W_n$	NA	EL
(-0.8, -0.7)	<i>grid</i> <sub>49</sub>	0.937	0.853	(-0.2, -0.1)	<i>grid</i> <sub>49</sub>	0.927	0.871
	<i>grid</i> <sub>100</sub>	0.934	0.905		<i>grid</i> <sub>100</sub>	0.943	0.927
	<i>grid</i> <sub>169</sub>	0.941	0.936		<i>grid</i> <sub>169</sub>	0.934	0.931
	<i>grid</i> <sub>256</sub>	0.947	0.934		<i>grid</i> <sub>256</sub>	0.946	0.938
	<i>grid</i> <sub>400</sub>	0.962	0.950		<i>grid</i> <sub>400</sub>	0.946	0.936
(0.8, 0.7)	<i>grid</i> <sub>49</sub>	0.928	0.887	(0.2, 0.1)	<i>grid</i> <sub>49</sub>	0.927	0.876
	<i>grid</i> <sub>100</sub>	0.939	0.927		<i>grid</i> <sub>100</sub>	0.937	0.909
	<i>grid</i> <sub>169</sub>	0.930	0.927		<i>grid</i> <sub>169</sub>	0.927	0.922
	<i>grid</i> <sub>256</sub>	0.934	0.944		<i>grid</i> <sub>256</sub>	0.939	0.940
	<i>grid</i> <sub>400</sub>	0.938	0.940		<i>grid</i> <sub>400</sub>	0.920	0.932
(-0.8, 0.7)	<i>grid</i> <sub>49</sub>	0.934	0.887	(0.2, -0.1)	<i>grid</i> <sub>49</sub>	0.940	0.883
	<i>grid</i> <sub>100</sub>	0.920	0.902		<i>grid</i> <sub>100</sub>	0.938	0.928
	<i>grid</i> <sub>169</sub>	0.942	0.927		<i>grid</i> <sub>169</sub>	0.944	0.926
	<i>grid</i> <sub>256</sub>	0.949	0.939		<i>grid</i> <sub>256</sub>	0.952	0.954
	<i>grid</i> <sub>400</sub>	0.948	0.942		<i>grid</i> <sub>400</sub>	0.938	0.932

**Table 5** Coverage probabilities of the NA and EL confidence regions with  $v_{it} \sim t(5)$  under model 2

$(\rho, \lambda)$	$W_n$	NA	EL	$(\rho, \lambda)$	$W_n$	NA	EL
(-0.8, -0.7)	<i>grid</i> <sub>49</sub>	0.843	0.787	(-0.2, -0.1)	<i>grid</i> <sub>49</sub>	0.853	0.797
	<i>grid</i> <sub>100</sub>	0.817	0.847		<i>grid</i> <sub>100</sub>	0.842	0.862
	<i>grid</i> <sub>169</sub>	0.863	0.887		<i>grid</i> <sub>169</sub>	0.837	0.891
	<i>grid</i> <sub>256</sub>	0.850	0.910		<i>grid</i> <sub>256</sub>	0.848	0.906
	<i>grid</i> <sub>400</sub>	0.864	0.920		<i>grid</i> <sub>400</sub>	0.846	0.912
(0.8, 0.7)	<i>grid</i> <sub>49</sub>	0.825	0.795	(0.2, 0.1)	<i>grid</i> <sub>49</sub>	0.850	0.801
	<i>grid</i> <sub>100</sub>	0.843	0.868		<i>grid</i> <sub>100</sub>	0.819	0.837
	<i>grid</i> <sub>169</sub>	0.832	0.885		<i>grid</i> <sub>169</sub>	0.854	0.899
	<i>grid</i> <sub>256</sub>	0.824	0.903		<i>grid</i> <sub>256</sub>	0.853	0.916
	<i>grid</i> <sub>400</sub>	0.824	0.894		<i>grid</i> <sub>400</sub>	0.838	0.928
(-0.8, 0.7)	<i>grid</i> <sub>49</sub>	0.820	0.794	(0.2, -0.1)	<i>grid</i> <sub>49</sub>	0.839	0.781
	<i>grid</i> <sub>100</sub>	0.857	0.874		<i>grid</i> <sub>100</sub>	0.844	0.844
	<i>grid</i> <sub>169</sub>	0.820	0.895		<i>grid</i> <sub>169</sub>	0.848	0.891
	<i>grid</i> <sub>256</sub>	0.844	0.907		<i>grid</i> <sub>256</sub>	0.828	0.875
	<i>grid</i> <sub>400</sub>	0.844	0.918		<i>grid</i> <sub>400</sub>	0.846	0.908

**Table 6** Coverage probabilities of the NA and EL confidence regions with  $v_{it} + 4 \sim \chi^2_4$  under model 2

$(\rho, \lambda)$	$W_n$	NA	EL	$(\rho, \lambda)$	$W_n$	NA	EL
(-0.8, -0.7)	<i>grid</i> <sub>49</sub>	0.840	0.811	(-0.2, -0.1)	<i>grid</i> <sub>49</sub>	0.835	0.809
	<i>grid</i> <sub>100</sub>	0.845	0.867		<i>grid</i> <sub>100</sub>	0.830	0.874
	<i>grid</i> <sub>169</sub>	0.854	0.909		<i>grid</i> <sub>169</sub>	0.857	0.883
	<i>grid</i> <sub>256</sub>	0.852	0.904		<i>grid</i> <sub>256</sub>	0.845	0.907
	<i>grid</i> <sub>400</sub>	0.872	0.930		<i>grid</i> <sub>400</sub>	0.888	0.928
(0.8, 0.7)	<i>grid</i> <sub>49</sub>	0.792	0.784	(0.2, 0.1)	<i>grid</i> <sub>49</sub>	0.802	0.823
	<i>grid</i> <sub>100</sub>	0.831	0.850		<i>grid</i> <sub>100</sub>	0.817	0.842
	<i>grid</i> <sub>169</sub>	0.849	0.899		<i>grid</i> <sub>169</sub>	0.833	0.878
	<i>grid</i> <sub>256</sub>	0.872	0.912		<i>grid</i> <sub>256</sub>	0.873	0.911
	<i>grid</i> <sub>400</sub>	0.854	0.928		<i>grid</i> <sub>400</sub>	0.884	0.922
(-0.8, 0.7)	<i>grid</i> <sub>49</sub>	0.825	0.809	(0.2, -0.1)	<i>grid</i> <sub>49</sub>	0.821	0.807
	<i>grid</i> <sub>100</sub>	0.831	0.855		<i>grid</i> <sub>100</sub>	0.839	0.866
	<i>grid</i> <sub>169</sub>	0.833	0.878		<i>grid</i> <sub>169</sub>	0.845	0.891
	<i>grid</i> <sub>256</sub>	0.863	0.914		<i>grid</i> <sub>256</sub>	0.864	0.911
	<i>grid</i> <sub>400</sub>	0.862	0.920		<i>grid</i> <sub>400</sub>	0.880	0.930

**Table 7** Coverage probabilities of the NA and EL confidence regions with  $v_{it} \sim 0.1N(0, 4) + 0.9N(0, 1)$  under model 2

$(\rho, \lambda)$	$W_n$	NA	EL	$(\rho, \lambda)$	$W_n$	NA	EL
(-0.8, -0.7)	<i>grid</i> <sub>49</sub>	0.908	0.844	(-0.2, -0.1)	<i>grid</i> <sub>49</sub>	0.906	0.848
	<i>grid</i> <sub>100</sub>	0.930	0.906		<i>grid</i> <sub>100</sub>	0.924	0.912
	<i>grid</i> <sub>169</sub>	0.956	0.938		<i>grid</i> <sub>169</sub>	0.946	0.946
	<i>grid</i> <sub>256</sub>	0.954	0.938		<i>grid</i> <sub>256</sub>	0.948	0.948
	<i>grid</i> <sub>400</sub>	0.938	0.944		<i>grid</i> <sub>400</sub>	0.942	0.954
(0.8, 0.7)	<i>grid</i> <sub>49</sub>	0.910	0.874	(0.2, 0.1)	<i>grid</i> <sub>49</sub>	0.918	0.848
	<i>grid</i> <sub>100</sub>	0.950	0.934		<i>grid</i> <sub>100</sub>	0.936	0.930
	<i>grid</i> <sub>169</sub>	0.938	0.938		<i>grid</i> <sub>169</sub>	0.938	0.920
	<i>grid</i> <sub>256</sub>	0.950	0.936		<i>grid</i> <sub>256</sub>	0.924	0.928
	<i>grid</i> <sub>400</sub>	0.942	0.942		<i>grid</i> <sub>400</sub>	0.958	0.956
(-0.8, 0.7)	<i>grid</i> <sub>49</sub>	0.906	0.854	(0.2, -0.1)	<i>grid</i> <sub>49</sub>	0.932	0.860
	<i>grid</i> <sub>100</sub>	0.918	0.918		<i>grid</i> <sub>100</sub>	0.942	0.926
	<i>grid</i> <sub>169</sub>	0.936	0.940		<i>grid</i> <sub>169</sub>	0.918	0.918
	<i>grid</i> <sub>256</sub>	0.942	0.934		<i>grid</i> <sub>256</sub>	0.948	0.938
	<i>grid</i> <sub>400</sub>	0.954	0.938		<i>grid</i> <sub>400</sub>	0.960	0.958

**Table 8** Coverage probabilities of the NA and EL confidence regions with  $v_{it} \sim 0.1t(3) + 0.9t(5)$  under model 2

$(\rho, \lambda)$	$W_n$	NA	EL	$(\rho, \lambda)$	$W_n$	NA	EL
(-0.8, -0.7)	<i>grid</i> <sub>49</sub>	0.850	0.788	(-0.2, -0.1)	<i>grid</i> <sub>49</sub>	0.866	0.796
	<i>grid</i> <sub>100</sub>	0.828	0.846		<i>grid</i> <sub>100</sub>	0.838	0.860
	<i>grid</i> <sub>169</sub>	0.870	0.902		<i>grid</i> <sub>169</sub>	0.838	0.900
	<i>grid</i> <sub>256</sub>	0.838	0.930		<i>grid</i> <sub>256</sub>	0.832	0.892
	<i>grid</i> <sub>400</sub>	0.858	0.912		<i>grid</i> <sub>400</sub>	0.878	0.906
(0.8, 0.7)	<i>grid</i> <sub>49</sub>	0.826	0.776	(0.2, 0.1)	<i>grid</i> <sub>49</sub>	0.856	0.790
	<i>grid</i> <sub>100</sub>	0.840	0.832		<i>grid</i> <sub>100</sub>	0.814	0.844
	<i>grid</i> <sub>169</sub>	0.866	0.904		<i>grid</i> <sub>169</sub>	0.814	0.896
	<i>grid</i> <sub>256</sub>	0.828	0.908		<i>grid</i> <sub>256</sub>	0.834	0.912
	<i>grid</i> <sub>400</sub>	0.870	0.926		<i>grid</i> <sub>400</sub>	0.868	0.914
(-0.8, 0.7)	<i>grid</i> <sub>49</sub>	0.836	0.776	(0.2, -0.1)	<i>grid</i> <sub>49</sub>	0.848	0.822
	<i>grid</i> <sub>100</sub>	0.838	0.866		<i>grid</i> <sub>100</sub>	0.852	0.868
	<i>grid</i> <sub>169</sub>	0.838	0.868		<i>grid</i> <sub>169</sub>	0.842	0.880
	<i>grid</i> <sub>256</sub>	0.872	0.904		<i>grid</i> <sub>256</sub>	0.842	0.900
	<i>grid</i> <sub>400</sub>	0.850	0.906		<i>grid</i> <sub>400</sub>	0.884	0.928

distributed or not. These results show that the EL based confidence regions generally outperform the NA based confidence regions when  $n$  is large enough.

### 4 A real data example

In order to illustrate the proposed method in Sect. 2, we conducted a real data analysis. The data come from 288 prefecture-level cities in China, collected from National Bureau of Statistics of China and Anjuke. There were three variables: the logarithm of housing price per square meter ( $y_t$ ), the logarithm of income per household ( $x_t$ ) and the urbanization rate ( $z$ ) from the years of 2010 to 2017. In order to ensure the stability and eliminate the influence of dimension, we first did difference and standardization on the above data, and then considered fitting the data via the following model:

**Table 9** Analysis results for the average price of commercial housing data (with ALs shown in brackets)

Variable	Estimation	Confidence interval (EL)	Confidence interval (NA)
$\beta$	0.7015	[0.6521, 0.7442] (0.0921)	[0.6495, 0.7535] (0.1040)
$\gamma$	0.1677	[0.1519, 0.2103] (0.0584)	[0.1319, 0.2035] (0.0716)
$\rho$	0.0093	[-0.0204, 0.0411] (0.0615)	[-0.0511, 0.0697] (0.1208)
$\lambda$	0.3743	[0.2775, 0.3830] (0.1055)	[0.3137, 0.4349] (0.1212)
$\sigma_v^2$	0.0110	[0.0105, 0.0119] (0.0014)	[0.0103, 0.0118] (0.0015)

$y_t = \rho y_{t-1} + x_t \beta + z_t \gamma + \epsilon_t$ ,  $\epsilon_t = \lambda W_n \epsilon_t + v_t$ ,  $t = 1, 2, \dots, 8$ , where  $n = 288$  and the spatial weighting matrix  $W_n$  was selected by the method in Sect. 3.

We separately employed the EL method in Sect. 2 and the NA method in Sect. 3 to obtain the confidence intervals for parameters  $\beta, \gamma, \rho, \lambda$  and  $\sigma_v^2$  with confidence level 0.95, which were shown in Table 9.

Table 9 shows that the estimator of the spatial parameter is  $\lambda = 0.3743$ , and 0 is not in its confidence interval, which implies that there exists a spatial relationship among the disturbances. The results also show that the lengths of the EL based intervals are uniformly shorter than those of the NA based intervals, which implies that the EL based method performs better than the NA based method for the real data.

### Appendix

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

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

where  $\epsilon_{ni}$  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 above assumptions (C1) and (C2), the mean and variance of  $\tilde{Q}_n$  are given as (e.g. Kelejjan & Prucha, 2001)

$$\begin{aligned} v_{\tilde{Q}} &= \sum_{i=1}^n a_{nii} \sigma_{ni}^2, \\ \sigma_{\tilde{Q}}^2 &= 2 \sum_{i=1}^n \sum_{j=1}^n a_{nij}^2 \sigma_{ni}^2 \sigma_{nj}^2 + \sum_{i=1}^n b_{ni}^2 \sigma_{ni}^2 \\ &\quad + \sum_{i=1}^n \{ a_{nii}^2 (\mu_{ni}^{(4)} - 3\sigma_{ni}^4) + 2b_{ni} a_{nii} \mu_{ni}^{(3)} \}, \end{aligned} \tag{25}$$

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

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

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

**Lemma 2** Let  $\xi_1, \xi_2, \dots, \xi_n$  be a sequence of stationary random variables, with  $E|\xi_1|^s < \infty$  for some constants  $s > 0$ . Then

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

**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  $\xi_1, \xi_2, \dots, \xi_n$  are independent in using Borel–Cantelli lemma. □

**Lemma 3** Suppose that Assumptions A1–A5 are satisfied. Then as  $n \rightarrow \infty$ ,

$$Z_n = \max_{1 \leq i \leq nT} |\omega_i(\psi)| = o_p((nT)^{2/(4+n_1)}) \text{ a.s.}, \tag{26}$$

$$\Sigma_{p+q+3}^{-1/2} \sum_{i=1}^{nT} \omega_i(\psi) \xrightarrow{d} N(0, I_{p+q+3}), \tag{27}$$

$$(nT)^{-1} \sum_{i=1}^{nT} \omega_i(\psi) \omega'_i(\psi) = (nT)^{-1} \Sigma_{p+q+3} + o_p(1), \tag{28}$$

$$\sum_{i=1}^{nT} |\omega_i(\psi)|^3 = O_p(nT). \tag{29}$$

**Proof** Note that

$$\begin{aligned} Z_n \leq \max_{1 \leq i \leq nT} \left\{ \max_{1 \leq i \leq nT} |a_{i,1} e_i|, \max_{1 \leq i \leq nT} |e_i^2 - \sigma_v^2|, \max_{1 \leq i \leq nT} \left| a_{i,2} e_i + h_{ii,1} (e_i^2 - \sigma_v^2) \right. \right. \\ \left. \left. + 2e_i \sum_{j=1}^{i-1} h_{ij,1} e_j \right|, \max_{1 \leq i \leq nT} \left| h_{ii,2} (e_i^2 - \sigma_v^2) + 2e_i \sum_{j=1}^{i-1} h_{ij,2} e_j \right| \right\}. \end{aligned}$$

By Conditions A1–A3 and Lemma 2, we have

$$\max_{1 \leq i \leq nT} |a_{i,k} e_i| = o_p((nT)^{1/(4+n_1)}), \quad \max_{1 \leq i \leq nT} |e_i^2 - \sigma_v^2| = o_p((nT)^{2/(4+n_1)}).$$

In addition, by Lemma B.2. in Su and Yang (2015),  $A'_v(I_T \otimes B')$  and  $(I_T \otimes (BAB'))$  are uniformly bounded in both row and column sums, it follows that



$$\begin{aligned} \max_{1 \leq i \leq nT} |h_{ii,k}(e_i^2 - \sigma_v^2)| &= \max_{1 \leq i \leq nT} |h_{ii,k}| o_p((nT)^{2/(4+\eta_1)}) = o_p((nT)^{2/(4+\eta_1)}). \\ \max_{1 \leq i \leq nT} \left| e_i \sum_{j=1}^{i-1} h_{ij,k} e_j \right| &\leq \left( \max_{1 \leq i \leq nT} |e_i| \right)^2 \cdot \max_{1 \leq i \leq nT} \left( \sum_{j=1}^{i-1} |h_{ij,k}| \right) = o_p((nT)^{2/(4+\eta_1)}), k = 1, 2. \end{aligned}$$

Thus  $Z_n = o_p((nT)^{2/(4+\eta_1)})$ . (26) is proved. □

We now prove (27). For any given  $l = (l'_1, l_2, l_3, l_4)' \in R^{p+q+3}$  with  $\|l\| = 1$ , where  $l_1 \in R^{p+q}, l_2, l_3, l_4 \in R$ , it is clear that

$$\begin{aligned} l' \omega_i(\psi) &= l'_1 a_{i,1} e_i + l_2 (e_i^2 - \sigma_v^2) + l_3 \{ a_{i,2} e_i + h_{ii,1} (e_i^2 - \sigma_v^2) + 2e_i \sum_{j=1}^{i-1} h_{ij,1} e_j \} \\ &\quad + l_4 \{ h_{ii,2} (e_i^2 - \sigma_v^2) + 2e_i \sum_{j=1}^{i-1} h_{ij,2} e_j \} \\ &= (l_2 + l_3 h_{ii,1} + l_4 h_{ii,2}) (e_i^2 - \sigma_v^2) + 2e_i \sum_{j=1}^{i-1} (l_3 h_{ij,1} + l_4 h_{ij,2}) e_j \\ &\quad + (l'_1 a_{i,1} + l_3 a_{i,2}) e_i. \end{aligned}$$

Denote

$$Q_n = \sum_{i=1}^{nT} l' \omega_i(\psi) = \sum_{i=1}^{nT} \sum_{j=1}^{nT} u_{ij} e_i e_j + \sum_{i=1}^{nT} b_i e_i = e' U_{1n} e + U_{2n} e,$$

where

$$\begin{aligned} U_{1n} &= (u_{ij})_{(nT) \times (nT)}, \quad U_{2n} = (b_i)_{1 \times (nT)}, \\ u_{ii} &= l_2 + l_3 h_{ii,1} + l_4 h_{ii,2}, \quad u_{ij} = l_3 h_{ij,1} + l_4 h_{ij,2} (i \neq j), \quad b_i = l'_1 a_{i,1} + l_3 a_{i,2}. \end{aligned}$$

Note that

$$U_{1n} = l_2 I_{nT} + l_3 H_1 + l_4 H_2, \quad U_{2n} = l'_1 \tilde{X}'_1 (I_T \otimes B') + l_3 \tilde{X}'_2 (I_T \otimes B').$$

The conditional expectation and variance given  $X, Z$  are denoted as  $E^*$  and  $Var^*$ , respectively. Then from (15) and note that  $E(v) = 0$ , we know that the variance of  $Q_n$  is

$$\begin{aligned} \sigma_{Q_n}^2 &= Var \left( \sum_{i=1}^{nT} l' \omega_i(\psi) \right) \\ &= Var(v' U_{1n} v) + Var(U_{2n} v) + 2Cov(v' U_{1n} v, U_{2n} v), \end{aligned} \tag{30}$$

and

$$\begin{aligned} \text{Var}^*(Q_n) &= (\vartheta_4 - 3\sigma_v^4) \|\text{vec}_D(U_{1n})\|^2 + \sigma_v^4 [\text{tr}(U_{1n}U'_{1n}) + \text{tr}(U_{1n}^2)], \\ &+ \sigma_v^2 U_{2n}U'_{2n} + 2\vartheta_3 U_{2n} \text{vec}_D(U_{1n}). \end{aligned} \tag{31}$$

Further,

$$\begin{aligned} \|\text{vec}_D(U_{1n})\|^2 &= \|\text{vec}_D(l_2 I_{nT} + l_3 H_1 + l_4 H_2)\|^2 \\ &= \|l_2 \text{vec}_D(I_{nT}) + l_3 \text{vec}_D(H_1) + l_4 \text{vec}_D(H_2)\|^2 \\ &= l_2^2 nT + l_3^2 \|\text{vec}_D(H_1)\|^2 + l_4^2 \|\text{vec}_D(H_2)\|^2 + 2l_2 l_3 \mathbf{1}'_{nT} \text{vec}_D(H_1) \\ &+ 2l_2 l_4 \mathbf{1}'_{nT} \text{vec}_D(H_2) + 2l_3 l_4 \text{vec}'_D(H_1) \text{vec}_D(H_2) \\ &= \tilde{l}' G_1 \tilde{l}, \end{aligned} \tag{32}$$

where  $\tilde{l} = (l_2, l_3, l_4)'$ ,  $G_1 = \begin{pmatrix} nT & \mathbf{1}'_{nT} \text{vec}_D(H_1) & \mathbf{1}'_{nT} \text{vec}_D(H_2) \\ * & \|\text{vec}_D(H_1)\|^2 & \text{vec}'_D(H_1) \text{vec}_D(H_2) \\ * & * & \|\text{vec}_D(H_2)\|^2 \end{pmatrix}$ . And

$$\begin{aligned} \text{tr}(U_{1n}U'_{1n}) &= l_2^2 nT + 2l_2 l_3 \text{tr}(H_1) + 2l_2 l_4 \text{tr}(H_2) + l_3^2 \text{tr}(H_1^2) \\ &+ 2l_3 l_4 \text{tr}(H_1 H_2) + l_4^2 \text{tr}(H_2^2) \\ &= \tilde{l}' G_2 \tilde{l}, \end{aligned} \tag{33}$$

where  $G_2 = \begin{pmatrix} nT & \text{tr}(H_1) & \text{tr}(H_2) \\ * & \text{tr}(H_1^2) & \text{tr}(H_1 H_2) \\ * & * & \text{tr}(H_2^2) \end{pmatrix}$ . Moreover,

$$U_{2n}U'_{2n} = l'_1 \tilde{X}'_1 \Omega^{-1} \tilde{X}_1 l_1 + l'_3 \tilde{X}'_2 \Omega^{-1} \tilde{X}_2 + 2l'_1 l_3 \tilde{X}'_1 \Omega^{-1} \tilde{X}_2. \tag{34}$$

It is easy to show (e.g. Su & Yang, 2015) that,  $n^{-1} \tilde{X}'_1 \Omega^{-1} \tilde{X}_1$ ,  $n^{-1} \tilde{X}'_2 \Omega^{-1} \tilde{X}_2$ , and  $n^{-1} \tilde{X}'_1 \Omega^{-1} \tilde{X}_2$  converge in probability to their expectations. We have

$$\begin{aligned} U_{2n}U'_{2n} &= l'_1 E\left(\tilde{X}'_1 \Omega^{-1} \tilde{X}_1\right) l_1 + l'_3 E\left(\tilde{X}'_2 \Omega^{-1} \tilde{X}_2\right) \\ &+ 2l'_1 l_3 E\left(\tilde{X}'_1 \Omega^{-1} \tilde{X}_2\right) + o_p(n), \end{aligned} \tag{35}$$

and

$$\begin{aligned} U_{2n} \text{vec}_D(U_{1n}) &= l'_1 l_2 \tilde{X}'_1 (I_T \otimes B') \mathbf{1}_{nT} + l'_1 l_3 \tilde{X}'_1 (I_T \otimes B') \text{vec}_D(H_1) \\ &+ l'_1 l_4 \tilde{X}'_1 (I_T \otimes B') \text{vec}_D(H_2) + l_3 l_2 \tilde{X}'_2 (I_T \otimes B') \mathbf{1}_{nT} \\ &+ l'_3 \tilde{X}'_2 (I_T \otimes B') \text{vec}_D(H_1) + l_3 l_4 \tilde{X}'_2 (I_T \otimes B') \text{vec}_D(H_2) \\ &= \tilde{l}' (\mathbf{1}_{nT}, \text{vec}_D(H_1), \text{vec}_D(H_2))' (I_T \otimes B) E(\tilde{X}_1) l_1 \\ &+ \tilde{l}' l_3 (\mathbf{1}_{nT}, \text{vec}_D(H_1), \text{vec}_D(H_2))' (I_T \otimes B) E(\tilde{X}_2) + o_p(n). \end{aligned} \tag{36}$$

Combine (31)–(36), we have

$$\begin{aligned}
 \text{Var}^*(Q_n) &= (\vartheta_4 - 3\sigma_v^4) \|\text{vec}_D(U_{1n})\|^2 + \sigma_v^4 [\text{tr}(U_{1n}U'_{1n}) + \text{tr}(U_{1n}^2)] \\
 &\quad + \sigma_v^2 U_{2n}U'_{2n} + 2\vartheta_3 U_{2n}\text{vec}_D(U_{1n}) \\
 &= (\vartheta_4 - 3\sigma_v^4) \tilde{l}' G_1 \tilde{l} + 2\sigma_v^4 \tilde{l}' G_2 \tilde{l} + l'_1 \sigma_v^2 E(\tilde{X}'_1 \Omega^{-1} \tilde{X}_1) l_1 \\
 &\quad + l_3^2 \sigma_v^2 E(\tilde{X}'_2 \Omega^{-1} \tilde{X}_2) + 2l'_1 l_3 \sigma_v^2 E(\tilde{X}'_1 \Omega^{-1} \tilde{X}_2) \\
 &\quad + 2\tilde{l}' \vartheta_3 (\mathbf{1}_{nT}, \text{vec}_D(H_1), \text{vec}_D(H_2))' (I_T \otimes B) E(\tilde{X}_1) l_1 \\
 &\quad + 2\tilde{l}' l_3 \vartheta_3 (\mathbf{1}_{nT}, \text{vec}_D(H_1), \text{vec}_D(H_2))' (I_T \otimes B) E(\tilde{X}_2) + o_p(n) \\
 &= l' \Sigma_{p+q+3} l + o_p(n),
 \end{aligned} \tag{37}$$

where  $\Sigma_{p+q+3}$  is given in (24). From Condition A4, one can see that  $(nT)^{-1} \text{Var}^*(Q_n) \geq c_1 > 0$ . From Lemma 1, we have

$$\frac{Q_n - E^*(Q_n)}{\sqrt{\text{Var}^*(Q_n)}} \xrightarrow{d^*} N(0, 1), \tag{38}$$

where  $d^*$  stands for convergence in distribution given  $X, Z$ . Noting that  $(nT)^{-1} \text{Var}^*(Q_n) \geq c_1 > 0$  and

$$\text{Var}^*(Q_n) = \sigma_{Q_n}^2 + o_p(n),$$

one can show that

$$\frac{\sigma_{Q_n}^2}{\text{Var}^*(Q_n)} \xrightarrow{P} 1. \tag{39}$$

Combing  $E^*(Q_n) = 0$ , (38) and (39), we thus have

$$\frac{Q_n}{\sigma_{Q_n}} \xrightarrow{d} N(0, 1).$$

Then (27) holds true.

Next we will prove (28), i. e.

$$(nT)^{-1} \sum_{i=1}^{nT} (l' \omega_i(\psi))^2 = (nT)^{-1} \sigma_{Q_n}^2 + o_p(1). \tag{40}$$

Let

$$\begin{aligned}
 N_{in} &= l' \omega_i(\psi) \\
 &= u_{ii}(e_i^2 - \sigma_v^2) + 2 \sum_{j=1}^{i-1} u_{ij} e_i e_j + b_i e_i \\
 &= u_{ii}(e_i^2 - \sigma_v^2) + R_i e_i,
 \end{aligned} \tag{41}$$

where  $R_i = 2 \sum_{j=1}^{i-1} u_{ij}e_j + b_i$ . Let  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ ,  $\mathcal{F}_i = \sigma(e_1, e_2, \dots, e_i)$ ,  $1 \leq i \leq nT$ . Then  $\{N_{in}, \mathcal{F}_i, 1 \leq i \leq nT\}$  form a martingale difference array given  $X, Z$ . From (30) and (37), one can see that

$$\sigma_{Q_n}^2 = \sum_{i=1}^{nT} E^*(N_{in}^2) + o_p(n).$$

It follows that

$$\begin{aligned} & (nT)^{-1} \sum_{i=1}^{nT} \{\ell' \omega_i(\psi)\}^2 - (nT)^{-1} \sigma_{Q_n}^2 \\ &= (nT)^{-1} \sum_{i=1}^{nT} (N_{in}^2 - E^*(N_{in}^2)) + o_p(1) \\ &= (nT)^{-1} \sum_{i=1}^{nT} \{N_{in}^2 - E^*(N_{in}^2 | \mathcal{F}_{i-1}) + E^*(N_{in}^2 | \mathcal{F}_{i-1}) - E^*(N_{in}^2)\} + o_p(1) \\ &= (nT)^{-1} S_{n1} + (nT)^{-1} S_{n2} + o_p(1), \end{aligned} \tag{42}$$

where  $S_{n1} = \sum_{i=1}^{nT} \{N_{in}^2 - E^*(N_{in}^2 | \mathcal{F}_{i-1})\}$ ,  $S_{n2} = \sum_{i=1}^{nT} \{E^*(N_{in}^2 | \mathcal{F}_{i-1}) - E^*(N_{in}^2)\}$ . Next we will show that: (1)

$$(nT)^{-1} S_{n1} = o_p(1), \tag{43}$$

and (2)

$$(nT)^{-1} S_{n2} = o_p(1). \tag{44}$$

To show (1) and (2), it is sufficient to show that  $(nT)^{-2} E(S_{n1}^2) \rightarrow 0$  and  $(nT)^{-2} E(S_{n2}^2) \rightarrow 0$ , respectively. Obviously,

$$N_{in}^2 = u_{ii}^2(e_i^2 - \sigma_v^2)^2 + R_i^2 e_i^2 + 2u_{ii}R_i(e_i^2 - \sigma_v^2)e_i.$$

Thus

$$E^*(N_{in}^2 | \mathcal{F}_{i-1}) = u_{ii}^2 E(e_i^2 - \sigma_v^2)^2 + R_i^2 \sigma_v^2 + 2u_{ii}R_i \vartheta_3.$$

It follows that

$$\begin{aligned}
 (nT)^{-2}E(S_{n1}^2) &= (nT)^{-2} \sum_{i=1}^{nT} E\{N_{in}^2 - E^*(N_{in}^2|\mathcal{F}_{i-1})\}^2 \\
 &= (nT)^{-2} \sum_{i=1}^{nT} E[u_{ii}^2\{(e_i^2 - \sigma_v^2)^2 - E(e_i^2 - \sigma_v^2)^2\} + R_i^2(e_i^2 - \sigma_v^2) \\
 &\quad + 2u_{ii}R_i(e_i^3 - \sigma_v^2e_i - \vartheta_3)]^2 \\
 &\leq C(nT)^{-2} \sum_{i=1}^{nT} E[u_{ii}^4\{(e_i^2 - \sigma_v^2)^2 - E(e_i^2 - \sigma_v^2)^2\}^2] \\
 &\quad + C(nT)^{-2} \sum_{i=1}^{nT} E\{R_i^4(e_i^2 - \sigma_v^2)^2\} \\
 &\quad + C(nT)^{-2} \sum_{i=1}^{nT} E\{u_{ii}^2R_i^2(e_i^3 - \sigma_v^2e_i - \vartheta_3)^2\}.
 \end{aligned} \tag{45}$$

By Conditions A1–A3, we have

$$\begin{aligned}
 &(nT)^{-2} \sum_{i=1}^{nT} E[u_{ii}^4\{(e_i^2 - \sigma_v^2)^2 - E(e_i^2 - \sigma_v^2)^2\}^2] \\
 &\leq C(nT)^{-2} \sum_{i=1}^{nT} u_{ii}^4 \leq C(nT)^{-2} \sum_{i=1}^{nT} |l_2 + l_3h_{ii,1} + l_4h_{ii,2}|^4 \\
 &\leq C(nT)^{-2} \sum_{i=1}^{nT} |l_2 + l_3h_{ii,1} + l_4h_{ii,2}|^4 \leq Cn^{-1} \rightarrow 0,
 \end{aligned} \tag{46}$$

and

$$\begin{aligned}
 (nT)^{-2} \sum_{i=1}^{nT} E\{R_i^4(e_i^2 - \sigma_v^2)^2\} &= (nT)^{-2} \sum_{i=1}^{nT} E[E^*\{R_i^4(e_i^2 - \sigma_v^2)^2\}] \\
 &\leq C(nT)^{-2} \sum_{i=1}^{nT} E\left(\sum_{j=1}^{i-1} u_{ij}e_j + b_i\right)^4 \\
 &\leq C(nT)^{-2} \sum_{i=1}^{nT} E\left(\sum_{j=1}^{i-1} u_{ij}e_j\right)^4 + C(nT)^{-2} \sum_{i=1}^{nT} Eb_i^4 \\
 &\leq C(nT)^{-2} \sum_{i=1}^{nT} \sum_{j=1}^{i-1} u_{ij}^4 \vartheta_4 + C(nT)^{-2} \sum_{i=1}^{nT} \left(\sum_{j=1}^{i-1} u_{ij}^2 \sigma_v^2\right)^2 \\
 &\quad + C(nT)^{-2} \sum_{i=1}^{nT} E(l_1 a_{i,1} + l_3 a_{i,2})^4 \\
 &\leq C(nT)^{-2} \sum_{i=1}^{nT} \sum_{j=1}^{i-1} |l_3 h_{ij,1} + l_4 h_{ij,2}|^4 \\
 &\quad + C(nT)^{-2} \sum_{i=1}^{nT} \left(\sum_{j=1}^{i-1} |l_3 h_{ij,1} + l_4 h_{ij,2}|^2\right)^2 \\
 &\quad + C(nT)^{-2} \sum_{i=1}^{nT} E(l_1 a_{i,1} + l_3 a_{i,2})^4 \leq Cn^{-1} \rightarrow 0.
 \end{aligned}
 \tag{47}$$

Similarly, we can prove that

$$(nT)^{-2} \sum_{i=1}^{nT} E\{u_{ii}^2 R_i^2(e_i^3 - \sigma^2 e_i - \vartheta_3)^2\} \rightarrow 0.
 \tag{48}$$

From (45)–(48), we have  $(nT)^{-2} E(S_{n1}^2) \rightarrow 0$ . Furthermore,

$$\begin{aligned}
 E^*(N_{in}^2) &= E^*\{E^*(N_{in}^2 | \mathcal{F}_{i-1})\} = u_{ii}^2 E(e_i^2 - \sigma_v^2)^2 + \sigma_v^2 E^*(R_i^2) + 2u_{ii} \vartheta_3 E^*(R_i) \\
 &= u_{ii}^2 E(e_i^2 - \sigma_v^2)^2 + \sigma_v^2 \left(4 \sum_{j=1}^{i-1} u_{ij}^2 \sigma_v^2 + b_i^2\right) + 2u_{ii} \vartheta_3 b_i.
 \end{aligned}$$

Thus,

$$\begin{aligned}
 (nT)^{-2}E(S_{n_2}^2) &= (nT)^{-2}E \left[ \sum_{i=1}^{nT} \{E^*(N_{in}^2 | \mathcal{F}_{i-1}) - E^*(N_{in}^2)\} \right]^2 \\
 &= (nT)^{-2}E \left[ \sum_{i=1}^{nT} \left\{ R_i^2 \sigma_v^2 - \sigma_v^2 \left( 4 \sum_{j=1}^{i-1} u_{ij}^2 \sigma_v^2 + b_i^2 \right) + 2u_{ii} \vartheta_3 (R_i - b_i) \right\} \right]^2 \\
 &= (nT)^{-2} \sum_{i=1}^{nT} E \left[ \sigma_v^2 \left\{ \left( 2 \sum_{j=1}^{i-1} u_{ij} e_j \right)^2 - 4 \sum_{j=1}^{i-1} u_{ij}^2 \sigma_v^2 \right\} + 4 \left( \sum_{j=1}^{i-1} u_{ij} e_j \right) b_i \sigma_v^2 \right. \\
 &\quad \left. + 2u_{ii} \vartheta_3 \left( 2 \sum_{j=1}^{i-1} u_{ij} e_j \right) \right]^2 \\
 &\leq C(nT)^{-2} \sum_{i=1}^{nT} E \left[ \sigma_v^2 \left\{ \left( \sum_{j=1}^{i-1} u_{ij} e_j \right)^2 - \sum_{j=1}^{i-1} u_{ij}^2 \sigma_v^2 \right\} \right]^2 \\
 &\quad + C(nT)^{-2} \sum_{i=1}^{nT} E \left\{ \left( \sum_{j=1}^{i-1} u_{ij} e_j \right) b_i \sigma_v^2 \right\}^2 \\
 &\quad + C(nT)^{-2} \sum_{i=1}^{nT} E \left\{ 2u_{ii} \vartheta_3 \left( \sum_{j=1}^{i-1} u_{ij} e_j \right) \right\}^2.
 \end{aligned} \tag{49}$$

Note that

$$\begin{aligned}
 (nT)^{-2} \sum_{i=1}^{nT} E \left[ \sigma_v^2 \left\{ \left( \sum_{j=1}^{i-1} u_{ij} e_j \right)^2 - \sum_{j=1}^{i-1} u_{ij}^2 \sigma_v^2 \right\} \right]^2 \\
 \leq (nT)^{-2} \sigma_v^4 \sum_{i=1}^{nT} E \left( \sum_{j=1}^{i-1} u_{ij} e_j \right)^4 \\
 \leq C(nT)^{-2} \sum_{i=1}^{nT} \sum_{j=1}^{i-1} u_{ij}^4 \vartheta_4 + C(nT)^{-2} \sum_{i=1}^{nT} \left( \sum_{j=1}^{i-1} u_{ij}^2 \sigma_v^2 \right)^2 \\
 \leq Cn^{-1} \rightarrow 0,
 \end{aligned} \tag{50}$$

$$\begin{aligned}
 (nT)^{-2} \sum_{i=1}^{nT} E \left\{ \left( \sum_{j=1}^{i-1} u_{ij} e_j \right) b_i \sigma_v^2 \right\}^2 &= (nT)^{-2} \sigma_v^6 \sum_{i=1}^{nT} E(b_i^2) \sum_{j=1}^{i-1} u_{ij}^2 \\
 &\leq C(nT)^{-2} \rightarrow 0,
 \end{aligned} \tag{51}$$

and

$$\begin{aligned}
 (nT)^{-2} \sum_{i=1}^{nT} E \left\{ 2u_{ii} \vartheta_3 \left( \sum_{j=1}^{i-1} u_{ij} e_j \right) \right\}^2 &= 4\vartheta_3^2 \sigma_v^2 (nT)^{-2} \sum_{i=1}^{nT} u_{ii}^2 \sum_{j=1}^{i-1} u_{ij}^2 \\
 &\leq C(nT)^{-1} \rightarrow 0,
 \end{aligned}
 \tag{52}$$

where we have used Conditions A2 and A3. From (49)–(52), we have  $(nT)^{-2} ES_{n2}^2 \rightarrow 0$ . The proof of (28) is thus complete.

Finally, we will prove (29). Note that

$$\begin{aligned}
 \sum_{i=1}^{nT} E \|\omega_i(\psi)\|^3 &\leq \sum_{i=1}^{nT} E \|a_{i,1} e_i\|^3 + \sum_{i=1}^{nT} E |e_i^2 - \sigma_v^2|^3 \\
 &+ \sum_{i=1}^{nT} E |a_{i,2} e_i + h_{ii,1}(e_i^2 - \sigma_v^2) + 2e_i \sum_{j=1}^{i-1} h_{ij,1} e_j|^3 \\
 &+ \sum_{i=1}^{nT} E |h_{ii,2}(e_i^2 - \sigma_v^2) + 2e_i \sum_{j=1}^{i-1} h_{ij,2} e_j|^3.
 \end{aligned}
 \tag{53}$$

By Conditions A2 and A3, we have

$$\sum_{i=1}^{nT} E \|a_{i,1} e_i\|^3 \leq CnT \left( \max_{1 \leq i \leq nT} E \|a_{i,1}\|^3 \right) E |e_i|^3 = O(nT),
 \tag{54}$$

$$\sum_{i=1}^{nT} E |e_i^2 - \sigma_v^2|^3 = O(nT),
 \tag{55}$$

$$\begin{aligned}
 &\sum_{i=1}^{nT} E \left| a_{i,2} e_i + h_{ii,1}(e_i^2 - \sigma_v^2) + 2e_i \sum_{j=1}^{i-1} h_{ij,1} e_j \right|^3 \\
 &\leq C \sum_{i=1}^{nT} E |a_{i,2} e_i|^3 + C \sum_{i=1}^{nT} E |h_{ii,1}(e_i^2 - \sigma_v^2)|^3 + C \sum_{i=1}^{nT} E \left| 2e_i \sum_{j=1}^{i-1} h_{ij,1} e_j \right|^3 \\
 &\leq C \sum_{i=1}^{nT} E |a_{i,2}|^3 E |e_i|^3 + C \sum_{i=1}^{nT} E |h_{ii,1}(e_i^2 - \sigma_v^2)|^3 \\
 &\quad + C \sum_{i=1}^{nT} E |e_i|^3 \sum_{j=1}^{i-1} E |h_{ij,1} e_j|^3 + C \sum_{i=1}^{nT} E |e_i|^3 \left\{ \sum_{j=1}^{i-1} E (h_{ij,1} e_j)^2 \right\}^{3/2} \\
 &= O(nT).
 \end{aligned}
 \tag{56}$$

Similarly,



$$\sum_{i=1}^{nT} E \left| h_{ii,2}(e_i^2 - \sigma_v^2) + 2e_i \sum_{j=1}^{i-1} h_{ij,2}e_j \right|^3 = O(nT). \quad (57)$$

From (53)–(57), we have

$$\sum_{i=1}^{nT} E \|\omega_i(\psi)\|^3 = O(nT). \quad (58)$$

Further, using (58) and Markov inequality, we obtain  $\sum_{i=1}^{nT} \|\omega_i(\psi)\|^3 = O_p(nT^2)$ . Thus (29) is proved.

**Proof of Theorem 1** Using Lemma 3 and following the proof of Theorem 1 in Qin (2021), one can easily show that Theorem 1 holds true.  $\square$

**Acknowledgements** This work was partially supported by the National Natural Science Foundation of China (12061017, 12161009). The authors are thankful to the referees for constructive suggestions.

## References

- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic Press.
- Anselin, L. (2001). Spatial econometrics. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics* (pp. 310–330). Blackwell Publishers Ltd.
- Baltagi, B., Egger, P., & Pfaffermayr, M. (2013). A generalized spatial panel model with random effects. *Econometric Reviews*, 32, 650–685.
- Baltagi, B., & Li, D. (2006). Prediction in the panel data model with spatial correlation: The case of liquor. *Spatial Economic Analysis*, 1, 175–185.
- Baltagi, B., Song, S. K., Jung, B. C., & Koh, W. (2007). Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics*, 140, 5–51.
- Baltagi, B., Song, S. H., & Koh, W. (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics*, 117, 123–150.
- Bandyopadhyay, S., Lahiri, S. N., & Nordman, D. J. (2015). A frequency domain empirical likelihood method for irregularly spaced spatial data. *The Annals of Statistics*, 43(2), 519–545.
- Chen, X., & Conley, T. G. (2001). A new semiparametric spatial model for panel time series. *Journal of Econometrics*, 105, 59–83.
- Chen, S. X., & Keilegom, I. V. (2009). A review on empirical likelihood for regressions (with discussions). *Test*, 3, 415–447.
- Chen, J., & Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika*, 80, 107–116.
- Elhorst, J. P. (2003). Specification and estimation of spatial panel data models. *International Regional Science Review*, 26, 244–268.
- Elhorst, J. P. (2005). Unconditional maximum likelihood estimation of linear and loglinear dynamic models for spatial panels. *Geographical Analysis*, 37, 85–106.
- Elhorst, J. P. (2010). Dynamic panels with endogenous interaction effects when T is small. *Regional Science and Urban Economics*, 40, 272–282.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag.
- Hall, P., & La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *International Statistical Review*, 58, 109–127.

- Hsiao, C., Pesaran, M. H., & Tahmiscioglu, A. K. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics*, *109*, 107–150.
- Jin, F., & Lee, L. F. (2019). GEL estimation and tests of spatial autoregressive models. *Journal of Econometrics*, *208*, 585–612.
- Kapoor, M., Kelejian, H. H., & Prucha, I. R. (2007). Panel data models with spatially correlated error components. *Journal of Econometrics*, *140*, 97–130.
- Kelejian, H. H., & Prucha, I. R. (2001). On the asymptotic distribution of the Moran  $I$  test statistic with applications. *Journal of Econometrics*, *104*, 219–257.
- Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *The Annals of Statistics*, *25*, 2084–2102.
- Lee, L. F., & Yu, J. (2010a). A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory*, *26*, 564–597.
- Lee, L. F., & Yu, J. (2010b). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, *154*(2), 165–185.
- Lee, L. F., & Yu, J. (2010c). Some recent developments in spatial panel data models. *Regional Science and Urban Economics*, *40*, 255–271.
- Lee, L. F., Yu, J., (2015a). Spatial Panel Data Models. In: Baltagi B (ed) *The Oxford Handbooks: Panel Data*. Oxford University Press, Oxford, England.
- Lee, L. F., & Yu, J. (2015b). Estimation of fixed effects panel regression models with separable and non-separable space-time filters. *Journal of Econometrics*, *184*, 174–192.
- Li, Y., & Qin, Y. (2020). Empirical likelihood for panel data models with spatial errors. *Communications in Statistics-Theory and Methods*. <https://doi.org/10.1080/03610926.2020.1780449>.
- Mutl, J. (2006). *Dynamic Panel Data Models with Spatially Correlated Disturbances*. College Park: University of Maryland. (Ph.D. thesis).
- Mutl, J., & Pfaffermayr, M. (2011). The Hausman test in a Cliff and Ord panel model. *The Economic Journal*, *14*, 48–76.
- Nordman, D. J. (2008a). A blockwise empirical likelihood for spatial lattice data, *Statist. Sinica*, *18*, 1111–1129.
- Nordman, D. J. (2008b). An empirical likelihood method for spatial regression. *Metrika*, *68*, 351–363.
- 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. *The Annals of Statistics*, *18*, 90–120.
- Owen, A. B. (2001). *Empirical Likelihood*. Chapman & Hall.
- Parent, O., & LeSage, J. P. (2011). A spaceCtime filter for panel data models containing random effects. *Computational Statistics & Data Analysis*, *55*, 475–490.
- Pesaran, M. H. (2004). General Diagnostic Tests for Cross Section Dependence in Panels, Working Paper No. 1229, University of Cambridge.
- Qin, Y. (2021). Empirical likelihood for spatial autoregressive models with spatial autoregressive disturbances. *Sankhyā A: The Indian Journal of Statistics*, *83*, 1–25.
- Qin, J., & Lawless, J. (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics*, *22*, 300–325.
- Su, L., & Yang, Z. (2007). *QML Estimation of Dynamic Panel Data Models with Spatial Errors*, Working Paper. Singapore Management University.
- Su, L., & Yang, Z. (2015). QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics*, *185*, 230–258.
- Wu, C. B. (2004). Weighted empirical likelihood inference. *Statistics & Probability Letters*, *66*, 67–79.
- Yang, Z., Li, C., & Tse, Y. K. (2006). Functional form and spatial dependence in dynamic panels. *Economics Letters*, *91*, 138–145.
- Yu, J., de Jong, R., & Lee, L. F. (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both  $n$  and  $T$  are large. *Journal of Econometrics*, *146*, 118–134.
- Yu, J., & Lee, L. F. (2010). Estimation of unit root spatial dynamic panel data models. *Econometric Theory*, *26*, 1332–1362.
- Zhong, B., & Rao, J. N. K. (2000). Empirical likelihood inference under stratified random sampling using auxiliary population information. *Biometrika*, *87*, 929–938.