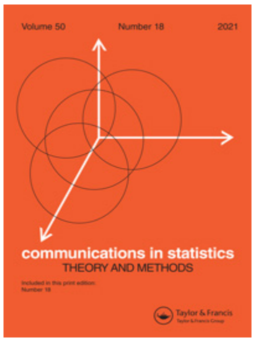

Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Theory and Methods

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/lsta20

Empirical likelihood and GMM for spatial models

Yongsong Qin

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2021, VOL. 50, NO. 18, 4367–4385
https://doi.org/10.1080/03610926.2020.1716252

Empirical likelihood and GMM for spatial models

Yongsong Qin

Department of Statistics, Guangxi Normal University, Guilin, Guangxi, China

ABSTRACT

We link the empirical likelihood (EL) and GMM for three major spatial models: spatial autoregressive model with spatial autoregressive disturbances (SARAR model), linear regression model with spatial autoregressive errors (SE model) and spatial autoregressive model (SAR model). It is shown that for every GMM estimator (GMME), there is an empirical likelihood (EL) estimator which has the same asymptotic variance as the GMME. Specifically, we show that there exists an EL estimator which is asymptotically efficient as the best GMME proposed by Liu *et al.* [Liu, X. D., L. F. Lee, and C. R. Bollinger. 2010. An efficient GMM estimator of spatial autoregressive models. Journal of Econometrics 159 (2):303–19] and the EL confidence regions for the parameters in above models can be constructed without the estimation of asymptotic variances.

ARTICLE HISTORY
Received 3 April 2019
Accepted 7 January 2020

KEYWORDS
Spatial model; GMM; empirical likelihood; confidence region

AMS 2010 SUBJECT CLASSIFICATION:
Primary: 62G05;
Secondary: 62E20

1. Introduction

Spatial econometrics models have found many applications in various fields of economics such as regional, urban and public economics where spatial dependence among cross-sectional units are involved (e.g., Cliff and Ord 1973; Anselin 1988). The study of spatial econometrics models has been an active field of statistical research for the last 30 years. In this article, we focus on the following spatial autoregressive model with spatial autoregressive disturbances (SARAR model):

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

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

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

This model is introduced by Cliff and Ord (1973). This model has been extensively studied for more than 30 years. Excellent surveys and developments in testing and

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

estimation of this model can be found in Cliff and Ord (1973), Anselin (1988), Cressien (1993), Anselin and Bera (1998), Kelejian and Prucha (2001) and Liu, Lee, and Bollinger (2010), among others. For the model (1.1), there are two special cases: $\rho_1 = 0$ and $\rho_2 = 0$. In the former case, the model is called linear regression model with spatial autoregressive errors (SE model). In the later case, the model is called spatial autoregressive model (SAR model).

There exist two major estimation approaches for the parameters in the above spatial models. One is the maximum likelihood (ML) method (e.g., Anselin 1988). The other is the computationally more efficient approach, the generalized method of moments (GMM) by Liu, Lee, and Bollinger (2010). Liu, Lee, and Bollinger (2010) have obtained the best GMM estimator within the class of GMM estimators based on linear and quadratic moment conditions. It is shown in Liu, Lee, and Bollinger (2010) that the best GMM estimator is asymptotically efficient as the ML estimator under normality. In this article, we propose to use the empirical likelihood (EL) method introduced by Owen (1988, 1990) to estimate and construct confidence region for the parameters in the SARAR, SE and SAR models. As a nonparametric method, the EL method does not require to specify the distribution form of the population in study. Moreover, the shape and orientation of the EL confidence region are determined by data and the confidence region is obtained without covariance estimation. There is a lot of excellent research work for EL method. Here, we only mention a small part of them. A comprehensive review on EL for regressions can be found in Chen and Keilegom (2009). More references on EL method can be found in Owen (2001) and Qin and Lawless (1994), among others.

The EL method depends on the GMM in choosing optimum estimation equations. The main challenge in using the EL method is that the estimating equations based on GMM for the SARAR, SE and SAR models contain linear-quadratic forms of ϵ_n. The idea to solve this problem is to introduce martingale sequences to transform the quadratic forms into linear forms of martingale sequences. We show that for every GMME in Liu, Lee, and Bollinger (2010), there is an EL estimator which has the same asymptotic variance as the GMME. Specifically, it is shown that there exists an EL estimator which is asymptotically efficient as the best GMME proposed by Lee Liu, Lee, and Bollinger (2010). More significantly, in this article, the EL confidence regions for the parameters in the SARAR, SE and SAR models are constructed without the estimation of asymptotic variances. We anticipate deeper and richer literature in this direction. The theory of EL method in this article is developed under the assumption that the model is correctly specified. As noted by Schennach (2007), the EL estimator possesses some undesirable properties when the model is misspecified.

The remaining of this article is organized as follows. Section 2 states the main results. Results from a simulation study are presented in Section 3. All the technical details are given in Section 4.

2. Main results

In the following we will study EL for SARAR, SE and SAR models, respectively.

2.1. EL for SARAR models

Let $\theta = (\rho_2, \rho_1, \beta^\tau)^\tau$ and use $\theta_0 = (\rho_{20}, \rho_{10}, \beta_0^\tau)^\tau$ to denote the true value of θ. Furthermore, let $S_n(\rho_1) = I_n - \rho_1 W_n$ and $R_n(\rho_2) = I_n - \rho_2 M_n$. For simplicity, denote $S_n = S_n(\rho_{10})$ and $R_n = R_n(\rho_{20})$.

For the estimation of the model (1.1), we change the model at θ_0 into the form: $R_n(S_n Y_n - X_n\beta_0) = \epsilon_n$. Then denote $\epsilon_n(\theta) = R_n(\rho_2)\{S_n(\rho_1)Y_n - X_n\beta\}$ for any possible value θ. Therefore, $\epsilon_n(\theta_0) = \epsilon_n$. We will use the linear-quadratic moment functions proposed in Liu, Lee, and Bollinger (2010) to construct EL score functions. Let Q_n be an $n \times q$ matrix of instrumental variables (IVs) constructed as functions of X_n, W_n and M_n. Let $\mathcal{P}_1$ be the class of constant (i.e., nonrandom) $n \times n$ matrices with a zero trace. With the selected matrices $P_{sn} \in \mathcal{P}_1, 1 \le s \le m, m \ge 1,$ and IV matrix Q_n, the following moment functions are used to form GMM estimators by Liu, Lee, and Bollinger (2010):

$$g_n(\theta) = \Big(\epsilon_n^\tau(\theta)Q_n, \epsilon_n^\tau(\theta)P_{1n}\epsilon_n(\theta), \epsilon_n^\tau(\theta)P_{2n}\epsilon_n(\theta), \ldots, \epsilon_n^\tau(\theta)P_{mn}\epsilon_n(\theta)\Big)^\tau \tag{2.1}$$

In this article, the following symmetrized form is used:

$$\begin{aligned} g_n(\theta) = (&\epsilon_n^\tau(\theta)Q_n, \epsilon_n^\tau(\theta)\tilde{P}_{1n}\epsilon_n(\theta), \epsilon_n^\tau(\theta)\tilde{P}_{2n}\epsilon_n(\theta), \\ &\cdots, \epsilon_n^\tau(\theta)\tilde{P}_{mn}\epsilon_n(\theta))^\tau \end{aligned} \tag{2.2}$$

where $\tilde{P}_{sn} = (P_{sn} + P_{sn}^\tau)/2, 1 \le s \le m.$ It is clear that $Eg_n(\theta_0) = 0$ as $E\{\epsilon_n^\tau(\theta_0)\tilde{P}_{sn}\epsilon_n(\theta_0)\} = \sigma_0^2 tr(\tilde{P}_{sn}) = 0,$ where σ_0^2 denotes the true value of σ^2. We use $\tilde{P}_{ijsn}$ and b_i^τ to denote the (i, j) element of the matrix $\tilde{P}_{sn}$ and the ith row of the matrix Q_n, respectively, and adapt the convention that any sum with an upper index of less than one is zero. Let $\epsilon_n = (\epsilon_{n1}, \epsilon_{n2}, \ldots, \epsilon_{nn})^\tau$. To deal with the quadratic forms of ϵ_n in $g_n(\theta_0)$, we follow Kelejian and Prucha (2001) to introduce a martingale difference array for every quadratic form. Define the σ-fields: $\mathcal{F}_0 = \{\emptyset, \tilde{\Omega}\}, \mathcal{F}_i = \sigma(\epsilon_{n1}, \epsilon_{n2}, \ldots, \epsilon_{ni}), 1 \le i \le n$. Let

$$\tilde{Y}_{isn} = \tilde{P}_{iisn}(\epsilon_{ni}^2 - \sigma_0^2) + 2\epsilon_{ni}\sum_{j=1}^{i-1}\tilde{P}_{ijsn}\epsilon_{nj} \tag{2.3}$$

Then $\mathcal{F}_{i-1} \subseteq \mathcal{F}_i, \tilde{Y}_{isn}$ is $\mathcal{F}_i-$ measurable and $E(\tilde{Y}_{isn}|\mathcal{F}_{i-1}) = 0$. Thus $\{\tilde{Y}_{isn}, \mathcal{F}_i, 1 \le i \le n\}$ form a martingale difference array and by $P_{ns} \in \mathcal{P}_1$,

$$\epsilon_n^\tau\tilde{P}_{sn}\epsilon_n = \sum_{i=1}^{n}\tilde{Y}_{isn}, 1 \le s \le m \tag{2.4}$$

In this way, a quadratic form of ϵ_n is changed into a linear form of a martingale difference sequence, which enables the application of EL method.

In this article, we focus on the EL estimator of θ. However, we need to construct an joint estimator of θ and σ^2 first. Let $\psi = (\theta^\tau, \sigma^2)^\tau$ and use $\psi_0 = (\theta_0^\tau, \sigma_0^2)^\tau$ to denote the true values of ψ. Based on (2.2) and (2.4), we propose the following EL ratio statistic for $\psi \in R^{k+3}$:

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

where $\{p_i\}$ satisfy

$$p_i \geq 0, 1 \leq i \leq n, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \omega_i(\psi) = 0 \tag{2.5}$$

where

$$\omega_i(\psi) = \begin{pmatrix} b_i \epsilon_{ni}(\theta) \\ \tilde{P}_{ii1n}\{\epsilon_{ni}^2(\theta) - \sigma^2\} + 2\epsilon_{ni}(\theta) \sum_{j=1}^{i-1} \tilde{P}_{ij1n} \epsilon_{nj}(\theta) \\ \tilde{P}_{ii2n}\{\epsilon_{ni}^2(\theta) - \sigma^2\} + 2\epsilon_{ni}(\theta) \sum_{j=1}^{i-1} \tilde{P}_{ij2n} \epsilon_{nj}(\theta) \\ \vdots \\ \tilde{P}_{iimn}\{\epsilon_{ni}^2(\theta) - \sigma^2\} + 2\epsilon_{ni}(\theta) \sum_{j=1}^{i-1} \tilde{P}_{ijmn} \epsilon_{nj}(\theta) \end{pmatrix}$$

where $\epsilon_{ni}(\theta)$ is the ith component of $\epsilon_n(\theta) = R_n(\rho_2)\{S_n(\rho_1)Y_n - X_n\beta\}$ for any possible θ. Suppose that 0 is inside the convex hull of the points $\{\omega_i(\psi), 1 \leq i \leq n\}$ for given ψ. Following Owen (1990), one can show that

$$\log L_n(\psi) = -\sum_{i=1}^{n} \log\{1 + \lambda^\tau(\psi)\omega_i(\psi)\} \tag{2.6}$$

where $\lambda(\psi)$ is the solution of the following equation:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\omega_i(\psi)}{1 + \lambda^\tau(\psi)\omega_i(\psi)} = 0 \tag{2.7}$$

Let $\hat{\psi}_n$ be the maximizer of $L_n(\psi)$ over the parameter space Ψ, which is called the EL estimator of ψ. For any two vectors $f(x) = (f_1(x), f_2(x), \ldots, f_r(x))^\tau$ and $x = (x_1, x_2, \ldots, x_s)^\tau$, define

$$\frac{\partial f(x)}{\partial x} = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & \cdots & \frac{\partial f_1(x)}{\partial x_s} \\ \frac{\partial f_2(x)}{\partial x_1} & \frac{\partial f_2(x)}{\partial x_2} & \cdots & \frac{\partial f_2(x)}{\partial x_s} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_r(x)}{\partial x_1} & \frac{\partial f_r(x)}{\partial x_2} & \cdots & \frac{\partial f_r(x)}{\partial x_s} \end{pmatrix}, \frac{\partial f(x)}{\partial x^\tau} = \left\{\frac{\partial f(x)}{\partial x}\right\}^\tau$$

Differentiating $\log L_n(\psi)$, we have the likelihood equation:

$$\frac{1}{n}\sum_{i=1}^{n} \frac{1}{1 + \lambda^\tau(\psi)\omega_i(\psi)} \left(\frac{\partial \omega_i(\psi)}{\partial \psi}\right)^\tau \lambda(\psi) = 0 \tag{2.8}$$

Let $\lambda(\psi) = \lambda$. The EL estimator $\hat{\psi}_n$ of ψ is also defined as the solution of the following Equations (2.9) and (2.10):

$$U_{1n}(\psi, \lambda) \hat{=} \frac{1}{n}\sum_{i=1}^{n} \frac{\omega_i(\psi)}{1+\lambda^\tau \omega_i(\psi)} = 0 \tag{2.9}$$

$$U_{2n}(\psi, \lambda) \hat{=} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1+\lambda^\tau \omega_i(\psi)} \left(\frac{\partial \omega_i(\psi)}{\partial \psi}\right)^\tau \lambda = 0 \tag{2.10}$$

In other words, $\hat{\psi}_n$ and $\hat{\lambda} = \lambda(\hat{\psi}_n)$ satisfy $U_{sn}(\hat{\psi}_n, \hat{\lambda}) = 0, s = 1, 2$. The first $k+2$ components of $\hat{\psi}_n$ is $\hat{\theta}_n$, which is the EL estimator of θ.

To obtain the asymptotic distribution of $\hat{\theta}_n$, we need following assumptions.

A1. $\{\epsilon_{ni}, 1 \le i \le n\}$ are independent and identically distributed random variables with mean 0, variance $\sigma_0^2 > 0$ and $E|\epsilon_{n1}|^{4+\eta_1} < \infty$ for some $\eta_1 > 0$.

A2. The elements of X_n are uniformly bounded, X_n has the full column rank k, and $\lim_{n\to\infty} \frac{1}{n} X_n^\tau X_n$ exists and is nonsingular.

A3. S_n^{-1} and R_n^{-1} exist. W_n, M_n, S_n^{-1} and R_n^{-1} are uniformly bounded in both row and column sums in absolute value.

A4. $P_{sn}, 1 \le s \le m$, are uniformly bounded in both row and column sums in absolute value, and the elements of Q_n are uniformly bounded.

A5. $\lim_{n\to\infty} \frac{1}{n}\Omega_n = \Omega$ exists and is a nonsingular matrix, where $\Omega_n = var\{g_n(\theta_0)\}$.

A6. $\lim_{n\to\infty} \frac{1}{n} D_n = D$ exists and $Rank(D)$ is the dimension of θ, where $D_n = -E(\frac{\partial g_n(\theta_0)}{\partial \theta})$.

Remark 1. Conditions A1 to A6 are common assumptions for SARAR models, which are also employed by Liu, Lee, and Bollinger (2010).

Use $Vec_D(A)$ to denote the column vector formed with the diagonal elements of A. Let

$$\begin{gathered} A^{(s)} = A + A^\tau, \mu_j = E(\epsilon_{n1}^j), j = 3, 4 \\ \bar{X}_n = R_n X_n, H_n = M_n R_n^{-1}, G_n = W_n S_n^{-1}, \bar{G}_n = R_n G_n R_n^{-1} \end{gathered} \tag{2.11}$$

The first result in this article establishes the asymptotic normality of $\hat{\theta}_n$.

Proposition 1. *Suppose that Assumptions (A1)-(A6) are satisfied and $P_{sn} \in \mathcal{P}_1$, $1 \le s \le m$. As $n \to \infty$,*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, \Sigma)$$

where

$$\Sigma = \left\{ (\lim_{n\to\infty} \frac{1}{n} D_n)^\tau (\lim_{n\to\infty} \frac{1}{n}\Omega_n)^{-1} (\lim_{n\to\infty} \frac{1}{n} D_n) \right\}^{-1}$$

$$D_n = \begin{pmatrix} 0 & Q_n^\tau \bar{G}_n \bar{X}_n \beta_0 & Q_n^\tau \bar{X}_n \\ \sigma_0^2 tr(P_{1n}^{(s)} H_n) & \sigma_0^2 tr(P_{1n}^{(s)} \bar{G}_n) & 0 \\ \vdots & \vdots & \vdots \\ \sigma_0^2 tr(P_{mn}^{(s)} H_n) & \sigma_0^2 tr(P_{mn}^{(s)} \bar{G}_n) & 0 \end{pmatrix}$$

and

$$\Omega_n = \Omega_{n1} + \sigma_0^4 \begin{pmatrix} \frac{1}{\sigma_0^2} Q_n^\tau Q_n & 0 & \cdots & 0 \\ 0 & tr(P_{1n}P_{1n}^{(s)}) & \cdots & tr(P_{1n}P_{mn}^{(s)}) \\ \vdots & \vdots & & \vdots \\ 0 & tr(P_{mn}P_{1n}^{(s)}) & \cdots & tr(P_{mn}P_{mn}^{(s)}) \end{pmatrix}$$

where

$$\Omega_{n1} = \begin{pmatrix} 0 & \mu_3 Q_n^\tau \omega_{mn} \\ \mu_3 \omega_{mn}^\tau Q_n & (\mu_4 - 3\sigma_0^4)\omega_{mn}^\tau \omega_{mn} \end{pmatrix}$$

with $\omega_{mn} = (Vec_D(P_{1n}), \ldots, Vec_D(P_{mn}))$.

Remark 2. Compared with Proposition 1 in Liu, Lee, and Bollinger (2010), $\hat{\theta}_n$ has the same limiting distribution as the optimum GMM estimator (OGMME).

The best GMM estimator (BGMME) is also obtained by choosing the best IVs Q_n and weighting matrices $\{P_{sn}\}$ by Liu, Lee, and Bollinger (2010). To state this result, we need more notations as follows. Define $\bar{X}_n^*$ as the submatrix of $\bar{X}_n$ with the intercept column deleted (if no intercept column in $\bar{X}_n, \bar{X}_n^* = \bar{X}_n$). Suppose the number of columns in $\bar{X}_n^*$ is k^*. For $1 \le j \le k^*$, use $\bar{X}_{nj}$ and $\bar{X}_{nj}^*$ to denote the jth columns of $\bar{X}_n$ and $\bar{X}_n^*$, respectively. Let $A^{(t)} = A - \frac{1}{n} tr(A) I_n$ for a $n \times n$ matrix A. Use $D(A)$ to denote a diagonal matrix with diagonal elements being A if A is a vector, or diagonal elements of A if A is a square matrix. Let

$$P_{1n} = \bar{G}_n^{(t)}, P_{2n} = D(\bar{G}_n^{(t)}), P_{3n} = (D(\bar{G}_n \bar{X}_n \beta_0))^{(t)}, P_{4n} = H_n^{(t)}, P_{5n} = D(H_n^{(t)})$$
$$P_{j+5,n} = (D(\bar{X}_{nj}^*))^{(t)}, j = 1, 2, \ldots, k^*, Q_n = (Q_{1n}, Q_{2n}, Q_{3n}, Q_{4n}, Q_{5n})$$

where

$$Q_{1n} = \bar{X}_n^*, Q_{2n} = \bar{G}_n \bar{X}_n \beta_0, Q_{3n} = 1_n, Q_{3n} = Vec_D(\bar{G}_n^{(t)}), Q_{5n} = Vec_D(H_n^{(t)})$$

with 1_n being the n-dimensional (column) vector with 1 as its components. It is shown in Liu, Lee, and Bollinger (2010), that above $\{P_{sn}, 1 \le s \le k^* + 5\}$ and Q_n provide the set of best IVs and weighting matrices. In this case, the resulting EL estimator $\hat{\theta}_n$ has the same asymptotic variance Σ_B as the BGMME, where Σ_B is given in (3) in Liu, Lee, and Bollinger (2010), which is apparently a special case of Σ in Proposition 1.

As there are unknown parameters θ_0 in the IVs and weighting matrices, in practice, we can initially give the $\sqrt{n}$-consistent estimators $\hat{\theta}_n$ of θ_0 to obtain the estimated IVs and weighting matrices: $\hat{P}_{in} = P_{in}(\hat{\theta}_n), i = 1, 2, \ldots, k^* + 5, \hat{Q}_{jn} = Q_{jn}(\hat{\theta}_n), j = 1, 2, \ldots, 5$. Then following the above procedures, we use these estimated IVs and weighting matrices to construct the likelihood function $\hat{L}_n(\psi)$ and obtain the EL estimator $\hat{\theta}_{bn}$ of θ_0.

We now state the asymptotic normality of $\hat{\theta}_{bn}$.

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

$$\sqrt{n}(\hat{\theta}_{bn} - \theta_0) \xrightarrow{d} N(0, \Sigma_b)$$

where Σ_b is given in (3) in Liu, Lee, and Bollinger (2010).

Remark 3. $\hat{\theta}_{bn}$ has the same limiting distribution as the best GMM estimator (BGMME).

To show the advantage of the proposed EL approach to statistical inference, we consider the properties of the EL ratios induced from $\hat{L}_n(\psi)$. Let $\ell_n(\theta_0) = 2\log\{\hat{L}_n(\hat{\psi}_{bn})\} - 2\log\{\hat{L}_n(\theta_0, \hat{\sigma}^2_{bn})\}$. The following result establishes the asymptotic distribution of $\ell_n(\theta_0)$.

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

$$\ell_n(\theta_0) \xrightarrow{d} \chi^2_{k+2}$$

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

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

$$\{\theta : \ell_n(\theta) \le z_\alpha(k+2)\}$$

2.2. EL for SE models

For an SE model, $\rho_1 = 0$. Let $\theta = (\rho_2, \beta^\tau)^\tau$ and use $\theta_0 = (\rho_{20}, \beta_0^\tau)^\tau$ to denote the true value of θ. Let $R_n(\rho_2) = I_n - \rho_2 M_n$ and $R_n = R_n(\rho_{20})$. Denote $\epsilon_n(\theta) = R_n(\rho_2)(Y_n - X_n\beta)$ for any possible value θ. Let $\psi = (\theta^\tau, \sigma^2)^\tau$ and use $\psi_0 = (\theta_0^\tau, \sigma_0^2)^\tau$ to denote the true values of ψ. Let

$$P_{1n} = H_n^{(t)}, P_{2n} = D(H_n^{(t)}), P_{j+2,n} = (D(\bar{X}^*_{nj}))^{(t)}, j = 1, 2, ..., k^*$$
$$Q_n = (Q_{1n}, Q_{2n}, Q_{3n}), \text{ with } Q_{1n} = \bar{X}^*_n, Q_{2n} = 1_n, Q_{3n} = Vec_D(H_n^{(t)})$$

It is shown in Liu, Lee, and Bollinger (2010), that above $\{P_{sn}, 1 \le s \le k^* + 2\}$ and Q_n provide the set of best IVs and weighting matrices. Following the procedure in Section 2.1, the resulting EL estimator $\hat{\theta}_n$ of θ_0 has the same asymptotic variance $\Sigma_{B\rho_2}$ as the BGMME, where $\Sigma_{B\rho_2}$ is given in (5) in Liu, Lee, and Bollinger (2010).

In practice, we first give the $\sqrt{n}$-consistent estimators $\hat{\theta}_n$ of θ_0 to obtain the estimated IVs and weighting matrices: $\hat{P}_{in} = P_{in}(\hat{\theta}_n), i = 1, 2, ..., k^* + 2, \hat{Q}_{jn} = Q_{jn}(\hat{\theta}_n), j = 1, 2, 3.$ Then following the procedure in Section 2.1, we can construct the likelihood function $\hat{L}_n(\psi)$ and obtain the EL estimator $\hat{\theta}_{bn}$ of θ_0, which has the same limiting distribution as the BGMME.

Proposition 3. *Suppose that Assumptions (A1)-(A6) are satisfied. Then as $n \to \infty$,*

$$\sqrt{n}(\hat{\theta}_{bn} - \theta_0) \xrightarrow{d} N(0, \Sigma_{B\rho_2})$$

where $\Sigma_{B\rho_2}$ is given in (5) in Liu, Lee, and Bollinger (2010).

Following the procedure in Section 2.1, we let $\ell_n(\theta_0) = 2\log\{\hat{L}_n(\hat{\psi}_{bn})\} - 2\log\{\hat{L}_n(\theta_0, \hat{\sigma}^2_{bn})\}$.

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

$$\ell_n(\theta_0) \xrightarrow{d} \chi^2_{k+1}$$

where χ^2_{k+1} is a chi-squared distributed random variable with $k+1$ degrees of freedom.

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

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

2.3. EL for SAR models

$\rho_2 = 0$ for an SAR model. Let $\theta = (\rho_1, \beta^\tau)^\tau$ and use $\theta_0 = (\rho_{10}, \beta_0^\tau)^\tau$ to denote the true value of θ. Let $S_n(\rho_1) = I_n - \rho_1 W_n$ and $S_n = S_n(\rho_{10})$. Denote $\epsilon_n(\theta) = S_n(\rho_1)Y_n - X_n\beta$ for any possible value θ. Let $\psi = (\theta^\tau, \sigma^2)^\tau$ and use $\psi_0 = (\theta_0^\tau, \sigma_0^2)^\tau$ to denote the true values of ψ. Let

$$P_{1n} = G_n^{(t)}, P_{2n} = D(G_n^{(t)}), P_{3n} = (D(G_n X_n \beta_0))^{(t)}$$
$$P_{j+3,n} = (D(X_{nj}^*))^{(t)}, j = 1, 2, \ldots, k^*, Q_n = (Q_{1n}, Q_{2n}, Q_{3n}, Q_{4n})$$

with

$$Q_{1n} = X_n^*, Q_{2n} = G_n X_n \beta_0, Q_{3n} = 1_n, Q_{4n} = Vec_D(G_n^{(t)})$$

The above $\{P_{sn}, 1 \le s \le k^* + 3\}$ and Q_n provide the set of best IVs and weighting matrices. Following the procedure in Section 2.1, the resulting EL estimator $\hat{\theta}_n$ of θ_0 has the same asymptotic variance $\Sigma_{B\rho_1}$ as the BGMME, where $\Sigma_{B\rho_1}$ is given in (6) in Liu, Lee, and Bollinger (2010).

In practice, we first give the $\sqrt{n}$-consistent estimators $\hat{\theta}_n$ of θ_0 to obtain the estimated IVs and weighting matrices: $\hat{P}_{in} = P_{in}(\hat{\theta}_n), i = 1, 2, \ldots, k^* + 3, \hat{Q}_{jn} = Q_{jn}(\hat{\theta}_n), j = 1, 2, 3, 4$. Then following the procedure in Section 2.1, we can construct the likelihood function $\hat{L}_n(\psi)$ and obtain the EL estimator $\hat{\theta}_{bn}$ of θ_0, which has the same limiting distribution as the BGMME.

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

$$\sqrt{n}(\hat{\theta}_{bn} - \theta_0) \xrightarrow{d} N(0, \Sigma_{B\rho_1})$$

where $\Sigma_{B\rho_1}$ is given in (6) in Liu, Lee, and Bollinger (2010).

Following the procedure in Section 2.1, we let $\ell_n(\theta_0) = 2\log\{\hat{L}_n \quad (\hat{\psi}_{bn})\} -2\log\{\hat{L}_n(\theta_0, \hat{\sigma}^2_{bn})\}$.

Theorem 3. *Suppose that Assumptions (A1)–(A6) are satisfied. Then as $n \to \infty$,*

$$\ell_n(\theta_0) \xrightarrow{d} \chi^2_{k+1}$$

where χ^2_{k+1} is a chi-squared distributed random variable with $k+1$ degrees of freedom.

Table 1. The mean, (SD) and [RMSE] for the SE model.

$n = 98$	$\rho_{20} = 0.3$ Normal	$\beta_1 = 1$	$\beta_2 = -1$
EL	0.317(0.132)[0.133]	0.998(0.149)[0.149]	−0.999(0.152)[0.151]
BGMM	0.329(0.143)[0.146]	0.997(0.151)[0.151]	−0.999(0.153)[0.153]
$n = 490$			
EL	0.302(0.048)[0.039]	1.000(0.054)[0.054]	−0.998(0.056)[0.057]
BGMM	0.305(0.056)[0.056]	1.000(0.064)[0.064]	−0.997(0.064)[0.064]
$n = 98$	Gamma		
EL	0.320(0.127)[0.130]	1.001(0.106)[0.106]	−1.005(0.109)[0.112]
BGMM	0.331(0.138)[0.141]	1.003(0.113)[0.113]	−1.005(0.115)[0.115]
$n = 490$			
EL	0.305(0.053)[0.054]	0.998(0.051)[0.051]	−1.002(0.050)[0.050]
BGMM	0.307(0.055)[0.056]	0.998(0.049)[0.049]	−1.001(0.049)[0.049]

From Theorem 3, the EL-based confidence region for θ with asymptotically correct coverage probability α can be constructed as

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

3. Simulations

To compare the performance of the proposed EL estimators in this article and the BGMME, we use the same model as in Liu, Lee, and Bollinger (2010):

$$Y_n = \rho_{10} W_n Y_n + X_{n1}\beta_{10} + X_{n2}\beta_{20} + u_{(n)}, u_{(n)} = \rho_{20} M_n u_{(n)} + \epsilon_{(n)}$$

where $\beta_{10} = 1, \beta_{20} = -1$ $X_{nj} \sim N(0, I_n), j = 1, 2,$ and X_{n1} and X_{n2} are mutually independent. ϵ_{ni} are independently drawn from the following two populations: (a) $\epsilon_{ni} \sim N(0, 2)$; (b) $\epsilon_{ni} \sim Gamma(2, 1) - 2$. Let W_A be the weight matrix from the study of crimes across 49 districts in Columbus, Ohio in Anselin (1988). Then we let $W_n = M_n$ and W_n be the two weight matrices: (a) $W_{98} = I_2 \otimes W_{49}$; (b) $W_{490} = I_{10} \otimes W_{49}$, where $\otimes$ is the Kronecker product.

In the simulations, the number of repetitions is 1,000 for each case. We report the mean, standard deviation (SD) and root mean square errors (RMSE) of the 1, 000 EL estimators $\hat{\theta}_{bn}$, where the initial estimators of θ_0 are the same as in Liu, Lee, and Bollinger (2010). The simulation results for BGMME done by Liu, Lee, and Bollinger (2010) are also listed here for comparison purpose. In addition, we also report the coverage probabilities (CP) of BGMME and EL based confidence intervals with a confidence level $\alpha = 0.95$. We have also done the simulations where the regressors are non-random with similar results to those reported here.

Tables 1–4 report the simulation results of mean, SD and RMSE for SE model ($\rho_{10} = 0, \rho_{20} = 0.3$), SAR model ($\rho_{10} = 0.3, \rho_{20} = 0$), SARAR model ($\rho_{10} = 0.3$, $\rho_{20} = 0.3$) and SARAR model ($\rho_{10} = 0.8, \rho_{20} = 0.85$), respectively. From these results, we can see that both the EL and BGMME estimators perform well. In addition, with a larger sample size $n = 490$, both estimators perform similarly. However, for the smaller sample size $n = 49$, the EL estimator outperforms the GMME.

Tables 5–8 report the simulation results of CP for SE model ($\rho_{10} = 0, \rho_{20} = 0.3$), SAR model ($\rho_{10} = 0.3, \rho_{20} = 0$), SARAR model ($\rho_{10} = 0.3, \rho_{20} = 0.3$) and SARAR

Table 2. The mean, (SD) and [RMSE] for the SAR model.

$n = 98$	$\rho_{10} = 0.3$ Normal	$\beta_1 = 1$	$\beta_2 = -1$
EL	0.311(0.107)[0.108]	0.988(0.147)[0.149]	−0.993(0.150)[0.151]
BGMM	0.320(0.117)[0.119]	0.987(0.150)[0.151]	−0.991(0.154)[0.155]
$n = 490$			
EL	0.302(0.048)[0.048]	0.997(0.066)[0.064]	−0.995(0.065)[0.063]
BGMM	0.301(0.047)[0.047]	0.997(0.065)[0.065]	−0.994(0.064)[0.065]
$n = 98$	Gamma		
EL	0.309(0.101)[0.102]	0.998(0.115)[0.115]	−0.999(0.114)[0.114]
BGMM	0.319(0.102)[0.104]	0.996(0.114)[0.114]	−0.999(0.115)[0.115]
$n = 490$			
EL	0.305(0.039)[0.040]	0.997(0.049)[0.050]	−1.000(0.045)[0.045]
BGMM	0.305(0.041)[0.041]	0.997(0.050)[0.050]	−1.000(0.050)[0.050]

Table 3. The mean, (SD) and [RMSE] for the SARAR model.

$n = 98$	$\rho_{10} = 0.3$ Normal	$\rho_{20} = 0.3$	$\beta_1 = 1$	$\beta_2 = -1$
EL	0.270(0.217)[0.211]	0.315(0.316)[0.315]	0.978(0.159)[0.159]	−0.976(0.156)[0.158]
BGMM	0.243(0.309)[0.315]	0.318(0.324)[0.324]	0.976(0.161)[0.163]	−0.974(0.162)[0.164]
$n = 490$				
EL	0.286(0.099)[0.099]	0.305(0.106)[0.108]	0.997(0.063)[0.066]	−0.995(0.065)[0.067]
BGMM	0.287(0.098)[0.099]	0.306(0.109)[0.110]	0.997(0.064)[0.064]	−0.994(0.064)[0.065]
$n = 98$	Gamma			
EL	0.282(0.265)[0.268]	0.311(0.295)[0.296]	0.985(0.131)[0.131]	−0.986(0.132)[0.133]
BGMM	0.251(0.295)[0.299]	0.315(0.301)[0.301]	0.984(0.130)[0.131]	−0.986(0.130)[0.131]
$n = 490$				
EL	0.299(0.068)[0.068]	0.300(0.083)[0.083]	0.997(0.049)[0.048]	−0.988(0.050)[0.050]
BGMM	0.299(0.069)[0.069]	0.299(0.087)[0.087]	0.997(0.050)[0.050]	−1.000(0.049)[0.049]

Table 4. The mean, (SD) and [RMSE] for the SARAR model (continued).

$n = 98$	$\rho_{10} = 0.8$ Normal	$\rho_{20} = 0.85$	$\beta_1 = 1$	$\beta_2 = -1$
EL	0.730(0.326)[0.328]	0.812(0.321)[0.326]	0.952(0.162)[0.163]	−0.913(0.159)[0.161]
BGMM	0.731(0.341)[0.3343]	0.812(0.330)[0.332]	0.951(0.163)[0.164]	−0.915(0.163)[0.164]
$n = 490$				
EL	0.743(0.099)[0.099]	0.825(0.125)[0.127]	0.961(0.075)[0.076]	−0.934(0.072)[0.073]
BGMM	0.744(0.098)[0.099]	0.825(0.128)[0.130]	0.962(0.076)[0.076]	−0.935(0.074)[0.075]
$n = 98$	Gamma			
EL	0.710(0.266)[0.267]	0.816(0.312)[0.314]	0.903(0.165)[0.166]	−0.913(0.158)[0.160]
BGMM	0.684(0.315)[0.317]	0.754(0.325)[0.327]	0.871(0.232)[0.234]	−0.887(0.205)[0.206]
$n = 490$				
EL	0.760(0.123)[0.124]	0.831(0.102)[0.103]	0.926(0.065)[0.066]	−0.927(0.108)[0.110]
BGMM	0.712(0.155)[0.156]	0.752(0.165)[0.165]	0.882(0.162)[0.163]	−0.870(0.136)[0.137]

model ($\rho_{10} = 0.8, \rho_{20} = 0.85$), respectively. From these results, we can see that both the EL and BGMME confidence intervals perform well in terms of CP with a larger sample size $n = 490$. However, for the smaller sample size $n = 49$, the EL confidence intervals outperform those of the GMME.

In summary, the EL method is competitive in the statistical inferences for spatial models.

Table 5. The CP for the SE model.

$n=98$	$\rho_{20}=0.3$ Normal	$\beta_1=1$	$\beta_2=-1$
EL	0.82	0.84	0.85
BGMM	0.81	0.85	0.85
$n=490$			
EL	0.91	0.92	0.91
BGMM	0.92	0.92	0.91
$n=98$	Gamma		
EL	0.85	0.87	0.88
BGMM	0.80	0.83	0.84
$n=490$			
EL	0.91	0.90	0.92
BGMM	0.91	0.88	0.91

Table 6. The CP for the SAR model.

$n=98$	$\rho_{10}=0.3$ Normal	$\beta_1=1$	$\beta_2=-1$
EL	0.83	0.84	0.83
BGMM	0.84	0.82	0.83
$n=490$			
EL	0.90	0.89	0.91
BGMM	0.91	0.90	0.90
$n=98$	Gamma		
EL	0.83	0.84	0.83
BGMM	0.78	0.76	0.77
$n=490$			
EL	0.90	0.91	0.90
BGMM	0.89	0.91	0.91

Table 7. The CP for the SARAR model.

$n=98$	$\rho_{10}=0.3$ Normal	$\rho_{20}=0.3$	$\beta_1=1$	$\beta_2=-1$
EL	0.81	0.82	0.82	0.83
BGMM	0.74	0.73	0.75	0.72
$n=490$				
EL	0.90	0.91	0.89	0.89
BGMM	0.91	0.90	0.90	0.89
$n=98$	Gamma			
EL	0.80	0.82	0.82	0.83
BGMM	0.76	0.78	0.77	0.78
$n=490$				
EL	0.88	0.90	0.87	0.87
BGMM	0.87	0.90	0.89	0.86

4. Proofs

In the sequel, we will use $||a||$ to denote the L_2-norm of a vector a. As the proofs of the results in Section 2.1 are more involved than other results, we only give the proofs of Propositions 1 and 2 and Theorem 1.

Table 8. The CP for the SARAR model (continued).

$n = 98$	$\rho_{10} = 0.8$ Normal	$\rho_{20} = 0.85$	$\beta_1 = 1$	$\beta_2 = -1$
EL	0.81	0.82	0.83	0.83
BGMM	0.80	0.81	0.84	0.83
$n = 490$				
EL	0.86	0.87	0.86	0.86
BGMM	0.87	0.86	0.85	0.86
$n = 98$	Gamma			
EL	0.76	0.74	0.78	0.78
BGMM	0.72	0.71	0.72	0.71
$n = 490$				
EL	0.84	0.83	0.84	0.84
BGMM	0.82	0.81	0.83	0.83

We need to use Theorem 1 in Kelejian and Prucha (2001). We now state this result. Let

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

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

(C1) $\{\epsilon_{ni}, 1 \le i \le n\}$ 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 $\eta_2 > 0$.

Denote

$$\mu_{\tilde{Q}} = E(\tilde{Q}), \sigma_{\tilde{Q}}^2 = var(\tilde{Q})$$

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

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

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

Lemma 2. *Under the conditions of Proposition 1, as $n \to \infty$,*

$$n^{-1/2}\sum_{i=1}^{n}\omega_i(\psi_0)\xrightarrow{d}N(0,\Omega) \tag{4.1}$$

$$n^{-1}\sum_{i=1}^{n}\omega_i(\psi_0)\omega_i^{\tau}(\psi_0) = \Omega + o_p(1) \tag{4.2}$$

where $\Omega = \lim_{n\to\infty}\left(\frac{1}{n}\Omega_n\right)$ and $\Omega_n = var(g_n(\theta_0))$ is given in Proposition 1.

Proof. Since $\sum_{i=1}^n \omega_i(\psi_0) = g_n(\theta_0)$ under $\mathcal{P}_1$, (4.1) thus can be proved by applying Lemma 1. It remains to prove (4.2), i. e. for any $l = (l_1^\tau, l_2^\tau)^\tau \in R^{m+q}$,

$$n^{-1}\sum_{i=1}^n \{l^\tau \omega_i(\theta_0)\}^2 = n^{-1}\sigma_Q^2 + o_p(1) \tag{4.3}$$

where $\sigma_Q^2 = var(\sum_{i=1}^n l^\tau \omega_i(\theta_0))$. Let

$$\begin{aligned} Y_{in} = l^\tau \omega_i(\theta_0) &= u_{ii}(\epsilon_{ni}^2 - \sigma_0^2) + 2\sum_{j=1}^{i-1} u_{ij}\epsilon_{ni}\epsilon_{nj} + v_i\epsilon_{ni} \\ &= u_{ii}(\epsilon_{ni}^2 - \sigma_0^2) + B_i\epsilon_{ni} \end{aligned} \tag{4.4}$$

where $u_{ii} = l_1^\tau(\tilde{P}_{ii1n}, \ldots, \tilde{P}_{iimn})^\tau, u_{ij} = l_1^\tau(\tilde{P}_{ij1n}, \ldots, \tilde{P}_{ijmn})^\tau (i \neq j), v_i = l_2^\tau b_i, B_i = 2\sum_{j=1}^{i-1} u_{ij}$ $\epsilon_{nj} + v_i$. Let $\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_i = \sigma(\epsilon_{n1}, \epsilon_{n2}, \ldots, \epsilon_{ni}), 1 \le i \le n$. Then $\{Y_{in}, \mathcal{F}_i, 1 \le i \le n\}$ form a martingale difference array. Note that

$$\begin{aligned} &n^{-1}\sum_{i=1}^n \{l^\tau \omega_i(\theta_0)\}^2 - n^{-1}\sigma_Q^2 = n^{-1}\sum_{i=1}^n (Y_{in}^2 - EY_{in}^2) \\ &= n^{-1}\sum_{i=1}^n \{Y_{in}^2 - E(Y_{in}^2|\mathcal{F}_{i-1}) + E(Y_{in}^2|\mathcal{F}_{i-1}) - EY_{in}^2\} \\ &= n^{-1}S_{n1} + n^{-1}S_{n2} \end{aligned} \tag{4.5}$$

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

$$n^{-1}S_{n1} = o_p(1) \tag{4.6}$$

and

$$n^{-1}S_{n2} = o_p(1) \tag{4.7}$$

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

$$Y_{in}^2 = u_{ii}^2(\epsilon_{ni}^2 - \sigma_0^2)^2 + B_i^2\epsilon_{ni}^2 + 2u_{ii}B_i(\epsilon_{ni}^2 - \sigma_0^2)\epsilon_{ni}$$

Thus

$$E(Y_{in}^2|\mathcal{F}_{i-1}) = u_{ii}^2 E(\epsilon_{ni}^2 - \sigma_0^2)^2 + B_i^2\sigma_0^2 + 2u_{ii}B_i\mu_3$$

It follows that

$$\begin{aligned} n^{-2}ES_{n1}^2 &= n^{-2}\sum_{i=1}^n E\{Y_{in}^2 - E(Y_{in}^2|\mathcal{F}_{i-1})\}^2 \\ &= n^{-2}\sum_{i=1}^n E[u_{ii}^2\{(\epsilon_{ni}^2 - \sigma_0^2)^2 - E(\epsilon_{ni}^2 - \sigma_0^2)^2\} + B_i^2(\epsilon_{ni}^2 - \sigma_0^2) \\ &\quad + 2u_{ii}B_i(\epsilon_i^3 - \sigma_0^2\epsilon_{ni} - \mu_3)]^2 \\ &\le Cn^{-2}\sum_{i=1}^n E[u_{ii}^4\{(\epsilon_{ni}^2 - \sigma_0^2)^2 - E(\epsilon_{ni}^2 - \sigma_0^2)^2\}^2] + Cn^{-2}\sum_{i=1}^n E\{B_i^4(\epsilon_{ni}^2 - \sigma_0^2)^2\} \\ &\quad + Cn^{-2}\sum_{i=1}^n E\{u_{ii}^2B_i^2(\epsilon_i^3 - \sigma_0^2\epsilon_{ni} - \mu_3)^2\} \end{aligned} \tag{4.8}$$

By Condition A4, we have

$$n^{-2}\sum_{i=1}^{n}E\Big[u_{ii}^4\{(\epsilon_{ni}^2-\sigma_0^2)^2-E(\epsilon_{ni}^2-\sigma_0^2)^2\}^2\Big]\le Cn^{-1}\to 0 \tag{4.9}$$

and

$$\begin{aligned}
n^{-2}\sum_{i=1}^{n}E\{B_i^4(\epsilon_{ni}^2-\sigma_0^2)^2\} &\le Cn^{-2}\sum_{i=1}^{n}E\left(\sum_{j=1}^{i-1}u_{ij}\epsilon_{nj}+\nu_i\right)^4\\
&\le Cn^{-2}\sum_{i=1}^{n}E\left(\sum_{j=1}^{i-1}u_{ij}\epsilon_{nj}\right)^4+Cn^{-2}\sum_{i=1}^{n}\nu_i^4\\
&\le Cn^{-2}\sum_{i=1}^{n}\sum_{j=1}^{i-1}u_{ij}^4\mu_4+Cn^{-2}\sum_{i=1}^{n}\left(\sum_{j=1}^{i-1}u_{ij}^2\sigma_0^2\right)^2+Cn^{-2}\sum_{i=1}^{n}(l_1^{\tau}b_i+l_2b_i)^4\\
&\le Cn^{-1}\to 0
\end{aligned}\tag{4.10}$$

Similarly, one can show that

$$n^{-2}\sum_{i=1}^{n}E\{u_{ii}^2B_i^2(\epsilon_i^3-\sigma_0^2\epsilon_{ni}-\mu_3)^2\}\to 0 \tag{4.11}$$

From Equations (4.8) to (4.11), we have $n^{-2}ES_{n1}^2\to 0$. Furthermore,

$$\begin{aligned}
EY_{in}^2=E\{E(Y_{in}^2|\mathcal{F}_{i-1})\}&=u_{ii}^2E(\epsilon_{ni}^2-\sigma_0^2)^2+\sigma_0^2E(B_i^2)+2u_{ii}\mu_3E(B_i)\\
&=u_{ii}^2E(\epsilon_{ni}^2-\sigma_0^2)^2+\sigma_0^2\left(4\sum_{j=1}^{i-1}u_{ij}^2\sigma_0^2+\nu_i^2\right)+2u_{ii}\mu_3\nu_i
\end{aligned}$$

Thus,

$$\begin{aligned}
n^{-2}ES_{n2}^2&=n^{-2}E\left[\sum_{i=1}^{n}\{E(Y_{in}^2|\mathcal{F}_{i-1})-EY_{in}^2\}\right]^2\\
&=n^{-2}E\left[\sum_{i=1}^{n}\left\{B_i^2\sigma_0^2-\sigma_0^2\left(4\sum_{j=1}^{i-1}u_{ij}^2\sigma_0^2+\nu_i^2\right)+2u_{ii}\mu_3(B_i-\nu_i)\right\}\right]^2\\
&=n^{-2}\sum_{i=1}^{n}E\left[\sigma_0^2\left\{\left(2\sum_{j=1}^{i-1}u_{ij}\epsilon_{nj}\right)^2-4\sum_{j=1}^{i-1}u_{ij}^2\sigma_0^2\right\}+4\left(\sum_{j=1}^{i-1}u_{ij}\epsilon_{nj}\right)\nu_i\sigma_0^2\right.\\
&\quad\left.+2u_{ii}\mu_3\left(2\sum_{j=1}^{i-1}u_{ij}\epsilon_{nj}\right)\right]^2\\
&\le Cn^{-2}\sum_{i=1}^{n}E\left\{\sigma_0^2\left(\sum_{j=1}^{i-1}u_{ij}\epsilon_{nj}\right)^2-\sum_{j=1}^{i-1}u_{ij}^2\sigma_0^2\right\}^2+Cn^{-2}\sum_{i=1}^{n}E\left\{\left(\sum_{j=1}^{i-1}u_{ij}\epsilon_{nj}\right)\nu_i\sigma_0^2\right\}^2\\
&\quad+Cn^{-2}\sum_{i=1}^{n}E\left\{2u_{ii}\mu_3\left(\sum_{j=1}^{i-1}u_{ij}\epsilon_{nj}\right)\right\}^2
\end{aligned}\tag{4.12}$$

Note that

$$\begin{aligned}
n^{-2}\sum_{i=1}^{n}E\left[\sigma_0^2\left\{\left(\sum_{j=1}^{i-1}u_{ij}\epsilon_{nj}\right)^2-\sum_{j=1}^{i-1}u_{ij}^2\sigma_0^2\right\}\right]^2&\le n^{-2}\sigma_0^4\sum_{i=1}^{n}E\left(\sum_{j=1}^{i-1}u_{ij}\epsilon_{nj}\right)^4\\
&\le Cn^{-2}\sum_{i=1}^{n}\sum_{j=1}^{i-1}u_{ij}^4\mu_4+Cn^{-2}\sum_{i=1}^{n}\left(\sum_{j=1}^{i-1}u_{ij}^2\sigma_0^2\right)^2\le Cn^{-1}\to 0
\end{aligned}\tag{4.13}$$

$$n^{-2}\sum_{i=1}^{n} E\left\{\left(\sum_{j=1}^{i-1} u_{ij}\epsilon_{nj}\right)v_i\sigma_0^2\right\}^2 = n^{-2}\sigma_0^6\sum_{i=1}^{n} v_i^2\sum_{j=1}^{i-1} u_{ij}^2 \le Cn^{-2} \to 0 \tag{4.14}$$

and

$$n^{-2}\sum_{i=1}^{n} E\left\{2u_{ii}\mu_3\left(\sum_{j=1}^{i-1} u_{ij}\epsilon_{nj}\right)\right\}^2 = 4\mu_3^2\sigma_0^2 n^{-2}\sum_{i=1}^{n} u_{ii}^2\sum_{j=1}^{i-1} u_{ij}^2 \le Cn^{-1} \to 0 \tag{4.15}$$

From (4.12) to (4.15), we have $n^{-2}ES_{n2}^2 \to 0$. The proof of Equation (4.3) is thus complete.

As a consequence of Lemma 2 and the proof of Lemma 1 in Qin and Lawless (1994), we have the following result for the existence of a local maximizer of $L_n(\psi)$:

Lemma 3. *Under the conditions of Proposition 1, as $n \to \infty$, with probability tending to 1 the likelihood equations (2.9) and (2.10) have a solution $\hat{\psi}_n$ within the open ball $||\hat{\psi}_n - \psi_0|| < Cn^{-1/3}$, and $L_n(\psi)$ attains its local maximum at $\hat{\psi}_n$.*

We now prove Propositions 1 and 2 and Theorem 1.

Proof of Proposition 1. Taking derivatives about ψ and λ^τ, we have

$$\frac{\partial U_{1n}(\psi, 0)}{\partial \psi} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial \omega_i(\psi)}{\partial \psi} = \left(\frac{1}{n}\frac{\partial g_n(\theta)}{\partial \theta}, 0\right)$$

$$\frac{\partial U_{1n}(\psi, 0)}{\partial \lambda^\tau} = -\frac{1}{n}\sum_{i=1}^{n}\omega_i(\psi)\omega_i^\tau(\psi)$$

$$\frac{\partial U_{2n}(\psi, 0)}{\partial \psi} = 0, \frac{\partial U_{2n}(\psi, 0)}{\partial \lambda^\tau} = \left(\frac{1}{n}\frac{\partial g_n(\theta)}{\partial \theta}, 0\right)^\tau$$

Therefore, from Lemma 3 and Taylor expansion, we have

$$\begin{aligned}
0 &= U_{1n}(\hat{\psi}_n, \hat{\lambda}) \\
&= U_{1n}(\psi_0, 0) + \frac{\partial U_{1n}(\psi_0, 0)}{\partial \psi}(\hat{\psi}_n - \psi_0) + \frac{\partial U_{1n}(\psi_0, 0)}{\partial \lambda^\tau}\hat{\lambda} + O_p(||\hat{\psi}_n - \psi_0||^2) \\
&= \frac{1}{n}g_n(\theta_0) + \frac{1}{n}\frac{\partial g_n(\theta)}{\partial \theta}\bigg|_{\theta=\theta_0}(\hat{\theta}_n - \theta_0) - \frac{1}{n}\sum_{i=1}^{n}\omega_i(\psi_0)\omega_i^\tau(\psi_0)\hat{\lambda} + o_p(n^{-1/2}) \\
0 &= U_{2n}(\hat{\psi}_n, \hat{\lambda}) \\
&= U_{2n}(\psi_0, 0) + \frac{\partial U_{2n}(\psi_0, 0)}{\partial \psi}(\hat{\psi}_n - \psi_0) + \frac{\partial U_{2n}(\psi_0, 0)}{\partial \lambda^\tau}\hat{\lambda} + O_p(||\hat{\psi}_n - \psi_0||^2) \\
&= \left(\frac{1}{n}\frac{\partial g_n(\theta)}{\partial \theta}\bigg|_{\theta=\theta_0}, 0\right)^\tau\hat{\lambda} + o_p(n^{-1/2})
\end{aligned}$$

It follows that

$$S_n\begin{pmatrix}\hat{\lambda} \\ \hat{\theta}_n - \theta_0\end{pmatrix} = \begin{pmatrix}-\frac{1}{n}g_n(\theta_0) + o_p(n^{-1/2}) \\ o_p(n^{-1/2})\end{pmatrix}$$

where

$$S_n = \begin{pmatrix} -\frac{1}{n}\sum_{i=1}^{n}\omega_i(\psi_0)\omega_i^\tau(\psi_0) & \frac{1}{n}\frac{\partial g_n(\theta)}{\partial\theta}\Big|_{\theta=\theta_0} \\ \left(\frac{1}{n}\frac{\partial g_n(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\right)^\tau & 0 \end{pmatrix} \hat{=} \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & 0 \end{pmatrix}$$

Therefore,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = S_{22.1}^{-1} S_{21} S_{11}^{-1} \cdot \frac{1}{\sqrt{n}} g_n(\theta_0) + o_p(1) \tag{4.16}$$

It has been shown in the proof of Proposition 1 in Liu, Lee, and Bollinger (2010)), that

$$\frac{1}{n}\frac{\partial g_n(\theta)}{\partial\theta}\Big|_{\theta=\theta_0} = -\frac{1}{n}D_n + o_p(1) \tag{4.17}$$

with D_n in Proposition 1. From (4.16), (4.17) and Lemma 2, we have Proposition 1.

Proof of Proposition 2. Based on the set of best IVs and weighting matrices $\{P_{sn}, 1 \leq s \leq k^* + 5\}$ and Q_n, and the estimated IVs and weighting matrices, define

$$g_{bn}(\theta) = \left(\epsilon_n^\tau(\theta)Q_n, \epsilon_n^\tau(\theta)\tilde{P}_{1n}\epsilon_n(\theta), \epsilon_n^\tau(\theta)\tilde{P}_{2n}\epsilon_n(\theta), \ldots, \epsilon_n^\tau(\theta)\tilde{P}_{k^*+5,n}\epsilon_n(\theta)\right)^\tau$$
$$\hat{g}_{bn}(\theta) = \left(\epsilon_n^\tau(\theta)\hat{Q}_n, \epsilon_n^\tau(\theta)\hat{\tilde{P}}_{1n}\epsilon_n(\theta), \epsilon_n^\tau(\theta)\hat{\tilde{P}}_{2n}\epsilon_n(\theta), \ldots, \epsilon_n^\tau(\theta)\hat{\tilde{P}}_{k^*+5,n}\epsilon_n(\theta)\right)^\tau$$

$$\omega_{bi}(\psi) = \begin{pmatrix} b_i\epsilon_{ni}(\theta) \\ \tilde{P}_{ii1n}\{\epsilon_{ni}^2(\theta) - \sigma^2\} + 2\epsilon_{ni}(\theta)\sum_{j=1}^{i-1}\tilde{P}_{ij1n}\epsilon_{nj}(\theta) \\ \tilde{P}_{ii2n}\{\epsilon_{ni}^2(\theta) - \sigma^2\} + 2\epsilon_{ni}(\theta)\sum_{j=1}^{i-1}\tilde{P}_{ij2n}\epsilon_{nj}(\theta) \\ \vdots \\ \tilde{P}_{ii,k^*+5,n}\{\epsilon_{ni}^2(\theta) - \sigma^2\} + 2\epsilon_{ni}(\theta)\sum_{j=1}^{i-1}\tilde{P}_{ij,k^*+5,n}\epsilon_{nj}(\theta) \end{pmatrix}$$

$$\hat{\omega}_{bi}(\psi) = \begin{pmatrix} \hat{b}_i\epsilon_{ni}(\theta) \\ \hat{\tilde{P}}_{ii1n}\{\epsilon_{ni}^2(\theta) - \sigma^2\} + 2\epsilon_{ni}(\theta)\sum_{j=1}^{i-1}\hat{\tilde{P}}_{ij1n}\epsilon_{nj}(\theta) \\ \hat{\tilde{P}}_{ii2n}\{\epsilon_{ni}^2(\theta) - \sigma^2\} + 2\epsilon_{ni}(\theta)\sum_{j=1}^{i-1}\hat{\tilde{P}}_{ij2n}\epsilon_{nj}(\theta) \\ \vdots \\ \hat{\tilde{P}}_{ii,k^*+5,n}\{\epsilon_{ni}^2(\theta) - \sigma^2\} + 2\epsilon_{ni}(\theta)\sum_{j=1}^{i-1}\hat{\tilde{P}}_{ij,k^*+5,n}\epsilon_{nj}(\theta) \end{pmatrix}$$

where $\tilde{P}_{ijsn}$ and b_i^τ are the (i, j) element of the matrix $\tilde{P}_{sn}$ and the ith row of the matrix Q_n, respectively. $\hat{\tilde{P}}_{ij1n}$ and $\hat{b}_i^\tau$ are defined similarly. It is shown in the proof of Proposition 2 in Liu, Lee, and Bollinger (2010) that

$$\frac{1}{\sqrt{n}}\{\hat{g}_{bn}(\theta_0) - g_{bn}(\theta_0)\} = o_p(1)$$

Noting that $\sum_{i=1}^n \hat{\omega}_{bi}(\psi_0) = \hat{g}_{bn}(\theta_0)$ and $\sum_{i=1}^n \omega_{bi}(\psi_0) = g_{bn}(\theta_0)$, and combing with Lemma 2, we have

$$n^{-1/2}\sum_{i=1}^n \hat{\omega}_{bi}(\psi_0)\xrightarrow{d}N(0,\Omega) \tag{4.18}$$

where $\Omega = \lim_{n\to\infty}\left(\frac{1}{n}\Omega_n\right)$ and $\Omega_n = var(g_{bn}(\theta_0))$. It is also proved in the proof of Proposition 1 in Liu, Lee, and Bollinger (2010), that

$$\frac{1}{n}\frac{\partial \hat{g}_{bn}(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} = \frac{1}{n}\frac{\partial g_{bn}(\theta)}{\partial\theta}\bigg|_{\theta=\theta_0} + o_p(1) \tag{4.19}$$

Furthermore, following the proof of Equation (4.2), it can be shown that

$$n^{-1}\sum_{i=1}^n \hat{\omega}_{bi}(\psi_0)\hat{\omega}_{bi}^\tau(\psi_0) = \Omega + o_p(1) \tag{4.20}$$

From Equations (4.18) to (4.20) and the proof of Proposition 1, we can see that Proposition 2 holds true.

Proof of **Theorem 1**. We use the notations in the proof of Proposition 2 and let

$$\hat{U}_{1n}(\psi,\lambda)\hat{=}\frac{1}{n}\sum_{i=1}^n\frac{\hat{\omega}_{bi}(\psi)}{1+\lambda^\tau(\psi)\hat{\omega}_{bi}(\psi)}$$

$$\hat{U}_{2n}(\psi,\lambda)\hat{=}\frac{1}{n}\sum_{i=1}^n\frac{1}{1+\lambda^\tau\hat{\omega}_{bi}(\psi)}\left(\frac{\partial\hat{\omega}_{bi}(\psi)}{\partial\psi}\right)^\tau\lambda$$

Then $\hat{U}_{jn}(\hat{\psi}_{bn},\hat{\lambda}_b) = 0, j = 1,2,$ where $\hat{\lambda}_b = \lambda(\hat{\psi}_{bn})$. Denote the first $k+2$ components of $\hat{\psi}_{bn}$ as $\hat{\theta}_{bn}$. Following the proof of Proposition 1, it can be shown that

$$\hat{S}_n\begin{pmatrix}\hat{\lambda}_b\\ \hat{\theta}_{bn}-\theta_0\end{pmatrix} = \begin{pmatrix}-\frac{1}{n}\hat{g}_{bn}(\theta_0)+o_p(n^{-1/2})\\ o_p(n^{-1/2})\end{pmatrix}$$

where

$$\hat{S}_n = \begin{pmatrix}-\frac{1}{n}\sum_{i=1}^n\hat{\omega}_{bi}(\psi_0)\hat{\omega}_{bi}^\tau(\psi_0) & \frac{1}{n}\frac{\partial\hat{g}_{bn}(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\\ \left(\frac{1}{n}\frac{\partial\hat{g}_{bn}(\theta)}{\partial\theta}\Big|_{\theta=\theta_0}\right)^\tau & 0\end{pmatrix}\hat{=}\begin{pmatrix}\hat{S}_{11} & \hat{S}_{12}\\ \hat{S}_{21} & 0\end{pmatrix}$$

Therefore,

$$\hat{\lambda}_b = \hat{S}_{11}^{-1}(I+\hat{S}_{12}\hat{S}_{22.1}^{-1}\hat{S}_{21}\hat{S}_{11}^{-1})\cdot\frac{1}{n}\hat{g}_{bn}(\theta_0)+o_p(n^{-1/2}) \tag{4.21}$$

where $\hat{S}_{22.1} = -\hat{S}_{21}\hat{S}_{11}^{-1}\hat{S}_{12}$. Using $\hat{U}_{1n}(\hat{\psi}_{bn},\hat{\lambda}_b) = 0$ and following the usual arguments in EL approach, one can show that

$$\sum_{i=1}^{n}\hat{\omega}_{bi}(\hat{\psi}_{bn})=\left\{\sum_{i=1}^{n}\hat{\omega}_{bi}(\hat{\psi}_{bn})\hat{\omega}_{bi}^{\tau}(\hat{\psi}_{bn})\right\}\hat{\lambda}_b+o_p(n^{1/2}) \tag{4.22}$$

Using Taylor expansion, Equations (4.21) and (4.22), we have

$$\begin{aligned}
&2\sum_{i=1}^{n}\log\{1+\hat{\lambda}_b^{\tau}\hat{\omega}_{bi}(\hat{\psi}_{bn})\}\\
&=2\sum_{i=1}^{n}\hat{\lambda}_b^{\tau}\hat{\omega}_{bi}(\hat{\psi}_{bn})-\sum_{i=1}^{n}\{\hat{\lambda}_b^{\tau}\hat{\omega}_{bi}(\hat{\psi}_{bn})\}^2+o_p(1)\\
&=\hat{\lambda}_b^{\tau}\left\{\sum_{i=1}^{n}\hat{\omega}_{bi}(\hat{\psi}_{bn})\hat{\omega}_{bi}^{\tau}(\hat{\psi}_{bn})\right\}\hat{\lambda}_b+o_p(1)\\
&=\hat{\lambda}_b^{\tau}\left\{\sum_{i=1}^{n}\hat{\omega}_{bi}(\psi_0)\hat{\omega}_{bi}^{\tau}(\psi_0)\right\}\hat{\lambda}_b+o_p(1)\\
&=-n\{\tfrac{1}{n}\hat{g}_{bn}(\theta_0)\}^{\tau}\hat{S}_{11}^{-1}(I+\hat{S}_{12}\hat{S}_{22.1}^{-1}\hat{S}_{21}\hat{S}_{11}^{-1})\cdot\frac{1}{n}\hat{g}_{bn}(\theta_0)\\
&\quad+o_p(1)
\end{aligned} \tag{4.23}$$

Furthermore, let

$$\hat{U}_{3n}(\theta_0,\sigma^2,\lambda_1)\hat{=}\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{\omega}_{bi}(\theta_0,\sigma^2)}{1+\lambda_1^{\tau}\hat{\omega}_{bi}(\theta_0,\sigma^2)}$$

$$\hat{U}_{4n}(\theta_0,\sigma^2,\lambda_1)\hat{=}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\lambda^{\tau}\hat{\omega}_{bi}(\theta_0,\sigma^2)}\left(\frac{\partial\hat{\omega}_{bi}(\theta_0,\sigma^2)}{\partial\sigma^2}\right)^{\tau}\lambda_1$$

Then $\hat{U}_{jn}(\theta_0,\hat{\sigma}_{bn}^2,\hat{\lambda}_{b1})=0, j=3,4,$ where $\hat{\lambda}_{b1}=\lambda_1(\theta_0,\hat{\sigma}_{bn}^2)$. Note that

$$\frac{\hat{U}_{3n}(\theta_0,\sigma^2,0)}{\partial\sigma^2}=0,\ \frac{\hat{U}_{3n}(\theta_0,\sigma^2,0)}{\partial\lambda_1^{\tau}}=-\frac{1}{n}\sum_{i=1}^{n}\hat{\omega}_{bi}(\theta_0,\sigma^2)\hat{\omega}_{bi}^{\tau}(\theta_0,\sigma^2)$$

$$\frac{\hat{U}_{4n}(\theta_0,\sigma^2,0)}{\partial\sigma^2}=0,\ \frac{\hat{U}_{4n}(\theta_0,\sigma^2,0)}{\partial\lambda_1^{\tau}}=0$$

Then

$$\hat{\lambda}_{b1}=-\hat{S}_{11}^{-1}\cdot\frac{1}{n}\hat{g}_{bn}(\theta_0)+o_p(n^{-1/2})$$

Similar to the proof of Equation (4.23), we have

$$2\sum_{i=1}^{n}\log\{1+\hat{\lambda}_{b1}^{\tau}\hat{\omega}_{bi}(\theta_0,\hat{\sigma}_{bn}^2)\}=-n\left\{\frac{1}{n}\hat{g}_{bn}(\theta_0)\right\}^{\tau}\hat{S}_{11}^{-1}\cdot\frac{1}{n}\hat{g}_{bn}(\theta_0)+o_p(1) \tag{4.24}$$

Finally, from Equations (4.23) and (4.24), we have

$$\begin{aligned}
\ell_n(\theta_0)&=2\log\{\hat{L}_n(\hat{\psi}_{bn})\}-2\log\{\hat{L}_n(\theta_0,\hat{\sigma}_{bn}^2)\}\\
&=n\{\tfrac{1}{n}\hat{g}_{bn}(\theta_0)\}^{\tau}\hat{S}_{11}^{-1}\hat{S}_{12}\hat{S}_{22.1}\hat{S}_{21}\hat{S}_{11}^{-1}\cdot\frac{1}{n}\hat{g}_{bn}(\theta_0)+o_p(1)\\
&=\{(-\hat{S}_{11})^{-1/2}\tfrac{1}{\sqrt{n}}\hat{g}_{bn}(\theta_0)\}^{\tau}(-\hat{S}_{11})^{-1/2}\hat{S}_{12}\hat{S}_{22.1}^{-1}\hat{S}_{21}(-\hat{S}_{11})^{-1/2}\\
&\quad\times(-\hat{S}_{11})^{-1/2}\frac{1}{\sqrt{n}}\hat{g}_{bn}(\theta_0)+o_p(1)
\end{aligned} \tag{4.25}$$

Equations (4.18) and (4.20) imply that $(-\hat{S}_{11})^{-1/2}\frac{1}{\sqrt{n}}\hat{g}_{bn}(\theta_0)\xrightarrow{d}N(0, I_{2k^*+9})$. Furthermore, $(-\hat{S}_{11})^{-1/2}\hat{S}_{12}\hat{S}_{22.1}^{-1}\hat{S}_{21}(-\hat{S}_{11})^{-1/2}$ is symmetric and idempotent with trace $k+2$. We thus have Theorem 1.

Acknowledgments

The authors are thankful to the referees for constructive suggestions.

Funding

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

References

Anselin, L. 1988. *Spatial econometrics: Methods and models*. The Netherlands: Kluwer Academic Publishers.

Anselin, L., and A. K. Bera. 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. *Handbook of applied economics statistics*, ed. By A. Ullah and D. E. A. Giles, 237–290. New York: Marcel Dekker.

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

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

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

Kelejian, H. H., and I. R. Prucha. 2001. On the asymptotic distribution of the Moran *I* test statistic with applications. *Journal of Econometrics* 104:219–57. doi:10.1016/S0304-4076(01)00064-1.

Liu, X. D., L. F. Lee, and C. R. Bollinger. 2010. An efficient GMM estimator of spatial autoregressive models. *Journal of Econometrics* 159 (2):303–19. doi:10.1016/j.jeconom.2010.08.001.

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. 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.

Schennach, S. M. 2007. Point estimation with exponentially tilted empirical likelihood. *The Annals of Statistics* 35 (2):634–72. doi:10.1214/009053606000001208.