Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Journal of Econometrics 137 (2007) 396–413
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Finite sample properties of maximum likelihood estimator in spatial models

Yong Bao[a], Aman Ullah[b,*]

[a]Department of Economics, University of Texas, San Antonio, TX 78249, USA
[b]Department of Economics, University of California, Riverside, CA 92521, USA

Received 7 January 2004; accepted 5 August 2005
Available online 24 April 2006

Abstract

We investigate the finite sample properties of the maximum likelihood estimator for the spatial autoregressive model. A stochastic expansion of the score function is used to develop the second-order bias and mean squared error of the maximum likelihood estimator. We show that the results can be expressed in terms of the expectations of cross products of quadratic forms, or ratios of quadratic forms in a normal vector which can be evaluated using the top order invariant polynomial. Our numerical calculations demonstrate that the second-order behaviors of the maximum likelihood estimator depend on the degree of sparseness of the weights matrix.

JEL classification: C10; C21

Keywords: Bias; Mean squared error; Spatial autoregressive model

1. Introduction

Consider the following spatial lag model:

$$y = \rho Wy + \varepsilon, \tag{1}$$

where y is an $n \times 1$ vector of observations on the dependent spatial variable, Wy is the corresponding spatially lagged dependent variable for weights matrix W, which is assumed to be known a priori, ε is an $n \times 1$ vector of independent and identically distributed (IID)

*Corresponding author. Tel.: +1 951 827 5037x1591; fax: +1 951 787 5685.
E-mail addresses: yong.bao@utsa.edu (Y. Bao), aman.ullah@ucr.edu (A. Ullah).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2005.08.006

Gaussian error terms with zero mean and finite variance σ^2, and ρ is the spatial autoregressive parameter. The spatial lag model (1) is assumed to be an equilibrium model, and under the assumption that $A \equiv I - \rho W$ is invertible, where $I = I_n$ is the identity matrix of dimension n, we can write the equilibrium solution as $y = A^{-1}\varepsilon$, which gives $\mathrm{Var}(y) = \sigma^2 A^{-1} A^{-1\prime}$. This implies that each spatial observation is correlated with every other spatial observation.

It is well known that the ordinary least squares (OLS) estimator of ρ is inconsistent for the spatial model since the right-hand side variable in (1) is correlated with the error term. Under some regularity assumptions, which we collect in the Appendix, the average sample likelihood function

$$\mathscr{L}(\rho, \sigma^2) = \frac{1}{n} \ln |I - \rho W| - \frac{1}{2} \ln(2\pi\sigma^2) - \frac{\varepsilon'\varepsilon}{2n\sigma^2} \tag{2}$$

is well defined and continuous, and Lee (2004) first formally proved that the maximum likelihood (ML) estimator has the usual asymptotic properties, including $\sqrt{n}$-consistency, normality, and asymptotic efficiency.

The analytical finite sample properties of the ML estimator are however not known. To our knowledge, the existing finite sample analysis in this subject has been done through Monte Carlo simulations, for example, Anselin (1980, 1982) presented the small sample bias, mean squared error (MSE), and mean absolute percentage error of the ML estimator as well as several other estimators; Anselin and Florax (1995) investigated the finite sample properties of tests for spatial dependence ($\mathrm{H}_0 : \rho = 0$ versus $H_1 : \rho \neq 0$); Das et al. (2003) compared the efficiency of the ML estimator and several other estimators. An analytical investigation of the finite sample properties of the ML estimator is not yet available in the literature. The finite sample analytical results can help us understand the source of finite sample bias, for example, design a bias-corrected estimator, determine how big the sample size is needed so that the asymptotic theory can be used safely, and check the accuracy of Monte Carlo results. One difficulty in developing the analytical finite sample results for the spatial model is particularly due to the fact that y is not IID (note that $\mathrm{Var}(y) = \sigma^2 A^{-1} A^{-1\prime}$). This paper aims to fill this gap by developing the second-order bias and MSE of the ML estimator for the spatial autoregressive model. Since the score function is just a moment condition for the ML estimator, we follow a Nagar-type expansion (Nagar, 1959) to derive the second-order bias and MSE results.

The organization of this paper is as follows. In Section 2, we give the general results on the second-order bias and MSE of a class of $\sqrt{n}$-consistent estimators based on some suitable moment conditions, and where the sample observations are non-IID. The main results on the bias and MSE of the ML estimator for the spatial autoregressive model are presented in Section 3. As a starting case, we first assume that the error variance in (1) is known and we show that the results are in terms of the expectations of cross products of quadratic forms in ε. Next, we consider the more general setting where σ^2 and ρ are estimated simultaneously. We show that the results are in terms of the expectations of cross products of ratios of quadratic forms in ε and can be evaluated with the help of the top order invariant polynomial. Section 4 gives some numerical calculations based on our analytical results. It is shown that the bias and MSE behaviors of the ML estimator of the spatial autoregressive parameter depend on the degree of sparseness of the weights matrix. Consequently, the performance of bias correction varies. For very sparse weights matrices, our bias formula and its feasible version can accurately calculate the bias and the resulting bias-corrected estimator is almost unbiased. But more attentions need to be called for when we have dense weights matrices and the true parameter is negatively large. Section 5 contains our conclusion. All the proofs and technical details are collected in the Appendix.

2. Second-order moments of $\sqrt{n}$-consistent estimators

Let us consider a class of $\sqrt{n}$-consistent estimators identified by the moment condition

$$\hat{\theta} = \arg\{\psi_n(\theta) = 0\}, \tag{3}$$

where $\psi_n(\theta) = \psi_n(Z;\theta)$ is a known $k \times 1$ vector-valued function of the observable data $Z = \{Z_i\}_{i=1}^n$, IID or non-IID, and a parameter vector θ of k elements (of the same dimension as $\psi_n(\theta)$) such that $\mathbb{E}[\psi_n(\theta)] = 0$. Under some regularity assumptions, which are collected in Appendix A.1, and using a Taylor series expansion of the moment condition $\psi_n(\hat{\theta}) = 0$, we can write a second-order stochastic expansion of $\hat{\theta} - \theta$ (see Appendix A.1),

$$\hat{\theta} - \theta = a_{-1/2} + a_{-1} + a_{-3/2} + \mathrm{o_P}(n^{-3/2}), \tag{4}$$

where $a_{-s/2}$ represents terms of order $\mathrm{O_P}(n^{-s/2})$ for $s = 1, 2, 3$ and they are

$$\begin{aligned}
a_{-1/2} &= -Q\psi_n, \\
a_{-1} &= -QVa_{-1/2} - \tfrac{1}{2}Q\overline{H_2}(a_{-1/2} \otimes a_{-1/2}), \\
a_{-3/2} &= -QVa_{-1} - \tfrac{1}{2}QG(a_{-1/2} \otimes a_{-1/2}) - \tfrac{1}{2}Q\overline{H_2}[(a_{-1/2} \otimes a_{-1}) + (a_{-1} \otimes a_{-1/2})] \\
&\quad - \tfrac{1}{6}Q\overline{H_3}(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}),
\end{aligned} \tag{5}$$

in which $\overline{X} = \mathbb{E}(X)$ denotes the expectation of a random vector X, $\nabla^s A(\theta)$ is the matrix of sth order partial derivative of $A(\theta)$ and is obtained recursively (specifically, if $A(\theta)$ is a $k \times 1$ vector function, the jth element of the lth row of $\nabla^s A(\theta)$ (a $k \times k^s$ matrix) is the $1 \times k$ vector $a_{lj}^s(\theta) = \partial a_{lj}^{s-1}(\theta)/\partial\theta'$), $H_i = \nabla^i\psi_n$, $i = 1,2,3$, $\psi_n = \psi_n(\theta)$, $Q = \overline{H_1}^{-1}$, $V = H_1 - \overline{H_1}$, $d = Q\psi_n$, $G = H_2 - \overline{H_2}$, and $\otimes$ represents the Kronecker product. This leads to the second-order bias of $\hat{\theta}$, up to $\mathrm{O}(n^{-1})$,

$$B(\hat{\theta}) = \mathbb{E}(a_{-1}) \tag{6}$$

because $\mathbb{E}(a_{-1/2}) = 0$ since $\mathbb{E}(\psi_n) = 0$. Further, the second-order MSE of $\hat{\theta}$, up to $\mathrm{O}(n^{-2})$, is

$$M(\hat{\theta}) = \mathbb{E}(a_{-1/2}a'_{-1/2}) + \mathbb{E}(a_{-1/2}a'_{-1} + a_{-1}a'_{-1/2}) + \mathbb{E}(a_{-1}a'_{-1} + a_{-3/2}a'_{-1/2} + a_{-1/2}a'_{-3/2}). \tag{7}$$

If we substitute (5) into (6) and (7), one can write the second-order bias and MSE results in an alternate way, which we present in the Appendix (see (A.3) and (A.4)). If the series is IID, (6) and (7) degenerate into the results of Rilstone et al. (1996), while Bao and Ullah (2003) treated (6) and (7) explicitly for the case of time series data. When we have a scalar ($k = 1$), the bias result (6) (or (A.3)) reduces to

$$B(\hat{\theta}) = Q^2\overline{H_1\psi_n} - \tfrac{1}{2}Q^3\overline{H_2}\overline{\psi_n^2}, \tag{8}$$

and the MSE result (7) (or (A.4)) reduces to

$$M(\hat{\theta}) = 6Q^2\overline{\psi_n^2} - 8Q^3\overline{H_1\psi_n^2} + 3Q^4\overline{H_1^2\psi_n^2} + 4Q^4\overline{H_2}\,\overline{\psi_n^3} + Q^4\overline{H_2\psi_n^3} - 4Q^5\overline{H_2}\,\overline{H_1\psi_n^3} - \tfrac{1}{3}Q^5\overline{H_3}\,\overline{\psi_n^4} + \tfrac{5}{4}Q^6\overline{H_2}^2\overline{\psi_n^4}. \tag{9}$$

3. Main results

In this section, we consider the model in (1) with the average likelihood function in (2) and start with a simple case when σ^2 is assumed to be known a priori and our interest is in the spatial autoregressive parameter ρ. We show that the second-order bias and MSE results for the ML estimator of ρ, $\hat{\rho}_{\mathrm{ML}}$, can be expressed in terms of the expectations of cross products of quadratic forms in the normal vector ε. We then discuss the case when ρ and σ^2 are estimated simultaneously, where the results are more involved in terms of the expectations of cross products of ratios of quadratic forms in ε. But with the top order invariant polynomial theory, these expectations can be evaluated straightforwardly. We can verify that the smoothness assumptions (see Assumptions 4–6 in Appendix A.1) on the score functions to be presented in the following (see (10) and (13)) are satisfied for the spatial autoregressive model, while the $\sqrt{n}$-consistency can be obtained by the assumptions on W and ρ (see Assumptions 1–3 in Appendix A.1) by following Lee (2004).

3.1. σ^2 Known

Suppose that σ^2 in (2) is known, then we have only one parameter, ρ, to estimate, with the score function being be moment condition, $\psi_n = 0$, where ψ_n and its higher derivatives $H_i = \nabla^i\psi_n$, $i = 1, 2, 3$, are

$$\psi_n = \frac{1}{n}\left(B_1 - \frac{\varepsilon' M_1 \varepsilon}{2\sigma^2}\right), \quad H_1 = \frac{1}{n}\left(B_2 - \frac{\varepsilon' M_2 \varepsilon}{2\sigma^2}\right), \quad H_2 = \frac{1}{n}B_3, \quad H_3 = \frac{1}{n}B_4, \tag{10}$$

in which $B_i = \partial^i \ln|A|/\partial\rho^i$, $M_i = A^{-1\prime}[\partial^i(I - \rho(W' + W) + \rho^2 W'W)/\partial\rho^i]A^{-1}$. In particular, $B_1 = -\mathrm{tr}(A^{-1}W)$, $B_2 = -\mathrm{tr}[(A^{-1}W)^2]$, $B_3 = -2\,\mathrm{tr}[(A^{-1}W)^3]$, and $B_4 = -6\,\mathrm{tr}[(A^{-1}W)^4]$.

Let $\theta_{ij} = \mathbb{E}[(\varepsilon' M_1 y)^i \cdot (\varepsilon' M_2 \varepsilon)^j]$, then $Q = (\overline{H_1})^{-1} = n(B_2 - \sigma^{-2}\theta_{01}/2)^{-1}$ and $\overline{H_2} = H_2$, $\overline{H_3} = H_3$ since H_2 and H_3 are nonstochastic. Further, for the bias and MSE of $\hat{\rho}_{\mathrm{ML}}$ from (8) and (9), we have the following expectations:

$$\overline{H_1\psi_n} = \frac{1}{n^2}\left(B_1B_2 - \frac{B_2}{2\sigma^2}\theta_{10} - \frac{B_1}{2\sigma^2}\theta_{01} - \frac{1}{4\sigma^4}\theta_{11}\right),$$

$$\overline{\psi_n^2} = \frac{1}{n^2}\left(B_1^2 - \frac{B_1}{\sigma^2}\theta_{10} + \frac{1}{4\sigma^4}\theta_{20}\right),$$

$$\overline{\psi_n^3} = \frac{1}{n^3}\left(B_1^3 - \frac{3B_1^2}{2\sigma^2}\theta_{10} + \frac{3B_1}{4\sigma^4}\theta_{20} - \frac{1}{8\sigma^6}\theta_{30}\right),$$

$$\overline{\psi_n^4} = \frac{1}{n^4}\left(B_1^4 - \frac{2B_1^3}{\sigma^2}\theta_{10} + \frac{3B_1^2}{2\sigma^4}\theta_{20} - \frac{B_1}{2\sigma^6}\theta_{30} + \frac{1}{16\sigma^8}\theta_{40}\right),$$

$$\overline{H_1\psi_n^2} = \frac{1}{n^3}\left(B_1^2B_2 - \frac{B_1B_2}{\sigma^2}\theta_{10} + \frac{B_2}{4\sigma^4}\theta_{20} - \frac{B_1^2}{2\sigma^2}\theta_{01} + \frac{B_1}{2\sigma^4}{}_{11} - \frac{1}{8\sigma^6}\theta_{21}\right)$$

$$\overline{H_1^2\psi_n^2} = \frac{1}{n^4}\left(B_1^2B_2^2 - \frac{B_1B_2^2}{\sigma^2}\theta_{10} + \frac{B_2^2}{4\sigma^4}\theta_{20} - \frac{B_1^2B_2}{\sigma^2}\theta_{01} + \frac{B_1B_2}{\sigma^4}\theta_{11} - \frac{B_2}{4\sigma^6}\theta_{21} + \frac{B_1^2}{4\sigma^4}\theta_{02} - \frac{B_1}{4\sigma^6}\theta_{12} + \frac{1}{16\sigma^8}\theta_{22}\right),$$

$$\overline{H_1\psi_n^3} = \frac{1}{n^4}\left(B_1^3B_2 - \frac{3B_1^2B_2}{2\sigma^2}\theta_{10} + \frac{3B_1B_2}{4\sigma^4}\theta_{20} - \frac{B_2}{8\sigma^6}\theta_{30} - \frac{B_1^3}{2\sigma^2}\theta_{01} + \frac{3B_1^2}{4\sigma^4}\theta_{11} - \frac{3B_1}{8\sigma^6}\theta_{21} + \frac{1}{16\sigma^8}\theta_{31}\right),$$

$$\overline{H_2\psi_n^3} = \frac{1}{n^4}\left(B_1^3B_3 - \frac{3B_1^2B_3}{2\sigma^2}\theta_{10} + \frac{3B_1B_3}{4\sigma^4}\theta_{20} - \frac{B_3}{8\sigma^6}\theta_{30}\right). \tag{11}$$

The above expectations involve θ_{ij}, moments of cross products of quadratic forms in the normal vector $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$. Using Magnus (1978, 1979), for example, we can easily work out the explicit expressions for these θ_{ij}'s, which we give in Appendix A.2. In summary, given ρ and W, the algorithm to evaluate the second-order bias and MSE of $\hat{\rho}_{\mathrm{ML}}$ when σ^2 is known is as follows: (i) calculate θ_{ij} as given in Appendix A.2; (ii) plug θ_{ij} into (11); (iii) plug the results from (ii) to (8) and (9) to get the second-order bias and MSE of $\hat{\rho}_{\mathrm{ML}}$.

3.2. σ^2 Unknown

Now suppose that σ^2 is unknown and we have to estimate ρ and σ^2 simultaneously. As is common in practice, the ML estimation procedure is often based on a concentrated likelihood function in terms of the autoregressive parameter only, which simplifies the maximization procedure substantially, see Anselin (1980) and Anselin and Bera (1998). This also simplifies our derivation since it is easier to work with a scalar case than with a vector to derive the second-order results. Hence in what follows, we pursue this strategy by working with the concentrated likelihood function.

For the spatial autoregressive model (1), if ρ is known, the ML estimator of σ^2 is given by $\hat{\sigma}^2_{\mathrm{ML}} = (y - \rho Wy)'(y - \rho Wy)/n = y'Dy/n$, where $D = I - \rho(W + W') + \rho^2 W'W$. Substituting this into (2) gives us a concentrated likelihood function in terms of ρ only

$$\mathscr{L}(\rho) = \frac{1}{n}\sum_{i=1}^{n}\left[\ln(1 - \rho\omega_i) - \frac{1}{2}\ln\left(\frac{2\pi}{n}y'Dy\right) - \frac{n}{2}\right], \tag{12}$$

where ω_i's are the eigenvalues of W. Therefore, we have the score function ψ_n for estimating $\hat{\rho}_{\mathrm{ML}}$, as well as its higher-order derivatives as follows:

$$\psi_n = \frac{B_1}{n} - \frac{1}{2}\frac{\varepsilon' M_1 \varepsilon}{\varepsilon'\varepsilon},$$

$$H_1 = \frac{B_2}{n} - \frac{1}{2}\frac{\varepsilon' M_2 \varepsilon}{\varepsilon'\varepsilon} + \frac{1}{2}\left(\frac{\varepsilon' M_1 \varepsilon}{\varepsilon'\varepsilon}\right)^2,$$

$$H_2 = \frac{B_3}{n} + \frac{3}{2}\frac{\varepsilon' M_1\varepsilon \cdot \varepsilon' M_2\varepsilon}{(\varepsilon'\varepsilon)^2} - (\frac{\varepsilon' M_1\varepsilon}{\varepsilon'\varepsilon})^3,$$

$$H_3 = \frac{B_4}{n} + \frac{3}{2}\left(\frac{\varepsilon' M_2\varepsilon}{\varepsilon'\varepsilon}\right)^2 - 6\frac{(\varepsilon' M_1\varepsilon)^2 \cdot \varepsilon' M_2\varepsilon}{(\varepsilon'\varepsilon)^3} + 3\left(\frac{\varepsilon' M_1\varepsilon}{\varepsilon'\varepsilon}\right)^4. \tag{13}$$

Let $\lambda_{ij} = \mathbb{E}[(\varepsilon' M_1\varepsilon)^i \cdot (\varepsilon' M_2\varepsilon)^j/(\varepsilon'\varepsilon)^{i+j}]$, then we can write $Q = \overline{H_1}^{-1} = n(B_2 - \lambda_{01}/n + \lambda_{20}/n)^{-1}$, $\overline{H_2} = (B_3 + 3n\lambda_{11}/2 - n\lambda_{30})/n$, $\overline{H_3} = (B_4 + 3n\lambda_{02}/2 - 6n\lambda_{21} + 3n\lambda_{40})/n$. Further we have the following expectations:

$$\overline{H_1\psi_n} = \frac{1}{n^2}B_1B_2 - \frac{1}{2n}B_1\lambda_{01} + \frac{1}{2n}B_1\lambda_{20} - \frac{1}{2n}B_2\lambda_{10} + \frac{1}{4}\lambda_{11} - \frac{1}{4}\lambda_{30},$$

$$\overline{\psi_n^2} = \frac{1}{n^2}B_1^2 - \frac{1}{n}B_1\lambda_{10} + \frac{1}{4}\lambda_{20},$$

$$\overline{\psi_n^3} = \frac{1}{n^3}B_1^3 - \frac{3}{2n^2}B_1^2\lambda_{10} + \frac{3}{4n}B_1\lambda_{20} - \frac{1}{8}\lambda_{30},$$

$$\overline{\psi_n^4} = \frac{1}{n^4}B_1^4 - \frac{2}{n^3}B_1^3\lambda_{10} + \frac{3}{2n^2}B_1^2\lambda_{20} - \frac{1}{2n}B_1\lambda_{30} + \frac{1}{16}\lambda_{40},$$

$$\begin{aligned}\overline{H_1\psi_n^2} = {} & \frac{1}{n^3}B_1^2B_2 + \frac{1}{2n^2}B_1^2\lambda_{20} - \frac{1}{n^2}B_1B_2\lambda_{10} - \frac{1}{2n^2}B_1^2\lambda_{01} + \frac{1}{2n}B_1\lambda_{11} + \frac{1}{4n}B_2\lambda_{20} \\ & - \frac{1}{2n}B_1\lambda_{30} - \frac{1}{8}\lambda_{21} + \frac{1}{8}\lambda_{40},\end{aligned}$$

$$\begin{aligned}\overline{H_1^2\psi_n^2} = {} & \frac{1}{n^4}B_1^2B_2^2 + \frac{1}{n^3}B_1^2B_2\lambda_{20} - \frac{1}{n^3}B_1^2B_2\lambda_{01} - \frac{1}{n^3}B_1B_2^2\lambda_{10} + \frac{1}{4n^2}B_1^2\lambda_{02} + \frac{1}{4n^2}B_1^2\lambda_{40} \\ & - \frac{1}{2n^2}B_1^2\lambda_{21} + \frac{1}{4n^2}B_2^2\lambda_{20} + \frac{1}{n^2}B_1B_2\lambda_{11} - \frac{1}{n^2}B_1B_2\lambda_{30} + \frac{1}{2n}B_1\lambda_{31} - \frac{1}{4n}B_1\lambda_{12} \\ & - \frac{1}{4n}B_1\lambda_{50} - \frac{1}{4n}B_2\lambda_{21} + \frac{1}{4n}B_2\lambda_{40} + \frac{1}{16}\lambda_{22} + \frac{1}{16}\lambda_{60} - \frac{1}{8}\lambda_{41},\end{aligned}$$

$$\begin{aligned}\overline{H_1\psi_n^3} = {} & \frac{1}{n^4}B_1^3B_2 + \frac{1}{2n^3}B_1^3\lambda_{20} - \frac{1}{2n^3}B_1^3\lambda_{01} - \frac{3}{2n^3}B_1^2B_2\lambda_{10} + \frac{3}{4n^2}B_1^2\lambda_{11} - \frac{3}{4n^2}B_1^2\lambda_{30} \\ & + \frac{3}{4n^2}B_1B_2\lambda_{20} - \frac{1}{8n}B_2\lambda_{30} - \frac{3}{8n}B_1\lambda_{21} + \frac{3}{8n}B_1\lambda_{40} + \frac{1}{16}\lambda_{31} - \frac{1}{16}\lambda_{50},\end{aligned}$$

$$\begin{aligned}\overline{H_2\psi_n^3} = {} & \frac{1}{n^4}B_1^3B_3 + \frac{3}{2n^3}B_1^3\lambda_{11} - \frac{1}{n^3}B_1^3\lambda_{30} - \frac{3}{2n^3}B_1^2B_3\lambda_{10} - \frac{9}{4n^2}B_1^2\lambda_{21} + \frac{3}{2n^2}B_1^2\lambda_{40} \\ & + \frac{3}{4n^2}B_1B_3\lambda_{20} + \frac{9}{8n}B_1\lambda_{31} - \frac{3}{4n}B_1\lambda_{50} - \frac{1}{8n}B_3\lambda_{30} - \frac{3}{16}\lambda_{41} + \frac{1}{8}\lambda_{60},\end{aligned} \tag{14}$$

which, together with Q, $\overline{H_2}$, and $\overline{H_3}$ can be plugged into (8) and (9) to get the second-order bias and MSE of $\hat{\rho}_{\mathrm{ML}}$ in terms of λ_{ij}, moments of cross products of ratios of quadratic forms in the vector $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$. Note that from the definition of λ_{ij}, replacing ε with $\epsilon = \varepsilon/\sigma \sim \mathcal{N}(0, I)$ does not affect our results.

To evaluate these terms, we note that the moments of cross products of ratios of quadratic forms in a normal vector, λ_{ij}, are not that easy to derive compared when we have cross products of quadratic forms. To our knowledge, there has been no explicit effort in the literature attempting to derive the analytical formulae for $\mathbb{E}[(\epsilon'\epsilon)^{-f}\prod_{i=1}^{r}(\epsilon' A_i\epsilon)^{k_i}]$, where $f=\sum_{i=1}^{r}k_i$ and A_i are $n\times n$ and symmetric (in our case they are either M_1 or M_2). But we notice that this task can be accomplished if we follow the top order invariant polynomial approach of Smith (1989). In particular, we can use Eq. (5.4) of Smith (1989) directly

$$\mathbb{E}[(\epsilon'\epsilon)^{-f}\prod_{i=1}^{r}(\epsilon' A_i\epsilon)^{k_i}]=\frac{(\frac{1}{2})_f}{(\frac{1}{2}n)_f}C_{[f]}^{k[r]}(A_{[r]}), \tag{15}$$

where the Pochammer symbol $(a)_b=\Gamma(a+b)/\Gamma(a)$, $[a]$ denotes the top order partition $(a,0,\ldots,0)$, $a[A]$ denotes $[a_1],\ldots,[a_A]$ for the top order partitions of scalars a_i, $i=1,\ldots,A$, $A_{[a]}$ denotes $A_1,\ldots,A_a$ for symmetric matrices A_i, $i=1,\ldots,a$, and $C_{[f]}^{k[r]}(A_{[r]})$ is an invariant polynomial introduced by Davis (1979, 1980, 1981). Apparently, $C_{[f]}^{k[r]}(A_{[r]})$ in (15) appears in the special form of a top order invariant polynomial. Further notice that M_1 and M_2 are mutually commutative and symmetric. Therefore, Theorem (2.1) of Chikuse (1987) can be used directly and we discuss in Appendix A.3 how to evaluate (15) and present explicit expressions for the λ_{ij}'s that are needed in (14).

In summary, given ρ and W, the algorithm to evaluate the second-order bias and MSE of $\hat{\rho}_{\mathrm{ML}}$ is as follows: (i) calculate λ_{ij} as given in Appendix A.3; (ii) plug λ_{ij} into (14); (iii) plug the results from (ii) to (8) and (9) to get the second-order bias and MSE of $\hat{\rho}_{\mathrm{ML}}$.

Besides the second-order results on $\hat{\rho}_{\mathrm{ML}}$, we may be interested as well in the properties of $\hat{\sigma}^2_{\mathrm{ML}}=e'e/n$, where $e=\hat{A}y$, $\hat{A}=I-\hat{\rho}_{\mathrm{ML}}W$. At first glance, this does not seem to fit in with our framework, since, strictly speaking, $\hat{\sigma}^2_{\mathrm{ML}}$ is not based directly on some suitable moment condition on itself, but on the moment condition on $\hat{\rho}_{\mathrm{ML}}$, which in turn is plugged to get $\hat{\sigma}^2_{\mathrm{ML}}$. However, as long as we can write a stochastic expansion for $\hat{\rho}_{\mathrm{ML}}$, so can we do for $\hat{\sigma}^2_{\mathrm{ML}}$ and hence the second-order results follow immediately.

For example, to work out the second-order bias of $\hat{\sigma}^2_{\mathrm{ML}}$, from (4), (5), and (13), we can write $\hat{\rho}_{\mathrm{ML}}-\rho=a_{-1/2}+a_{-1}+\mathrm{o_P}(n^{-1})$, where $a_{-1/2}=-Q(B_1/n-\varepsilon' M_1\varepsilon/2\varepsilon'\varepsilon)=\mathrm{O_P}(n^{-1/2})$, $a_{-1}=a_{-1/2}-Qa_{-1/2}[B_2/n-\varepsilon' M_2\varepsilon/2\varepsilon'\varepsilon+(\varepsilon' M_1\varepsilon/\varepsilon'\varepsilon)^2/2]-Q\,\overline{H_2}\,a^2_{-1/2}/2=\mathrm{O_P}(n^{-1})$, $Q=\overline{H_1}^{-1}=n(B_2-\lambda_{01}/n+\lambda_{20}/n)^{-1}$, and $\overline{H_2}=(B_3+3n\lambda_{11}/2-n\lambda_{30})/n$. Then by substitution

$$\begin{aligned}\hat{\sigma}^2_{\mathrm{ML}}&=\frac{1}{n}y'[I-(\rho+a_{-1/2}+a_{-1})W]'[I-(\rho+a_{-1/2}+a_{-1})W]y+\mathrm{o_P}(n^{-1})\\&=\frac{1}{n}\varepsilon'\varepsilon+b_{-1/2}+b_{-1}++\mathrm{o_P}(n^{-1}),\end{aligned} \tag{16}$$

where $b_{-1/2}=a_{-1/2}\varepsilon' A^{-1\prime}(2\rho W'W-W-W')A^{-1}\varepsilon/n=\mathrm{O_P}(n^{-1/2})$ and $b_{-1}=\varepsilon' A^{-1\prime}[a_{-1}(2\rho W'W-W-W')+a^2_{-1/2}W'W]A^{-1}\varepsilon/n=\mathrm{O_P}(n^{-1})$. Noting that $\mathbb{E}(\varepsilon'\varepsilon/n)=\sigma^2$, we have the second-order bias of $\hat{\sigma}^2_{\mathrm{ML}}$, up to $\mathrm{O}(n^{-1})$, as

$$B(\hat{\sigma}^2_{\mathrm{ML}})=\mathbb{E}(b_{-1/2}+b_{-1}). \tag{17}$$

Again, $B(\hat{\sigma}^2_{\mathrm{ML}})$ can be represented as the expectations of cross products of ratios of quadratic forms in ε and we can follow the same steps outlined before to evaluate it. The second-order MSE follows by similar strategy.

4. Numerical calculations

In this section, we give some numerical calculations to illustrate how to evaluate the second-order bias and MSE of the ML estimator in the spatial model and the merits of bias correction. Following Kelejian and Prucha (1999), we consider three specifications of the weights matrix with different degree of sparseness, namely, the "one ahead and one behind," "three ahead and three behind," and "five ahead and five behind" matrices, denoted by $W_{J=2}$, $W_{J=6}$, and $W_{J=10}$, respectively.[1] We row-standardize the three matrices and set all the non-zero elements to be equal to each other. Apparently, the three weights matrices satisfy the assumptions on the weights matrix outlined in the Appendix. We normalize $\sigma^2 = 1$ and estimate ρ and σ^2 simultaneously. We focus on the results on $\hat{\rho}_{\mathrm{ML}}$.

Fig. 1 plots the theoretical second-order bias $B(\hat{\rho}_{\mathrm{ML}})$ and standard error $\sqrt{M(\hat{\rho}_{\mathrm{ML}}) - [B(\hat{\rho}_{\mathrm{ML}})]^2}$ of $\hat{\rho}_{\mathrm{ML}}$ using the true value of ρ, for $n = 30$, 100. When the sample size n increases, as we expect, the second-order bias and standard error reduce significantly. We observe that when we have a sparse matrix $W_{J=2}$, the second-order bias and standard error behave approximately symmetric between the positive and negative parts of allowable parameter space in the sense that the bias tends to be positive for $\rho < 0$ and negative for $\rho > 0$, while the standard error decreases for increasing $\rho > 0$ and it increases for decreasing $\rho < 0$. On the other hand, when the degrees of sparseness of the weights matrix increases ($J = 6$ or 10), we observe that the bias and standard error do not behave symmetrically. For most values of allowable parameter, the bias is negative and the standard error seems to monotonically decrease as ρ goes from the negative part to the positive part. This means that when observations are densely distributed across space and are negatively correlated, $\hat{\rho}_{\mathrm{ML}}$ usually underestimates the true ρ and is much less precise than the case when we have positively correlated spatial observations.

Next, we investigate the merit of bias correction given our analytical results. We consider seven values for ρ, namely, ± 0.9, ± 0.4, ± 0.2, 0 and two values of the sample size n, namely, 30 and 100. We conduct a Monte Carlo simulations to compare the behavior of the ML estimator and the bias-corrected one. The number of replications is 5000 when $n = 30$ and 1000 when $n = 100$. Also, we report the first-order asymptotic standard error.[2] Reported in Table 1 are first: the averaged $\hat{\rho}_{\mathrm{ML}}$, the bias-corrected one, $\hat{\rho}_{\mathrm{BC}} = \hat{\rho}_{\mathrm{ML}} - B(\hat{\rho}_{\mathrm{ML}})$, using the true parameter value, and the feasible version of $\hat{\rho}_{\mathrm{BC}}$, $\tilde{\rho}_{\mathrm{BC}} = \hat{\rho}_{\mathrm{ML}} - \hat{B}(\hat{\rho}_{\mathrm{ML}})$, using the estimated parameter value. What follows next are the true standard deviation of $\hat{\rho}_{\mathrm{ML}}$ (that is, across the replications), the (theoretical) asymptotic standard error, the averaged asymptotic standard error evaluated at $\hat{\rho}_{\mathrm{ML}}$, the (theoretical) second-order standard error, $M(\hat{\rho}_{\mathrm{ML}}) - B^2(\hat{\rho}_{\mathrm{ML}})$, evaluated at the true parameter value, and the average of the 5000/1000 replicated estimates of this evaluated at $\hat{\rho}_{\mathrm{ML}}$.

[1] A "one ahead and one behind" matrix has the ith row with non-zero elements only in positions $i-1$ and $i+1$, $i = 2, \ldots, n-1$, and the first row has non-zero elements only in positions 2 and n while for the last row the non-zeros occur only in positions 1 and $n-1$. By this, we define the weights matrix in a circular way. The average number of neighboring units J for the "one ahead and one behind" matrix is hence 2. Similarly, we can define the "two ahead and two behind" and "three ahead and three behind" matrices.

[2] In Lee (2004), the asymptotic variance is expressed as a limit of terms involving W. In practice, however, it is difficult to derive the limit given any W. Nevertheless, since we can easily take the expectation of H_1 in (13), the first-order standard error of $\hat{\rho}_{\mathrm{ML}}$ is given by $\sqrt{-Q^{-1}/n}$.

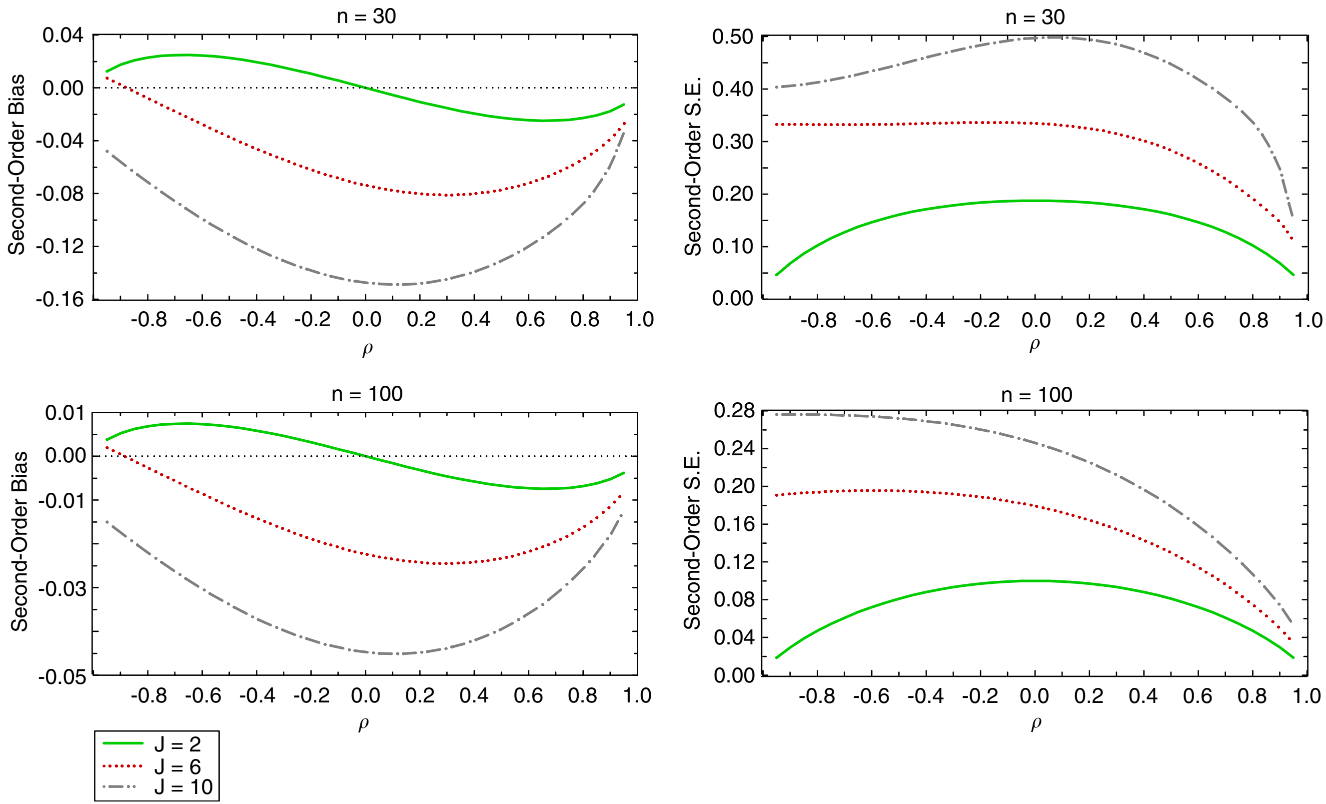

Fig. 1. Theoretical second-order bias and standard error, $n = 30, 100$.

Table 1
Bias correction for spatial model

n	J	ρ	$\hat{\rho}$	$\hat{\rho}_{BC}$	$\tilde{\rho}_{BC}$	$SD(\hat{\rho})$	$ASD(\hat{\rho})$	$A\hat{S}D(\hat{\rho})$	$ASD_2(\hat{\rho})$	$A\hat{S}D_2(\hat{\rho})$
30	2	−0.9	−0.882	−0.900	−0.900	0.056	0.046	0.051	0.069	0.074
		−0.4	−0.382	−0.401	−0.399	0.160	0.162	0.160	0.171	0.169
		−0.2	−0.188	−0.198	−0.197	0.175	0.182	0.178	0.184	0.181
		0	−0.001	−0.001	−0.001	0.182	0.189	0.183	0.188	0.184
		0.2	0.189	0.200	0.199	0.173	0.182	0.178	0.184	0.181
		0.4	0.382	0.402	0.400	0.155	0.162	0.160	0.171	0.170
		0.9	0.881	0.899	0.899	0.057	0.046	0.052	0.069	0.074
	6	−0.9	−0.810	−0.812	−0.804	0.233	0.371	0.369	0.333	0.333
		−0.4	−0.434	−0.388	−0.394	0.328	0.372	0.360	0.334	0.333
		−0.2	−0.253	−0.191	−0.199	0.336	0.354	0.346	0.336	0.332
		0	−0.071	0.003	−0.007	0.325	0.327	0.324	0.335	0.328
		0.2	0.124	0.205	0.196	0.302	0.289	0.293	0.325	0.319
		0.4	0.322	0.402	0.397	0.264	0.241	0.251	0.302	0.300
		0.9	0.853	0.892	0.897	0.106	0.062	0.081	0.147	0.163
	10	−0.9	−0.774	−0.718	−0.704	0.294	0.549	0.532	0.406	0.421
		−0.4	−0.457	−0.335	−0.353	0.398	0.503	0.492	0.460	0.450
		−0.2	−0.301	−0.163	−0.184	0.411	0.467	0.466	0.483	0.461
		0	−0.133	0.014	−0.007	0.417	0.422	0.431	0.497	0.469
		0.2	0.066	0.214	0.198	0.392	0.366	0.385	0.495	0.470
		0.4	0.263	0.402	0.395	0.356	0.300	0.330	0.469	0.459
		0.9	0.817	0.875	0.890	0.169	0.070	0.111	0.246	0.285
100	2	−0.9	−0.894	−0.899	−0.899	0.026	0.025	0.026	0.030	0.031
		−0.4	−0.395	−0.401	−0.401	0.084	0.087	0.086	0.088	0.088
		−0.2	−0.200	−0.203	−0.203	0.095	0.098	0.097	0.097	0.096
		0	0.001	0.001	0.001	0.100	0.101	0.100	0.100	0.099
		0.2	0.200	0.203	0.203	0.099	0.098	0.097	0.097	0.096
		0.4	0.390	0.396	0.395	0.086	0.087	0.087	0.088	0.088
		0.9	0.896	0.901	0.901	0.027	0.025	0.025	0.030	0.030
	6	−0.9	−0.861	−0.862	−0.861	0.143	0.199	0.199	0.192	0.192
		−0.4	−0.411	−0.397	−0.398	0.194	0.199	0.197	0.194	0.193
		−0.2	−0.218	−0.199	−0.200	0.184	0.190	0.189	0.189	0.188
		0	−0.010	0.013	0.012	0.170	0.175	0.174	0.179	0.178
		0.2	0.178	0.202	0.201	0.155	0.155	0.156	0.164	0.164
		0.4	0.378	0.402	0.402	0.134	0.129	0.131	0.143	0.144
		0.9	0.885	0.897	0.898	0.043	0.034	0.037	0.049	0.053
	10	−0.9	−0.838	−0.821	−0.818	0.190	0.294	0.290	0.276	0.275
		−0.4	−0.432	−0.395	−0.397	0.255	0.269	0.268	0.269	0.267
		−0.2	−0.233	−0.191	−0.193	0.244	0.250	0.250	0.260	0.259
		0	−0.044	0.001	−0.001	0.229	0.226	0.228	0.246	0.246
		0.2	0.150	0.195	0.194	0.212	0.196	0.201	0.226	0.227
		0.4	0.345	0.387	0.387	0.183	0.161	0.169	0.197	0.202
		0.9	0.880	0.898	0.900	0.052	0.041	0.047	0.074	0.080

Note: $\hat{\rho}$ is the averaged MLE of ρ over Monte Carlo replications; $\hat{\rho}_{BC}$ is the averaged bias-corrected MLE using the true value of ρ; $\tilde{\rho}_{BC}$ is the feasible version of $\hat{\rho}_{BC}$; $SD(\hat{\rho})$ is the standard error of the MLE across the replications; $ASD(\hat{\rho})$ is the theoretical asymptotical standard error using the true value of ρ; $A\hat{S}D(\hat{\rho})$ is the averaged asymptotical standard error using the estimated value of ρ; $ASD_2(\hat{\rho})$ theoretical second-order standard error using the true value of ρ; $A\hat{S}D_2(\hat{\rho})$ is the averaged second-order standard error using the estimated value of ρ.

Regarding the Monte Carlo results, we have the following observations. *Firstly*, for positive ρ, $\hat{\rho}_{\mathrm{ML}}$ tends to underestimate the parameter value and the bias is not negligible even in sample of size as large as 100. The bias-corrected estimator is almost unbiased, even in sample of size as small as 30. For small $\rho(>0)$, the usual asymptotic standard errors provide reasonable estimates of the dispersion of $\hat{\rho}_{\mathrm{ML}}$ and going for the second-order seems unnecessary. But when $\rho(>0)$ is moderately large, there is a tendency for the asymptotic standard errors to underestimate the dispersion of $\hat{\rho}_{\mathrm{ML}}$ and in those cases, the second-order standard errors tend to overestimate the dispersion. So when we use bias correction in those cases, trade-off between bias and dispersion might come. *Secondly*, for negative ρ, when the weights matrix is sparse ($J = 2$), $\hat{\rho}_{\mathrm{ML}}$ tends to underestimate (in absolute value) and the bias is not negligible, but bias correction works very well. However, when ρ is large (in absolute value) and the weights matrix is less sparse, bias of $\hat{\rho}_{\mathrm{ML}}$ is more severe and bias correction does not seem to work, though the severity reduces as the sample size increases and when ρ decreases (in absolute value). From our previous analysis, we know that when observations are densely distributed across space and are negatively correlated, the estimation lacks precision. This leads us to wonder about the distribution behavior of $\hat{\rho}_{\mathrm{ML}}$ in that range. Fig. 2 plots the kernel density of $\hat{\rho}_{\mathrm{ML}}$ for $\rho = -0.9$, -0.6, -0.4, and -0.2 and $n = 30$ under different degree of sparseness of the weights matrix.[3] For very negative parameter value, while for very sparsely distributed observations the ML estimator behaves quite closely to what the asymptotic theory says, it is apparently not the case when $J = 6$ and 10. Moreover, as the degree of denseness increases, the departure from the asymptotic theory is more severe. In particular, the distribution of $\hat{\rho}_{\mathrm{ML}}$ is quite skewed and most of the probability mass is clustered to the left tail. This may help explain why $\hat{\rho}_{\mathrm{ML}}$ is severely biased and even bias correction does not work when ρ is negatively very large. The second-order results (6) and (7) we develop are under the assumption that the estimator has convergence rate $\sqrt{n}$, though normality is not required per se. When ρ is negatively large and each spatial observation has many neighbors, Fig. 2 suggests that the estimates are quite noisy and we may need really large n to achieve convergence.[4] *Finally*, we note that the behavior of the standard errors of $\hat{\rho}_{\mathrm{ML}}$ when $\rho<0$ is similar to the case when ρ is positive: for moderately negative ρ, the asymptotic standard errors work well; for negatively large ρ, the asymptotic standard errors tend to underestimate the dispersion of $\hat{\rho}_{\mathrm{ML}}$ and the second-order standard errors tend to overestimate. Some of the above observed features of the ML estimator when ρ is negatively large and the weights matrix is dense are also documented in Kelejian and Prucha (1999) in their Monte Carlo studies.

5. Conclusion

We have developed the analytical finite sample bias and MSE of the ML estimators of the parameters of the spatial autoregressive model. We investigate first the case when the error variance is assumed to be known and we show that the second-order results for the

[3]We use the "leave-one-out" kernel density to estimate the distribution of $\hat{\rho}_{\mathrm{ML}}$ non-parametrically. The "leave-one-out" kernel density for a sample $\{x_i\}_{i=1}^n$ is defined as $f(x_i) = \sum_{j=1, j\neq i}^n K((x_j - x_i)/h)/nh$, where h is the bandwidth and $K(\cdot)$ is the kernel function. We use a standard normal kernel and choose the bandwidth $h = 1.06 s_x n^{-1/5}$, where s_x is the sample standard deviation of $\{x_i\}_{i=1}^n$.

[4]We tried with $n = 50$, 100, 200 and the behavior of $\hat{\rho}_{\mathrm{ML}}$, when $\rho(<0)$ is vary large (in absolute value) for $J = 6$, 10, is similar to Fig. 2. To save space, we do not report the results here.

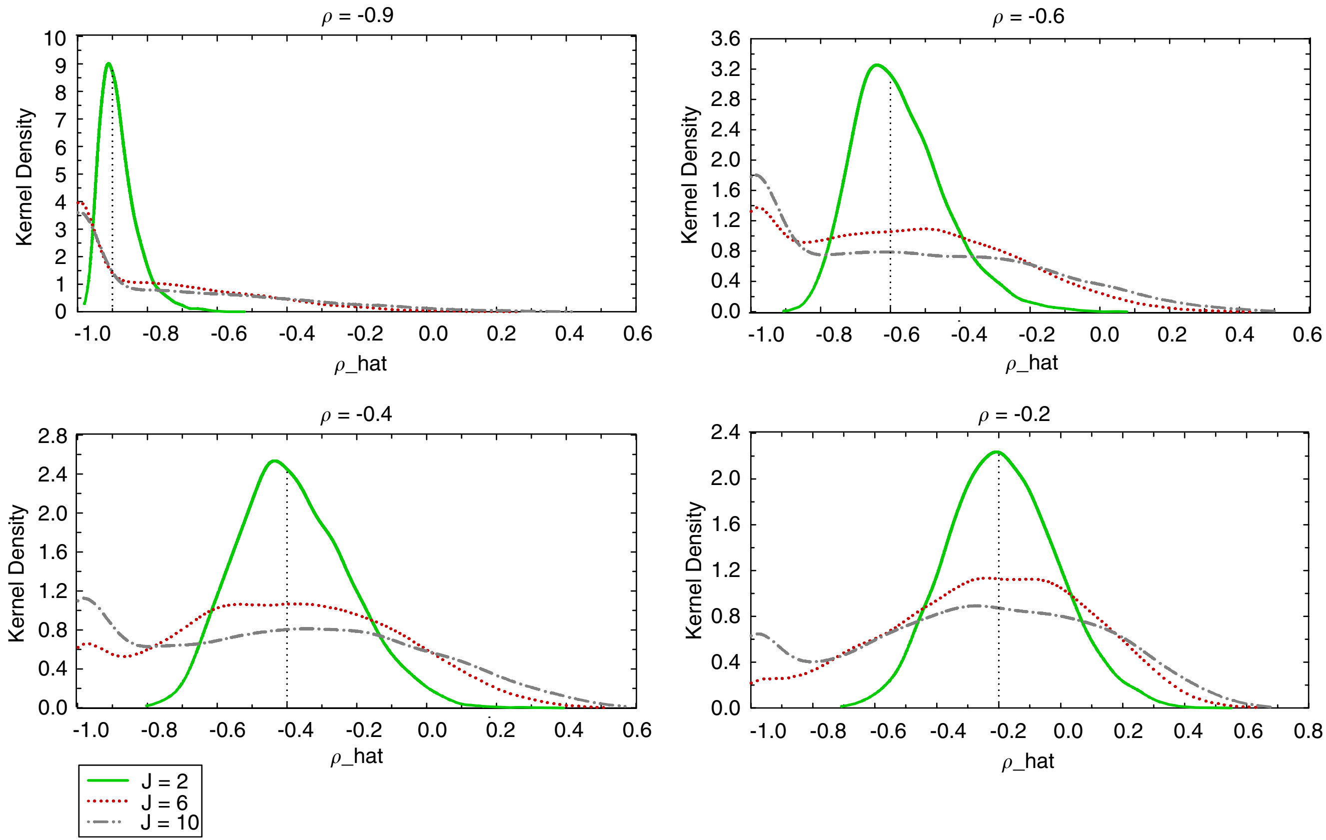

Fig. 2. Distribution of ρ_hat, $n = 30$.

estimator of the spatial autoregressive parameter can be compactly expressed in terms of the expectations of cross products of quadratic forms in a normal vector. When the spatial autoregressive parameter and the error variance are estimated simultaneously, the second-order results can be expressed in terms of the expectations of cross products of ratios of quadratic forms and with the help of the top order invariant polynomial theory, they can be easily calculated or programmed. Our numerical calculations show that when the weights matrix is sparse, bias correction is quite worthwhile since the resulting bias-corrected estimator is almost unbiased. The behavior of the ML estimator is quite inconsistent with what the asymptotic theory predicts, however, when the true parameter is negatively large and when the weights matrix is less sparse, where bias correction is not working either, though for the positive part of parameter space this problem does not happen. This provokes us to think about the distribution properties of the estimator and we find that in those cases the estimator is severely skewed and the probability mass is clustered to the left tail. To help better understand these phenomena, a thorough finite sample distribution theory is called for and this will be the subject of a future study.

Extensions to the cases when the error term also follows a spatial autoregressive process and when some exogenous regressors are added into the model are straightforward. In essence, as shown in Anselin (1980) and Lee (2004), the ML estimator for the spatial autoregressive parameter can always be based on the concentrated likelihood function (concentrating out the error variance and the coefficients of the exogenous regressors). We can show that the bias and MSE results can again be represented in terms of the expectations of cross products of ratios of quadratic forms in a normal vector and we can use the top order invariant polynomial to evaluate these expectations as we have done in this paper.

We may also consider other competitive estimators and compare their theoretical properties with those of the ML estimator. Note that the ML estimation requires a term $|I - \rho W|$, or its Ord (1975) decomposition $\prod_{i=1}^{N}(1 - \rho\omega_i)$, both of which are difficult for us to calculate accurately, especially when n is large. This prompts researchers to propose the moment estimator (Kelejian and Robinson, 1993; Conley, 1996; Kelejian and Prucha, 1998, 1999), which is claimed to have a computational advantage over the ML estimator when the sample size is large. While its analytical asymptotic properties have been established, the finite sample properties of the moment estimator are studied through Monte Carlo simulations. A theoretical investigation of its finite sample properties is not yet available and this can be the subject of another future study.

Acknowledgments

We are grateful to Ronald Gallant and two anonymous referees for their constructive and helpful comments that have led to improvement of our paper. We would also like to thank Fathali Firoozi and Bernard Gress for their helpful comments. Ullah thanks the UCR Academic Senate for research support. All remaining errors are our own.

Appendix A

A.1. Stochastic expansion and the second-order bias and MSE

In the following, we use tr$A$ to denote the trace, $\|A\|$ to denote the usual norm $[\mathrm{tr}(AA')]^{1/2}$, and $\|A\|_{\mathrm{Row}} = \max_{1\leqslant i\leqslant n}\sum_{j=1}^{n} \mathrm{abs}(a_{ij})$ to denote the row sum norm of an $n \times n$

matrix A with standard elements a_{ij}. We say that a matrix A is uniformly bounded in row sums if $\|A\|_{\mathrm{Row}} < c$, where c is a finite constant independent of n. For the spatial model $y = \rho W y + \varepsilon$, $A \equiv I - \rho W$ and the following assumptions are made:

Assumption 1. ρ is in the interior of a compact set $\Lambda \subseteq (-1, 1)$ and $\mathrm{abs}(|A|) > \alpha$ for some $\alpha > 0$ not depending on n or ρ.

Assumption 2. W is uniformly bounded in row sums, and A^{-1} is uniformly bounded in row sums, uniformly in $\rho \in \Lambda$.

Assumption 3. The elements w_{ij} of W are of order $\mathrm{O}(h_n^{-1})$ uniformly in all i,j, where the rate sequence $\{h_n\}$ can be bounded or divergent and $h_n/n \to 0$ as $n \to \infty$.

For the estimator $\hat{\theta} = \arg\{\psi_n(\theta) = 0\}$, the following assumptions are maintained:

Assumption 4. The sth order derivatives of $\psi_n(\vartheta)$ exist for ϑ in a neighborhood of θ and $\mathbb{E}(\|\nabla^s \psi_n(\theta)\|^2) < \infty$, for s up to 3.

Assumption 5. For ϑ in some neighborhood of $\theta, [\nabla \psi_n(\vartheta)]^{-1} = \mathrm{O_p}(1)$.

Assumption 6. $\|\nabla^s \psi_n(\vartheta) - \nabla^s \psi_n(\theta)\| \leqslant \|\vartheta - \theta\| M_n$ for ϑ in some neighborhood of θ, where $\mathbb{E}(|M_n|) < C < \infty$ for some positive constant C.

Essentially, Assumptions 1–3 put some constraints on the weights matrix W to limit the spatial correlation of y_i to a manageable degree and to guarantee that the likelihood function to be well defined and continuous. Under these three assumptions, together with the assumption that $\mathbb{E}(\mathrm{abs}(\varepsilon_i)^{4+\gamma})$ exists for some $\gamma > 0$, the ML estimators of the parameters of the spatial autoregressive model can be proved to be $\sqrt{n}$-consistent, see Lee (2004) for more technical details of the asymptotic results. Under the $\sqrt{n}$-consistency of $\hat{\theta}$ and Assumptions 4–6, Rilstone et al. (1996) and Bao and Ullah (2003) obtained the stochastic expansion in (4) except that they wrote $\psi_n(\theta) = \sum_{i=1}^n q_i(Z_i; \theta)/n$ and used Assumptions 4 and 6 in terms of $q_i(Z_i; \theta)$. We consider a general form of $\psi_n(\theta)$ in (3) and note that the existence and smoothness conditions on $\psi_n(\cdot)$ in Assumptions 4–6 are sufficient to obtain the stochastic expansion of $\psi_n(\hat{\theta})$ around θ. This is given below.

Given Assumptions 4–6 and the $\sqrt{n}$-consistency of $\hat{\theta}$, we implement a Taylor series expansion of the moment condition

$$\begin{aligned} 0 &= \psi_n(\hat{\theta}) \\ &= \psi_n(\theta) + \nabla \psi_n(\theta)(\hat{\theta} - \theta) + \tfrac{1}{2}\nabla^2 \psi_n(\theta)[(\hat{\theta} - \theta) \otimes (\hat{\theta} - \theta)] \\ &\quad + \tfrac{1}{6}\nabla^3 \psi_n(\theta)[(\hat{\theta} - \theta) \otimes (\hat{\theta} - \theta) \otimes (\hat{\theta} - \theta)] \\ &\quad + \tfrac{1}{6}[\nabla^3 \psi_n(\bar{\theta}) - \nabla^3 \psi(\theta)][(\hat{\theta} - \theta) \otimes (\hat{\theta} - \theta) \otimes (\hat{\theta} - \theta)], \end{aligned} \tag{A.1}$$

where $\bar{\theta}$ lies between $\hat{\theta}$ and θ. Next, noticing the Nagar-type expansion

$$[\nabla \psi_n(\theta)]^{-1} = [Q^{-1} + V]^{-1} = Q - QVQ + QVQVQ + \cdots, \tag{A.2}$$

we can solve for $\hat{\theta} - \theta$ as follows, $\hat{\theta} - \theta = a_{-1/2} + a_{-1} + a_{-3/2} + \mathrm{o_P}(n^{-3/2})$, which is as given in (4) and (5). Substituting (5) in (6) and (7) and noting that $d = Q\psi_n$ and $\mathbb{E}(\psi_n) = 0$, we can write (6) as

$$B(\hat{\theta}) = Q\overline{H_1 d} - \tfrac{1}{2}Q\overline{H_2}(\overline{d \otimes d}), \tag{A.3}$$

and MSE in (7) as

$$M(\hat{\theta}) = A_{-1} + A_{-3/2} + A_{-2}, \tag{A.4}$$

where $A_{-s/2} = \mathrm{O}(n^{-s/2})$, $s = 2, 3, 4$, and they are

$$A_{-1} = \overline{dd'},$$

$$A_{-3/2} = -Q[\overline{Vdd'} + \frac{1}{2}\overline{H_2}\overline{(d \otimes d)d'}] - [\overline{dd'V} + \frac{1}{2}\overline{d(d' \otimes d')}\overline{H_2}'],Q,$$

$$\begin{aligned} A_{-2} = {} & Q[\overline{Vdd'V'} + \tfrac{1}{4}\overline{H_2}\overline{(d \otimes d)(d' \otimes d')}\overline{H_2}' - \tfrac{1}{2}\overline{Vd(d' \otimes d')}\overline{H_2}' - \tfrac{1}{2}\overline{H_2}\overline{(d \otimes d)d'V'}]Q \\ & + Q[\overline{VQVdd'} - \tfrac{1}{2}\overline{VQ\overline{H_2}(d \otimes d)d'} + \tfrac{1}{2}\overline{G(d \otimes d)d'} - \tfrac{1}{2}\overline{H_2}\overline{(d \otimes (QVd))d'} \\ & + \tfrac{1}{4}\overline{H_2}\overline{((Q\overline{H_2}(d \otimes d)) \otimes d)d'} - \tfrac{1}{2}\overline{H_2}\overline{((QVd) \otimes d)d'} \\ & + \tfrac{1}{4}\overline{H_2}\overline{(d \otimes (Q\overline{H_2}(d \otimes d)))d'} - \tfrac{1}{6}\overline{H_3}\overline{(d \otimes d \otimes d)d'}] \\ & + [\overline{dd'VQV} - \tfrac{1}{2}\overline{d(d' \otimes d')\overline{H_2}'QV} + \tfrac{1}{2}\overline{d(d' \otimes d')G'} - \tfrac{1}{2}\overline{d((d'VQ) \otimes d')}\overline{H_2}' \\ & + \tfrac{1}{4}\overline{d(d' \otimes ((d' \otimes d')\overline{H_2}'Q))}\overline{H_2}' - \tfrac{1}{2}\overline{d(d' \otimes (d'VQ))}\overline{H_2}' \\ & + \tfrac{1}{4}\overline{d(((d' \otimes d')\overline{H_2}'Q) \otimes d')}\overline{H_2}' - \tfrac{1}{6}\overline{d(d' \otimes d' \otimes d')}\overline{H_3}']Q. \end{aligned}$$

The above results show that the bias and MSE results on the $\sqrt{n}$-consistent estimator $\hat{\theta}$ in Bao and Ullah (2003) go through for a general form $\psi_n(\theta)$ in (3) provided that Assumptions 4–6 hold. However, when these unified results on the bias and MSE are applied to different models, the techniques in deriving the specific results, especially obtaining the expectations involved, can be different for each model. For example, in the spatial model here the $\sqrt{n}$-consistency needs Assumptions 1–3, and the top order invariant polynomial is used to evaluate the expectations involved, which was not needed in the time series models considered in Bao and Ullah (2003).

A.2. Expectations of cross products of quadratic forms in $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$

Utilizing the results of Magnus (1978, 1979), among others, and noting that both M_1 and M_2 are symmetric, we have the following results on $\theta_{ij} = \mathbb{E}[(\varepsilon' M_1 y)^i \cdot (\varepsilon' M_2 \varepsilon)^j]$:

$$\theta_{01} = \sigma^2 \operatorname{tr} M_2,$$

$$\theta_{10} = \sigma^2 \operatorname{tr} M_1,$$

$$\theta_{20} = \sigma^4[(\operatorname{tr} M_1)^2 + 2 \operatorname{tr} M_1^2],$$

$$\theta_{30} = \sigma^6[(\operatorname{tr} M_1)^3 + 6 \operatorname{tr} M_1 \cdot \operatorname{tr} M_1^2 + 8 \operatorname{tr} M_1^3],$$

$$\theta_{40} = \sigma^8[(\operatorname{tr} M_1)^4 + 12(\operatorname{tr} M_1^2)^2 + 12 \operatorname{tr} M_1^2 \cdot (\operatorname{tr} M_1)^2 + 32 \operatorname{tr} M_1 \cdot \operatorname{tr} M_1^3 + 48 \operatorname{tr} M_1^4],$$

$$\theta_{02} = \sigma^4[(\operatorname{tr} M_2)^2 + 2 \operatorname{tr} M_2^2],$$

$$\theta_{11} = \sigma^4[\operatorname{tr} M_1 \cdot \operatorname{tr} M_2 + 2 \operatorname{tr} M_1 M_2],$$

$$\theta_{12} = \sigma^6[\operatorname{tr} M_1 \cdot (\operatorname{tr} M_2)^2 + 8 \operatorname{tr} M_1 M_2^2 + 2 \operatorname{tr} M_1 \cdot (\operatorname{tr} M_2^2) + 4 \operatorname{tr} M_2 \cdot \operatorname{tr} M_1 M_2],$$

$$\theta_{21} = \sigma^6[\operatorname{tr} M_2 \cdot (\operatorname{tr} M_1)^2 + 8 \operatorname{tr} M_2 M_1^2 + 2 \operatorname{tr} M_2 \cdot (\operatorname{tr} M_1^2) + 4 \operatorname{tr} M_1 \cdot \operatorname{tr} M_2 M_1],$$

$$\begin{aligned}\theta_{22} = \sigma^8[&(\operatorname{tr} M_1)^2 \cdot (\operatorname{tr} M_2)^2 + 16 \operatorname{tr} M_1 \cdot \operatorname{tr} M_1 M_2^2 + 16 \operatorname{tr} M_2 \cdot \operatorname{tr} M_1^2 M_2 \\ &+ 4 \operatorname{tr} M_1^2 \cdot \operatorname{tr} M_2^2 + 8(\operatorname{tr} M_1 M_2)^2 + 2 \operatorname{tr} M_1^2 \cdot (\operatorname{tr} M_2)^2 + 2 \operatorname{tr} M_2^2 \cdot (\operatorname{tr} M_1)^2 \\ &+ 8 \operatorname{tr} M_1 \cdot \operatorname{tr} M_2 \cdot \operatorname{tr} M_1 M_2 + 32 \operatorname{tr} M_1^2 M_2^2 + 16 \operatorname{tr} (M_1 M_2)^2],\end{aligned}$$

$$\begin{aligned}\theta_{31} = \sigma^8[&\operatorname{tr} M_2 \cdot (\operatorname{tr} M_1)^3 + 24 \operatorname{tr} M_1 \cdot \operatorname{tr} M_1^2 M_2 + 8 \operatorname{tr} M_2 \cdot \operatorname{tr} M_1^3 + 12 \operatorname{tr} M_1^2 \cdot \operatorname{tr} M_1 M_2 \\ &+ 6 \operatorname{tr} M_1 M_2 \cdot (\operatorname{tr} M_1)^2 + 6 \operatorname{tr} M_1 \cdot \operatorname{tr} M_2 \cdot \operatorname{tr} M_1^2 + 48 \operatorname{tr} M_1^3 M_2].\end{aligned}$$

A.3. Expectations of ratios of cross products of quadratic forms in $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$

From (15), $\lambda_{ij} = \mathbb{E}[(\epsilon' M_1 \epsilon)^i \cdot (\epsilon' M_2 \epsilon)^j / (\epsilon'\epsilon)^{i+j}] = ((\frac{1}{2})_f / (\frac{1}{2}n)_f) C_{[f]}^{k[r]}(A_{[r]}) = [(\frac{1}{2})_{i+j}/(\frac{1}{2}n)_{i+j}] C_{[i+j]}^{i,j}(M_1, M_2)$, where $f = i+j$, $k[r] = i,j$, and $A_{[r]} = M_1, M_2$. Now we utilize Theorem (2.1) of Chikuse (1987) and the algorithm of Smith (1993) to evaluate $C_{[i+j]}^{i,j}(M_1, M_2)$. To calculate all the λs needed, the following are enough for us for a general $C_{[p+q]}^{p,q}(X, Y)$:

(i) $p = 0$,

$$C_{[q]}^{0,q}(X, Y) = C_{[q]}(Y);$$

(ii) $p = 1$,

$$C_{[1+q]}^{1,q}(X, Y) = \frac{1}{2}\left[\frac{q!}{(\frac{1}{2})_{q+1}}\right] \sum_{i=0}^{q} \left[\frac{(\frac{1}{2})_{q-i}}{(q-i)!} \operatorname{tr}(XY^i) \times C_{[q-i]}(Y)\right];$$

(iii) $p = 2$,

$$\begin{aligned} &C_{[2+q]}^{2,q}(X, Y) \\ &= \frac{(\frac{1}{2})_2 (\frac{1}{2})_q}{(\frac{1}{2})_{q+2}} C_{[2]}(X) C_{[q]}(Y) + \frac{1}{4}\left[\frac{q!}{(\frac{1}{2})_{q+2}}\right] \sum_{i=1}^{\eta_1} \left\{ \frac{(\frac{1}{2})_{q-2i}}{(q-2i)!} [\operatorname{tr}(XY^i)]^2 C_{[q-2i]}(Y) \right\} \\ &+ \frac{1}{2}\left[\frac{q!}{(\frac{1}{2})_{q+2}}\right] \sum_{i=1}^{q} \left\{ \frac{(\frac{1}{2})_{q-i}}{(q-i)!} \left[\sum_{j=0}^{i} \operatorname{tr}(XY^j X Y^{i-j}) + \operatorname{tr}(X)\operatorname{tr}(XY^i)\right] C_{[q-i]}(Y) \right\} \\ &+ \frac{1}{2}\left[\frac{q!}{(\frac{1}{2})_{q+2}}\right] \sum_{i=1}^{\eta_2} \sum_{j=i+1}^{q-i} \frac{(\frac{1}{2})_{q-i-j}}{(q-i-j)!} \operatorname{tr}(XY^i)\operatorname{tr}(XY^j) C_{[q-i-j]}(Y), \end{aligned}$$

where in (iii) $\eta_1 = \frac{1}{2}q$ and $\eta_2 = \eta_1 - 1$ if q is even, $\eta_1 = \eta_2 = \frac{1}{2}(q-1)$ if q is odd, and the top oder zonal polynomial $C_{[q]}(Y)$ is given by

$$C_{[q]}(Y) = \frac{q!}{(\frac{1}{2})_q} d_q(Y),$$

in which $d_q(Y)$ follows a recursion

$$d_q(Y) = q^{-1}\sum_{j=0}^{q-1}(q-j)t_{q-j}d_j(Y), \quad t_i = \frac{1}{2}i^{-1}\,\mathrm{tr}(Y^i), \quad d_0(Y) = 1.$$

After simplifications, we put the following results for λ_{ij}, in which $n_{(i)} = \prod_{j=0}^{i-1}(n+2j)$,

$$n_{(1)}\lambda_{01} = \mathrm{tr}\, M_2,$$

$$n_{(1)}\lambda_{10} = \mathrm{tr}\, M_1,$$

$$n_{(2)}\lambda_{20} = 2\,\mathrm{tr}\, M_1^2 + (\mathrm{tr}\, M_1)^2,$$

$$n_{(2)}\lambda_{02} = 2\,\mathrm{tr}\, M_2^2 + (\mathrm{tr}\, M_2)^2,$$

$$n_{(3)}\lambda_{30} = 8\,\mathrm{tr}\, M_1^3 + 6\,\mathrm{tr}\, M_1 \cdot \mathrm{tr}\, M_1^2 + (\mathrm{tr}\, M_1)^3,$$

$$n_{(4)}\lambda_{40} = 48\,\mathrm{tr}\, M_1^4 + 32\,\mathrm{tr}\, M_1 \cdot \mathrm{tr}\, M_1^3 + 12\,\mathrm{tr}\, M_1^2 \cdot (\mathrm{tr}\, M_1)^2 + 12(\mathrm{tr}\, M_1^2)^2 + (\mathrm{tr}\, M_1)^4,$$

$$\begin{aligned} n_{(5)}\lambda_{50} = {} & 384\,\mathrm{tr}\, M_1^5 + 240\,\mathrm{tr}\, M_1 \cdot \mathrm{tr}\, M_1^4 + 160\,\mathrm{tr}\, M_1^2 \cdot \mathrm{tr}\, M_1^3 + 80(\mathrm{tr}\, M_1)^2 \cdot \mathrm{tr}\, M_1^3 \\ & + 60\,\mathrm{tr}\, M_1 \cdot (\mathrm{tr}\, M_1^2)^2 + 20\,\mathrm{tr}\, M_1^2 \cdot (\mathrm{tr}\, M_1)^3 + (\mathrm{tr}\, M_1)^5, \end{aligned}$$

$$\begin{aligned} n_{(6)}\lambda_{60} = {} & 3840\,\mathrm{tr}\, M_1^6 + 2304\,\mathrm{tr}\, M_1 \cdot \mathrm{tr}\, M_1^5 + 1440\,\mathrm{tr}\, M_1^2 \cdot \mathrm{tr}\, M_1^4 \\ & + 960\,\mathrm{tr}\, M_1 \cdot \mathrm{tr}\, M_1^2 \cdot \mathrm{tr}\, M_1^3 + 720(\mathrm{tr}\, M_1)^2 \cdot \mathrm{tr}\, M_1^4 + 640(\mathrm{tr}\, M_1^3)^2 \\ & + 180(\mathrm{tr}\, M_1)^2 \cdot (\mathrm{tr}\, M_1^2)^2 + 160(\mathrm{tr}\, M_1)^3 \cdot \mathrm{tr}\, M_1^3 + 120(\mathrm{tr}\, M_1^2)^3 \\ & + 30(\mathrm{tr}\, M_1)^4 \cdot \mathrm{tr}\, M_1^2 + (\mathrm{tr}\, M_1)^6, \end{aligned}$$

$$n_{(2)}\lambda_{11} = 2\,\mathrm{tr}\, M_1M_2 + \mathrm{tr}\, M_1 \cdot \mathrm{tr}\, M_2,$$

$$n_{(3)}\lambda_{12} = 8\,\mathrm{tr}\, M_1M_2^2 + 4\,\mathrm{tr}\, M_1M_2 \cdot \mathrm{tr}\, M_2 + 2\,\mathrm{tr}\, M_1 \cdot \mathrm{tr}\, M_2^2 + \mathrm{tr}\, M_1 \cdot (\mathrm{tr}\, M_2)^2,$$

$$n_{(3)}\lambda_{21} = 8\,\mathrm{tr}\, M_1^2M_2 + 4\,\mathrm{tr}\, M_1M_2 \cdot \mathrm{tr}\, M_1 + 2\,\mathrm{tr}\, M_2 \cdot \mathrm{tr}\, M_1^2 + \mathrm{tr}\, M_2 \cdot (\mathrm{tr}\, M_1)^2,$$

$$\begin{aligned} n_{(4)}\lambda_{31} = {} & 48\,\mathrm{tr}\, M_1^3M_2 + 24\,\mathrm{tr}\, M_1^2M_2 \cdot \mathrm{tr}\, M_1 + 12\,\mathrm{tr}\, M_1M_2 \cdot \mathrm{tr}\, M_1^2 + 8\,\mathrm{tr}\, M_1^3 \cdot \mathrm{tr}\, M_2 \\ & + 6\,\mathrm{tr}\, M_1 \cdot \mathrm{tr}\, M_2 \cdot \mathrm{tr}\, M_1^2 + 6\,\mathrm{tr}\, M_1M_2 \cdot (\mathrm{tr}\, M_1)^2 + (\mathrm{tr}\, M_1)^3 \cdot \mathrm{tr}\, M_2, \end{aligned}$$

$$\begin{aligned} n_{(4)}\lambda_{22} = {} & 32\,\mathrm{tr}\, M_1^2M_2^2 + 16\,\mathrm{tr}\, M_1 \cdot \mathrm{tr}\, M_1M_2^2 + 16\,\mathrm{tr}\, M_2 \cdot \mathrm{tr}\, M_1^2M_2 + 16\,\mathrm{tr}\, M_1M_2M_1M_2 \\ & + 8\,\mathrm{tr}\, M_1 \cdot \mathrm{tr}\, M_2 \cdot \mathrm{tr}\, M_1M_2 + 8(\mathrm{tr}\, M_1M_2)^2 + 4\,\mathrm{tr}\, M_1^2 \cdot \mathrm{tr}\, M_2^2 \\ & + 2\,\mathrm{tr}\, M_1^2 \cdot (\mathrm{tr}\, M_2)^2 + 2\,\mathrm{tr}\, M_2^2 \cdot (\mathrm{tr}\, M_1)^2 + (\mathrm{tr}\, M_1)^2 \cdot (\mathrm{tr}\, M_2)^2, \end{aligned}$$

$$\begin{aligned} n_{(5)}\lambda_{41} = {} & 384\,\mathrm{tr}\, M_1^4M_2 + 192\,\mathrm{tr}\, M_1 \cdot \mathrm{tr}\, M_1^3M_2 + 96\,\mathrm{tr}\, M_1^2M_2 \cdot \mathrm{tr}\, M_1^2 \\ & + 64\,\mathrm{tr}\, M_1M_2 \cdot \mathrm{tr}\, M_1^3 + 48\,\mathrm{tr}\, M_1^4 \cdot \mathrm{tr}\, M_2 + 48\,\mathrm{tr}\, M_1^2M_2 \cdot (\mathrm{tr}\, M_1)^2 \\ & + 48\,\mathrm{tr}\, M_1 \cdot \mathrm{tr}\, M_1^2 \cdot \mathrm{tr}\, M_1M_2 + 32\,\mathrm{tr}\, M_1 \cdot \mathrm{tr}\, M_1^3 \cdot \mathrm{tr}\, M_2 \\ & + 12(\mathrm{tr}\, M_1)^2 \cdot \mathrm{tr}\, M_1^2 \cdot \mathrm{tr}\, M_2 + 12(\mathrm{tr}\, M_1^2)^2 \cdot \mathrm{tr}\, M_2 \\ & + 8(\mathrm{tr}\, M_1)^3 \cdot \mathrm{tr}\, M_1M_2 + (\mathrm{tr}\, M_1)^4 \cdot \mathrm{tr}\, M_2. \end{aligned}$$

References

Anselin, L., 1980. Estimation methods for spatial autoregressive structures. Regional Science Dissertation and Monograph Series, vol. 8, Cornell University, Ithaca, NY.

Anselin, L., 1982. A note on small sample properties of estimators in a first-order spatial autoregressive model. Environment and Planning 14, 1023–1030.

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

Anselin, L., Florax, R., 1995. Small sample properties of tests for spatial dependence in regression models: some further results. In: Anselin, L., Florax, R. (Eds.), New Directions in Spatial Econometrics. Springer, Berlin.

Bao, Y., Ullah, A., 2003. The second-order bias and mean squared error of estimators in time series models. Working Paper 03-08, Department of Economics, University of California, Riverside.

Chikuse, Y., 1987. Methods for constructing top order invariant polynomials. Econometric Theory 3, 195–207.

Conley, T.G., 1996. Econometric modeling of cross-sectional dependence. Ph.D. Dissertation, Department of Economics, University of Chicago.

Das, D., Kelejian, H.H., Prucha, I.R., 2003. Finite sample properties of estimators of spatial autoregressive models with autoregressive errors. Papers in Regional Science 82, 1–26.

Davis, A.W., 1979. Invariant polynomials with two matrix arguments extending the zonal polynomials: application to multivariate distribution theory. Annals of the Institute of Statistical Mathematics A 31, 465–484.

Davis, A.W., 1980. Invariant polynomials with two matrix arguments, extending the zonal polynomials. In: Krishnaiah, P.R. (Ed.), Multivariate Analysis (V). North-Holland, Amsterdam.

Davis, A.W., 1981. On the construction of a class of invariant polynomials in several matrices, extending the zonal polynomials. Annals of the Institute of Statistical Mathematics A 33, 297–313.

Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics 17, 99–121.

Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533.

Kelejian, H.H., Robinson, D.P., 1993. A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a country expenditure model. Papers in Regional Science 72, 297–312.

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

Magnus, J.R., 1978. The moments of products of quadratic forms in normal variables. Statistica Neerlandica 32, 201–210.

Magnus, J.R., 1979. The expectation of products of quadratic forms in normal variables: the practice. Statistica Neerlandica 33, 131–136.

Nagar, A.L., 1959. The bias and moment matrix of the general *k*-class estimators of the parameters in simultaneous equations. Econometrica 27, 575–595.

Ord, J.K., 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association 70, 120–126.

Rilstone, P., Srivastava, V.K., Ullah, A., 1996. The second-order bias and mean squared error of nonlinear estimators. Journal of Econometrics 75, 369–385.

Smith, M.D., 1989. On the expectations of a ratio of quadratic forms in normal variables. Journal of Multivariate Analysis 31, 244–257.

Smith, M.D., 1993. Expectations of ratios of quadratic forms in normal variables: evaluating some top order invariant polynomials. Australian Journal of Statistics 35, 271–282.