

A General Method for Third-Order Bias and Variance Corrections on a Nonlinear Estimator*

Zhenlin Yang

School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903

emails: zlyang@smu.edu.sg

March, 2012

Abstract

Motivated by a recent study of Bao and Ullah (2007a) on finite sample properties of MLE in the pure SAR (spatial autoregressive) model, a general method for third-order bias and variance corrections on a nonlinear estimator is proposed based on stochastic expansion and bootstrap. Working with concentrated estimating equation simplifies greatly the high-order expansions for bias and variance; a simple bootstrap procedure overcomes a major difficulty in analytically evaluating expectations of various quantities in the expansions. The method is then studied in detail using a more general SAR model, with its effectiveness in correcting bias and improving inference fully demonstrated by extensive Monte Carlo experiments. Compared with the analytical approach, the proposed approach is much simpler and has a much wider applicability.

Key Words: Third-order bias; Third-order variance; Bootstrap; Concentrated estimating equation; Monte Carlo; Spatial layout; Stochastic expansion.

JEL Classification: C10, C15, C21

1 Introduction

Many econometric models share the following common features: (i) there is a nonlinear parameter that is the main source of bias in model estimation and main cause of difficulty in bias correction, (ii) there are many other parameters in the model but their estimates, given this nonlinear parameter, are either unbiased or can be easily bias-corrected, and (iii) the constrained estimates possess analytical expressions, leading to an analytical form for a concentrated estimating equation. These include spatial autoregressive model, spatial panel model with fixed effects, dynamic regression model, dynamic panel model with fixed

*I thank Aman Ullah, Jihai Yu, Yong Bao, two anonymous referees, an associate editor and the Editor Cheng Hsiao for their constructive comments that have led to significant improvements in the paper. The early version of the paper entitled: “*Bias-Corrected Estimation for Spatial Autocorrelation*”, was presented at the IVth World Conference of the Spatial Econometrics Association, 2010, and the Singapore Management University, 2009, and helpful comments were received. I am grateful to the support from a research grant (# C244/MSS9E005) from Singapore Management University and the research assistance from Yan Shen.

effects, Box-Cox regression, Weibull duration model, etc.. The biasness problem arising from the estimation of the nonlinear parameter has been widely recognized and a satisfactory treatment of it has been the main focus of many researchers in the last two decades (see, among others, Kiviet, 1995; Hahn and Kuersteiner, 2002; Hahn and Newey, 2004; Bun and Carree, 2005; Bao and Ullah, 2007a,b; Bao, 2010). Another important issue, the high-order correction on the variance of a bias-corrected estimator, has not been formally addressed.

Stochastic expansion (Rileston et al., 1996; Ullah, 2004) is seen to be a very useful tool for studying the finite sample properties of a nonlinear estimator (Bao and Ullah, 2007a,b; Bao, 2010). However, it has two major limitations in performing high-order bias and variance corrections: (i) it involves high dimension matrix manipulations and (ii) it requires analytical expressions of expectations of various quantities in the expansions, which are either very cumbersome to derive or simply do not even exist. We show in this paper that (i) can be overcome by focusing on the nonlinear parameter and working with the concentrated estimating equation, and (ii) can be overcome by a simple bootstrap procedure.

To illustrate the above ideas, consider first the spatial autoregressive (SAR) model:

$$Y_n = \lambda W_n Y_n + X_n \beta + u_n \quad (1)$$

where Y_n is a vector of observations on n spatial units, X_n is an $n \times p$ matrix of values of k exogenous regressors, W_n is a specified $n \times n$ spatial weights matrix, u_n is a vector of independent and identically distributed (iid) disturbances of zero mean and finite variance σ^2 , λ is a scalar spatial parameter, and β is a $p \times 1$ vector of regression coefficients.¹

Denote $\theta = \{\lambda, \beta', \sigma^2\}'$. The Gaussian log-likelihood function is,

$$\ell_n(\theta) = -\frac{n}{2} \log(2\pi\sigma^2) + \log |A_n(\lambda)| - \frac{1}{2\sigma^2} [A_n(\lambda)Y_n - X_n\beta]' [A_n(\lambda)Y_n - X_n\beta], \quad (2)$$

where $A_n(\lambda) = I_n - \lambda W_n$ and I_n is an $n \times n$ identity matrix. Maximizing $\ell(\theta)$ gives the maximum likelihood estimator (MLE) of θ if the errors are exactly normal, otherwise the quasi-MLE (QMLE). Given λ , the constrained QMLEs of β and σ^2 are

$$\hat{\beta}_n(\lambda) = (X_n' X_n)^{-1} X_n' A_n(\lambda) Y_n \quad \text{and} \quad \hat{\sigma}_n^2(\lambda) = n^{-1} Y_n' A_n'(\lambda) M_n A_n(\lambda) Y_n, \quad (3)$$

where $M_n = I_n - X_n (X_n' X_n)^{-1} X_n'$. These lead to the concentrated log-likelihood of λ as

$$\ell_n^c(\lambda) = -\frac{n}{2} [\log(2\pi) + 1] - \frac{n}{2} \log \hat{\sigma}_n^2(\lambda) + \log |A_n(\lambda)|. \quad (4)$$

Maximizing $\ell_n^c(\lambda)$ gives the unconstrained QMLE $\hat{\lambda}_n$ of λ . The unconstrained QMLEs of β and σ^2 are thus $\hat{\beta}_n \equiv \hat{\beta}_n(\hat{\lambda}_n)$ and $\hat{\sigma}_n^2 \equiv \hat{\sigma}_n^2(\hat{\lambda}_n)$. Write $\hat{\theta}_n = (\hat{\lambda}_n, \hat{\beta}_n', \hat{\sigma}_n^2)'$.

¹For theory and applications, see Cliff and Ord (1973, 1981), Ord (1975), Anselin (1988, 2001), Case (1991), Case, et al. (1993), Besley and Case (1995), Brueckner (1998), Anselin and Bera (1998), Kelejian and Prucha (1998, 1999, 2001), Bell and Bockstael (2000), Bertrand, et al. (2000), Topa (2001), Lee (2002, 2003, 2004a, 2007a,b), Mynbaev and Ullah (2008), Robinson (2010), Su and Jin (2010), Su (2012), etc..

To study the finite sample properties of $\hat{\theta}_n$ following the stochastic expansion approach, one needs to derive analytically the expectations of various quantities involving derivatives of $\ell_n(\theta)$ (up to fourth order for third-order bias and variance corrections). While finding the expectations is not a problem in theory as it involves only quadratic forms of u_n , the dimensionality of the problem (up to $(p+2)^3 \times (p+2)$) greatly complicates the results that in turn hinders their practical tractability (see Bao, 2010, for a second-order bias formula). We note that if λ were known, then $\hat{\beta}_n(\lambda)$ is unbiased and $\hat{\sigma}_n^2(\lambda)$ can be made unbiased by multiplying a factor $n/(n-p)$. This suggests that in estimating the SAR model the main source of bias and the main difficulty in correcting the bias are associated with the estimation of λ . Lee (2007a) made a similar remark based on his Monte Carlo results. Further, given λ the finite sample variances of $\hat{\beta}_n(\lambda)$ and $\hat{\sigma}_n^2(\lambda)$ both possess explicit expressions. Thus, for bias and variance corrections for the SAR model it may be only necessary to focus on the estimation of λ . A multidimensional problem is thus reduced to a scalar one, which greatly simplifies the higher-order stochastic expansions. However, working with the concentrated log-likelihood $\ell_n^c(\lambda)$ makes the analytical derivation harder as it now involves ratios of quadratic forms (see Section 3 for details). Thus, for these expansions to be of a general practical value, they must be supplemented with simple ways for evaluating various expectations involving ratios of quadratic forms.

The above arguments extend directly to all other models of similar features as the SAR model. Take, for example, the Box-Cox transformation model (Box and Cox, 1964): $h(Y_n, \lambda) = X_n\beta + u_n$, where all quantities are defined similarly as the SAR model (1), except that h denotes a known nonlinear monotonic transformation indexed by an unknown transformation parameter λ , applied to Y_n elementwise. The concentrated log-likelihood of λ takes the form $\ell_n^c(\lambda) = -\frac{n}{2}[\log(2\pi) + 1] - \frac{n}{2} \log \hat{\sigma}_n^2(\lambda) + \sum_{i=1}^n \log h_y(Y_{n,i}, \lambda)$, where $\hat{\sigma}_n^2(\lambda) = n^{-1}h'(Y_n, \lambda)M_n h(Y_n, \lambda)$ and $h_y(Y_{n,i}, \lambda) = \partial h_y(Y_{n,i}, \lambda)/\partial Y_{n,i}$. It is clear that the analytical expectations of various quantities involving the derivatives of $\ell_n^c(\lambda)$ are not obtainable, and working with the full likelihood in this case does not solve this problem.

The above discussions show clearly the need for a general method for high-order bias and variance corrections that avoids the analytical derivations of various expectations, and thus works for all models even when the analytical expectations are not obtainable. Noting that the derivatives of $\ell_n^c(\lambda)$ for both the SAR model and the Box-Cox model discussed above can be expressed as functions of the parameter vector θ and the error vector u_n with iid elements, naturally, their expectations can be bootstrapped (see Efron, 1979).

In this paper we present a general method for third-order bias and variance corrections under a fairly general model specification that encompasses all the models mentioned above. The proposed approach is hybrid – combining stochastic expansion and bootstrap, with the former providing tractable approximations to the bias and variance (up to third-order) of a nonlinear estimator, and the latter making these expansions practically implementable. A

key assumption followed in the literature is relaxed, resulting in different bias and variance formulas when concentrated estimating equation is used. The important issue: third-order correction on the standard error of a bias-corrected estimator, is formally studied.

When applied to the SAR model, the proposed approach quickly leads to a complete set of results for third-order bias and variance corrections, which extends Bao and Ullah (2007a) by (i) allowing regressors in the model, (ii) allowing nonnormal errors, and (iii) providing a third-order bias correction on $\hat{\lambda}_n$, and second- and third-order corrections on the variances of $\hat{\lambda}_n$ and the bias-corrected $\hat{\lambda}_n$. Compared with Bao (2010), where only a second-order bias formula for $\hat{\theta}_n$ is derived based on the full likelihood, our method can be viewed as a simpler alternative when only second-order bias correction on $\hat{\lambda}_n$ is concerned. In addition, our method provides a complete set of third-order results, including the third-order variance of the bias-corrected $\hat{\lambda}_n$. More importantly, the proposed approach is much simpler and has a much wider applicability than the analytical approach.

The rest of the paper is organized as follows. Section 2 presents the general method for third-order bias and variance corrections of a general nonlinear estimator. Section 3 presents the main theoretical results corresponding to the SAR model. Section 4 presents Monte Carlo results for the finite sample performance of the proposed method under the SAR model, along with a comparison with the analytical approach of Bao and Ullah (2007a). The effectiveness of the proposed method in correcting bias and improving inference is fully demonstrated. Section 5 concludes the paper and provides discussions on related issues.

2 A General Method for Bias and Variance Corrections

In this section, we first present revised third-order results by relaxing a key assumption, to suit the concentrated estimating equation, and then we outline the main idea of bootstrap method for estimating quantities in the bias and variance formulas.

2.1 Third-Order Bias and Variance of a Nonlinear Estimator

Bao and Ullah (2007a), extending Rilstone et al. (1996), considered a general class of \sqrt{n} -consistent estimators identified by the moment condition or estimating equation

$$\hat{\theta}_n = \arg\{\psi_n(\theta) = 0\}, \quad (5)$$

where $\psi_n(\theta) \equiv \psi_n(Z_n; \theta)$ is a $k \times 1$ vector-valued function of the observable data $Z_n = \{Z_i\}_{i=1}^n$ (iid or non-iid) and a parameter vector θ , of the same dimension as θ , and normalized to have order $O_p(n^{-1/2})$.² They obtained a third-order stochastic expansion for $\hat{\theta}_n$, and a

²This is in fact a generalized version of the well-known M-estimation (maximum likelihood type estimation) of Huber (1964). Obviously, the maximum likelihood or quasi-maximum likelihood, least squares, method of moments, and generalized method of moments are the special cases of this estimation method.

second-order bias and a third-order MSE for $\hat{\theta}_n$, assuming $E\psi_n(\theta) = 0$.

We note that the condition, $E\psi_n(\theta) = 0$, is neither necessary nor true in general for deriving high-order results based on a general estimation equation as in (5). It is required for achieving asymptotic efficiency but not for achieving consistency (see, e.g., Amemiya, 1985; White, 1994). Under the joint estimation, it is usually true if the model is correctly specified and the ML method is followed, but may fail if the model is misspecified. Under the constrained estimation, however, it is generally untrue whichever the estimation method is followed and whether or not the model is correctly specified, except in some special cases, e.g., the pure (no regressors) SAR model with σ^2 unknown (Bao and Ullah, 2007a). Thus, this condition needs to be relaxed, in particular under the constrained estimation framework.

To fix the idea, let $\theta = (\lambda, \alpha)'$ where λ is the scalar nonlinear parameter of which the estimation incurs biasness that is difficult to correct, and given λ the estimation of the parameter vector α has an analytical solution, either unbiased or easily bias-correctable. Let $\hat{\alpha}_n(\lambda)$ be the constrained estimator of α for a given λ value. Let $\theta_0 = (\lambda_0, \alpha_0)'$ be the true value of θ . Partition $\psi_n(\theta)$ according to $(\lambda, \alpha)'$, i.e., $\psi_n(\theta) = \{\psi_{\lambda n}(\lambda, \alpha), \psi'_{\alpha n}(\lambda, \alpha)\}'$. Define $\tilde{\psi}_n(\lambda) \equiv \psi_{\lambda n}(\lambda, \hat{\alpha}_n(\lambda))$. Then, the estimator $\hat{\lambda}_n$ of λ would typically be

$$\hat{\lambda}_n = \arg\{\tilde{\psi}_n(\lambda) = 0\}, \quad (6)$$

with $\tilde{\psi}_n(\lambda) = 0$ being referred to as the *concentrated estimating equation* (CEE), in contrast to the *joint estimating equation* (JEE) embedded in (5).³ Note that by nonlinear we mean the CEE, $\tilde{\psi}_n(\lambda) = 0$, does not have an analytical solution. In this paper, we focus on the cases where λ is a scalar. The CEE looks identical to the JEE when θ is a scalar, and thus the corresponding expansion is expected to have the same form, though the regularity conditions need to be strengthened. However, there is a major difference: the expectation of $\tilde{\psi}_n(\lambda_0)$ may not be zero even if $E\psi_n(\theta_0) = 0$. If $\hat{\alpha}_n(\lambda)$ is \sqrt{n} -consistent, it is typical that $E[\tilde{\psi}_n(\lambda_0)] = O(n^{-1})$, i.e., the expectation goes to zero at an n -rate. If this is true, then $E[\tilde{\psi}_n(\lambda_0)]$ constitutes an important term in the bias correction. In this case, the bias formula need to be modified. As a consequence, the higher-order approximations to the variance needs to be modified as well. The mean squared error (MSE), however, remains in the same form as it directly follows the stochastic expansions for $\hat{\lambda}_n$.

Let $H_{rn}(\lambda) = d^r \tilde{\psi}(\lambda)/d\lambda^r, r = 1, 2, 3$. Let $\tilde{\psi}_n \equiv \tilde{\psi}_n(\lambda_0)$, $H_{rn} \equiv H_{rn}(\lambda_0)$ and $H_{rn}^o = H_{rn} - EH_{rn}, r = 1, 2, 3$. Define $\Omega_n = -E(H_{1n})^{-1}$. Note that here and hereafter the expectation operator corresponds to the true model or the true parameter values θ_0 . Let

³Making inference about the parameter of interest in the presence of many parameters not of direct interest (called the nuisance parameters) is a standard statistical problem, and it is typical in these situations to replace the nuisance parameters by their estimators (constrained) in the object function or the estimating function. There is a vast literature on the satisfactory handling of nuisance parameters. Most of this work focused on the modification of the likelihood function and the concentrated likelihood function. See Laskar and King (1998) for a survey and a comparison of the various methods.

Λ be the parameter space of λ . So far we have not yet specified the form of the $\tilde{\psi}_n(\lambda)$ function, thus as general theories we need some generic smoothness conditions on $\tilde{\psi}_n(\lambda)$.

Assumption A: Λ is compact with λ_0 being an interior point. $E(\tilde{\psi}_n) = O(n^{-1})$, and $\hat{\lambda}_n$, as a solution of $\tilde{\psi}_n(\lambda) = 0$, is a \sqrt{n} -consistent estimator of λ_0 .

Assumption B: $\tilde{\psi}_n(\lambda)$ is differentiable up to r th order for λ in a neighborhood of λ_0 , $E(H_{rn}) = O(1)$, and $H_{rn}^\circ = O_p(n^{-\frac{1}{2}})$, $r = 1, 2, 3$.

Assumption C: $E(H_{1n})^{-1} = O(1)$, and $H_{1n}^{-1} = O_p(1)$.

Assumption D: $|H_{rn}(\lambda) - H_{rn}(\lambda_0)| \leq |\lambda - \lambda_0|U_n$ for λ in a neighborhood of λ_0 , $r = 1, 2, 3$, and $E(|U_n|) < C < \infty$ for some constant C .

The \sqrt{n} -consistency is a standard requirement for a higher-order stochastic expansion. In the context of CEE, the \sqrt{n} -consistency of $\hat{\lambda}_n$ implies $E(\tilde{\psi}_n) = o(n^{-1/2})$ but not zero in general due to the estimation of the nuisance parameters. If the estimators of the nuisance parameters are also \sqrt{n} -consistent, it can be argued that $E(\tilde{\psi}_n) = O(n^{-1})$. Further, the \sqrt{n} -consistency of $\hat{\lambda}_n$ implies $\tilde{\psi}_n = O_p(n^{-\frac{1}{2}})$. The Assumptions B and C are the tightened versions of the Assumptions 4 and 5 in Bao and Ullah (2007a). The conditions $E(H_{rn}) = O(1)$ and $H_{rn}^\circ = O_p(n^{-\frac{1}{2}})$ are needed so that H_{rn} in a relevant term can be replaced by $E(H_{rn})$ with the error $O_p(n^{-\frac{1}{2}})$ being absorbed into the overall error term.⁴ We are ready to state the general theorems. All the proofs are given in Appendix A.

Theorem 2.1: Under Assumptions A-D, we have a third-order stochastic expansion:

$$\hat{\lambda}_n - \lambda_0 = a_{-1/2} + a_{-1} + a_{-3/2} + O_p(n^{-2}), \quad (7)$$

where $a_{-s/2}$ represents terms of order $O_p(n^{-s/2})$ for $s = 1, 2, 3$, and they are $a_{-1/2} = \Omega_n \tilde{\psi}_n$, $a_{-1} = \Omega_n H_{1n}^\circ a_{-1/2} + \frac{1}{2} \Omega_n E(H_{2n})(a_{-1/2}^2)$, and $a_{-3/2} = \Omega_n H_{1n}^\circ a_{-1} + \frac{1}{2} \Omega_n H_{2n}^\circ (a_{-1/2}^2) + \Omega_n E(H_{2n})(a_{-1/2} a_{-1}) + \frac{1}{6} \Omega_n E(H_{3n})(a_{-1/2}^3)$.

The third-order stochastic expansion for $\hat{\lambda}_n$ based on CEE is seen to have an identical form as those in Rilstone et al. (1996) and Bao and Ullah (2007a,b) when there is only one parameter in the model. The same holds for the MSE expansion given below.

Corollary 2.1: Under Assumptions A-D, assume further that a quantity bounded in probability has a finite expectation. We have a third-order expansion for the MSE of $\hat{\lambda}_n$:

$$\text{MSE}(\hat{\lambda}_n) = m_{-1} + m_{-3/2} + m_{-2} + O(n^{-5/2}), \quad (8)$$

where $m_{-s/2} = O(n^{-s/2})$, $s = 2, 3, 4$, with $m_{-1} = E(a_{-1/2}^2)$, $m_{-3/2} = 2E(a_{-1/2} a_{-1})$, and $m_{-2} = 2E(a_{-1/2} a_{-3/2}) + E(a_{-1}^2)$.

⁴Under a specific model with a specific estimation method, these generic conditions may be replaced by a set of weaker and more primitive conditions. The Assumption A may be relaxed to allow for asymptotic (first-order) bias, and our methods can in principle be applied to do higher-order bias reduction for dynamic or nonlinear panel models with fixed effects, see Hahn and Kuersteiner (2002) and Hahn and Newey (2004).

The leading term $m_{-1} = \Omega_n^2 \mathbb{E}(\tilde{\psi}_n^2)$ gives the asymptotic variance of $\hat{\lambda}$, and $m_{-1} + m_{-3/2}$ and $m_{-1} + m_{-3/2} + m_{-2}$ give, respectively, the 2nd- and 3rd-order expansions for $\text{MSE}(\hat{\lambda}_n)$. Turning to the expansions for bias and variance of $\hat{\lambda}_n$, the relaxed Assumption A leads to results that are different from those based on JEE. First, we give the result for bias.

Corollary 2.2: *Under Assumptions A-D, assume further that a quantity bounded in probability has a finite expectation. We have a third-order expansion for the bias of $\hat{\lambda}_n$:*

$$\text{Bias}(\hat{\lambda}_n) = b_{-1} + b_{-3/2} + O(n^{-2}), \quad (9)$$

where $b_{-s/2} = O(n^{-s/2})$, $s = 2, 3$, with $b_{-1} = \mathbb{E}(a_{-1/2} + a_{-1})$ and $b_{-3/2} = \mathbb{E}(a_{-3/2})$.

Thus, b_{-1} alone gives a second-order expansion for the bias of $\hat{\lambda}_n$, and $b_{-1} + b_{-3/2}$ gives a third-order expansion. Note that $\mathbb{E}(a_{-1/2}) = \Omega_n \mathbb{E}(\tilde{\psi}_n)$. This term is $O(n^{-1})$ under CEE, and can be identically zero when JEE is used. Rilstone et al. (1996) and Bao and Ullah (2007a,b) considered in their general theory only second-order expansions for the bias. Their formulas correspond to our b_{-1} term only. Comparing with their second-order expansions for the bias, we see that b_{-1} contains an extra term, $2\Omega_n \mathbb{E}(\tilde{\psi}_n)$. When CEE is used, this term plays a pivotal role in bias and variance corrections. This point is confirmed by the additional Monte Carlo results for the SAR model, available from the author upon request.

Adding a third-order bias-correction term $b_{-3/2}$ into the formula gives us a choice for further improvement on the bias-correction procedure if necessary. With the results of Corollary 2.2, the second- and third-order bias-corrected estimators of λ are, respectively,

$$\hat{\lambda}_n^{\text{bc2}} = \hat{\lambda}_n - \hat{b}_{-1} \quad \text{and} \quad \hat{\lambda}_n^{\text{bc3}} = \hat{\lambda}_n - \hat{b}_{-1} - \hat{b}_{-3/2}. \quad (10)$$

Remark 1: The practical implementation of $\hat{\lambda}_n^{\text{bc3}}$ requires the estimation of $b_{-3/2}$, which greatly complicates the algebraic work and computer coding if the analytical approach is followed, but adds only a little if the bootstrap procedure introduced later is followed.

Remark 2: There is an issue on the validity of replacing b_{-1} by \hat{b}_{-1} for feasible third-order bias correction. Writing $b_{-1} \equiv b(\theta_0)$, then $\hat{b}_{-1} \equiv b(\hat{\theta}_n) = b(\theta_0) + b'_\theta(\theta_0)(\hat{\theta}_n - \theta_0) + O_p(n^{-2})$ provided the derivative $b'_\theta(\theta_0) \sim b(\theta_0) = O(n^{-1})$. Taking expectation on both sides, one sees that $\mathbb{E}(\hat{b}_{-1}) = b(\theta_0) + O(n^{-2})$, showing the validity of the method.

While it is important to have higher-order expansions for $\text{MSE}(\hat{\lambda}_n)$ for the purpose of efficiency comparison, it is more important to have higher-order expansions for the variances of $\hat{\lambda}_n$, $\hat{\lambda}_n^{\text{bc2}}$ and $\hat{\lambda}_n^{\text{bc3}}$ for inference purpose. For $\hat{\lambda}_n$, one is tempted to simply combine the expansions for bias and MSE to give second- and third-order expansions: $\text{Var}(\hat{\lambda}_n) = m_{-1} + m_{-3/2} + O(n^{-2})$, and $\text{Var}(\hat{\lambda}_n) = m_{-1} + m_{-3/2} + m_{-2} - b_{-1}^2 + O(n^{-5/2})$. Theoretically these are correct, but empirically they do not guarantee positiveness of the variance estimator when n is not large. We thus propose to have variance expansions directly out of (7).

Corollary 2.3: *Under Assumptions A-D, assume further that a quantity bounded in probability has a finite expectation. We have a third-order expansion for the variance of $\hat{\lambda}_n$:*

$$\text{Var}(\hat{\lambda}_n) = v_{-1} + v_{-3/2} + v_{-2} + O(n^{-5/2}), \quad (11)$$

where $v_{-1} = \text{Var}(a_{-1/2})$, $v_{-3/2} = 2\text{Cov}(a_{-1/2}, a_{-1})$, and $v_{-2} = \text{Var}(a_{-1} + a_{-3/2}) + 2\text{Cov}(a_{-1/2}, a_{-3/2})$, with $v_{-s/2} = O(n^{-s/2})$, $s = 2, 3, 4$.

The third-order expansions are presented by clearly separating out the terms of different order, thus allowing one to choose between the 2nd- or 3rd-order approximations according to the actual needs. With the results of Corollaries 2.2 and 2.3, one can correct $\hat{\lambda}_n$ and its standard error for an improved inference for λ . However, the bias-corrected estimators, $\hat{\lambda}_n^{\text{bc2}}$ and $\hat{\lambda}_n^{\text{bc3}}$, contain additional random elements due to the estimation of the bias. Therefore, for improving finite sample inference for λ it is more relevant to use $\text{Var}(\hat{\lambda}_n^{\text{bc2}})$ or $\text{Var}(\hat{\lambda}_n^{\text{bc3}})$ to calculate the standard errors of $\hat{\lambda}_n^{\text{bc2}}$ or $\hat{\lambda}_n^{\text{bc3}}$. Letting $b_\theta(\theta_0) \equiv \{b_\lambda(\theta_0), b'_\alpha(\theta_0)\}' \equiv \frac{\partial}{\partial \theta_0} b_{-1}$, we have the following important corollary.

Corollary 2.4: *Under Assumptions A-D, assume further that (i) a quantity bounded in probability has a finite expectation, and (ii) $b_\theta(\theta_0) = O(n^{-1})$, $b_\theta(\hat{\theta}_n) - b_\theta(\theta_0) = O_p(n^{-3/2})$, and $\frac{\partial}{\partial \hat{\theta}_n} \hat{b}_{-3/2} = O_p(n^{-3/2})$. We have the third-order expansions:*

$$\text{Var}(\hat{\lambda}_n^{\text{bc3}}) = (v_{-1} + v_{-3/2} + v_{-2})(1 - 2b_\lambda(\theta_0)) - 2b'_\alpha(\theta_0) \text{ACov}(\hat{\alpha}_n, \hat{\lambda}_n) + O(n^{-5/2}), \quad (12)$$

where ACov denotes asymptotic covariance. Further, $\text{Var}(\hat{\lambda}_n^{\text{bc2}}) = \text{Var}(\hat{\lambda}_n^{\text{bc3}}) + O(n^{-5/2})$.

Thus, the variances of $\hat{\lambda}_n^{\text{bc3}}$ and $\hat{\lambda}_n$ agree only up to second order, suggesting that for improving finite sample inference for λ , $\hat{\lambda}_n^{\text{bc3}}$ should be used in conjunction with $\text{Var}(\hat{\lambda}_n^{\text{bc3}})$.

2.2 A bootstrap method for estimating the bias and variance corrections

The second- or third-order corrections on the bias and variance of nonlinear estimators are practically tractable only if one could find a simple way to estimate the quantities like $E(H_{rn})$, $E(\tilde{\psi}_n^2)$, $E(H_{1n}\tilde{\psi}_n^2)$, etc. The analytical approach is to first find these expectations and then replace θ in the resulted expressions by its consistent estimator $\hat{\theta}_n$. However, finding these expectations analytically is either very cumbersome or impossible (see discussions in Sec. 5). Thus, alternative methods are highly desirable. We now introduce a simple bootstrap method for estimating these quantities. Consider a general model of the form

$$g(Z_n, \theta_0) = u_n$$

where u_n is the disturbance vector of iid (not necessarily normal) components. Assume that the key quantities $\tilde{\psi}_n$ and H_{rn} can be expressed as $\tilde{\psi}_n \equiv \tilde{\psi}_n(u_n, \theta_0)$ and $H_{rn} \equiv H_{rn}(u_n, \theta_0)$, $r = 1, 2, 3$. Let $\hat{u}_n = g(Z_n, \hat{\theta}_n)$ be the vector of estimated residuals based on

the data. Resample the elements of \hat{u}_n (by making n random draws with replacement) to give $\hat{u}_{n,b}$, and compute $\tilde{\psi}_n(u_{n,b}, \hat{\theta}_n)$ and $H_{rn}(u_{n,b}, \hat{\theta}_n)$. Repeat these steps for $b = 1, 2, \dots, B$ times, and the bootstrap estimates of these expected quantities are given as,

$$\widehat{\mathbb{E}}(\tilde{\psi}_n^i H_{rn}^j) = \frac{1}{B} \sum_{b=1}^B \tilde{\psi}_n^i(u_{n,b}, \hat{\theta}_n) H_{rn}^j(u_{n,b}, \hat{\theta}_n), \quad i, j = 1, 2, \dots, \quad r = 1, 2, 3,$$

leading to the bootstrap estimates of biases, MSEs, and variances. $b_\theta(\theta_0)$ in Corollary 2.4 can be estimated by numerical derivatives and bootstrap. See Sections 3 and 4 for details.

Note that in the entire bootstrap process, the same estimate $\hat{\theta}_n$ based on the original data is used when recalculating $\tilde{\psi}_n$ and H_{rn} based on each bootstrap sample $\hat{u}_{n,b}$. The reestimation of the model parameter θ is thus avoided, which makes this bootstrap procedure easy to be implemented. In summary, our proposed approach for bias and variance corrections takes the advantages of both stochastic expansions and bootstrap, neither of which alone allows us to handle a problem of this type comfortably. The usefulness and effectiveness of this approach is fully demonstrated in the following section using the SAR model, and the validity of this approach is justified therein.

3 Bias and Variance Corrections for SAR Model

We now consider the estimation of spatial lag parameter λ in the general SAR model specified in (1), to give a detailed demonstration of the applications of the general methods presented in the earlier section. The nature of the SAR model indeed renders it a special attention in terms of bias and variance corrections. First, the parameter λ enters into the model in a nonlinear manner, hence the estimation of it is likely to incur bias. Second, the degree of spatial dependence among the spatial units depends not only on the magnitude of the spatial parameter λ , but also on the number of neighbors each spatial unit has, or equivalently the number of non-zero elements that each row of the W_n matrix contains. A very important special case of this is that the number of neighbors, h_n say, grows with n (see, e.g., Case, 1991), and in this case, Lee (2004a) showed that the QML estimators of λ and β are no longer \sqrt{n} -consistent, but rather $\sqrt{n/h_n}$ -consistent. Thus, the effective sample size is n/h_n , and the bias and variance formulas given above need to be adjusted to allow for this possibility. Conceptually, this may be fairly straightforward as one may simply replace n by h_n/n everywhere in the expansion formulas. Theoretically, however, much needs to be done in terms of regularity conditions and formal proofs of the results. We do so in this paper by following the theoretical foundations laid out in Lee (2004a).

Bao and Ullah (2007a) made their first attempt to address the biasness issue by working with the pure SAR model with normal errors, and provided analytical formulas for the second-order bias and mean squared error (MSE) for the MLE $\hat{\lambda}_n$ based on the stochastic

expansion technique first introduced by Rilestone et al. (1996). Their results, though limited to the pure SAR model with normal errors, shed much light on a general solution to the biasness problem of the general SAR model (or a class of similar models). Bao (2010) followed up with this issue through the full likelihood and derived an analytical second-order bias for the QMLE $\hat{\theta}_n$, but the MSE formula and the third-order results were not given.

In dealing with the case of σ^2 unknown in the pure SAR model, Bao and Ullah (2007a, p.400) advocate the use of concentrated likelihood function of λ as (i) it simplifies the maximization procedure substantially, and (ii) it also simplifies the derivations for the higher-order results since it is much easier to work with a scalar case than with a vector. We concur with their view and stress further that (i) these simplifications are even greater if the SAR model involves exogenous regressors, and (ii) for the purpose of bias correction, λ is the parameter of primary interest as, given λ , the model reduces to a linear regression, and the constrained QMLEs $\hat{\beta}_n(\lambda)$ and $\hat{\sigma}_n^2(\lambda)$ is either unbiased or can be made unbiased.

3.1 The main results

The QMLE $\hat{\lambda}_n$ of the spatial parameter λ , which maximizes the concentrated log-likelihood function $\ell_n^c(\lambda)$ given in (4), can be equivalently defined as $\hat{\lambda}_n = \arg\{\tilde{\psi}_n(\lambda) = 0\}$, where $\tilde{\psi}_n(\lambda)$ is the derivative of $\frac{h_n}{n}\ell_n^c(\lambda)$ and has the form,

$$\tilde{\psi}_n(\lambda) = -h_n T_{0n}(\lambda) + h_n R_{1n}(\lambda), \quad (13)$$

with its r th derivative, $H_{rn}(\lambda) = \frac{d^r}{d\lambda^r} \tilde{\psi}_n(\lambda)$, $r = 1, 2, 3$, given as follows

$$h_n^{-1} H_{1n}(\lambda) = -T_{1n}(\lambda) - R_{2n}(\lambda) + 2R_{1n}^2(\lambda), \quad (14)$$

$$h_n^{-1} H_{2n}(\lambda) = -2T_{2n}(\lambda) - 6R_{1n}(\lambda)R_{2n}(\lambda) + 8R_{1n}^3(\lambda), \quad (15)$$

$$h_n^{-1} H_{3n}(\lambda) = -6T_{3n}(\lambda) + 6R_{2n}^2(\lambda) - 48R_{1n}^2(\lambda)R_{2n}(\lambda) + 48R_{1n}^4(\lambda), \quad (16)$$

where $T_{rn}(\lambda) = n^{-1}\text{tr}(G_n^{r+1}(\lambda))$, $r = 0, 1, 2, 3$, $G_n(\lambda) = W_n A_n^{-1}(\lambda)$,⁵

$$R_{1n}(\lambda) = \frac{Y_n' A_n'(\lambda) M_n W_n Y_n}{Y_n' A_n'(\lambda) M_n A_n(\lambda) Y_n} \quad \text{and} \quad R_{2n}(\lambda) = \frac{Y_n' W_n' M_n W_n Y_n}{Y_n' A_n'(\lambda) M_n A_n(\lambda) Y_n}. \quad (17)$$

Clearly the function $\tilde{\psi}_n(\lambda)$ defined in (13) leads to a concentrated estimating equation, and fits into the general framework described in Section 2. The difference is that the quantity h_n may alter the rate of convergence of $\hat{\lambda}_n$ in the first-order asymptotics (Lee, 2004a) and of course the magnitude of the quantities in the higher-order asymptotics. Paralleled with the general theories given in Section 2. we now present a complete and rigorous treatment for the SAR model, taking into account of the possibility that h_n is unbounded.

⁵The author is very grateful to Jihai Yu and a referee for pointing out errors in the expressions for $H_{2n}(\lambda)$ and $H_{3n}(\lambda)$. The corrections, though correspond to the higher-order terms, do lead to further improved and more coherent results in that $\hat{\lambda}_n^{\text{bc3}}$ performs consistently better than $\hat{\lambda}_n^{\text{bc2}}$. See Yang (2010a).

Recall $\tilde{\psi}_n = \tilde{\psi}_n(\lambda_0)$ and $H_{rn} = H_{rn}(\lambda_0), r = 1, 2, 3$. Similarly, let $A_n = A_n(\lambda_0)$, $G_n = G_n(\lambda_0)$, $T_{rn} = T_{rn}(\lambda_0), r = 0, 1, 2, 3$, and $R_{rn} = R_{rn}(\lambda_0), r = 1, 2$. Let $\eta_n = G_n X_n \beta_0$. With the specification of the SAR model and the quasi-maximum likelihood estimation method, the generic regularity conditions listed in Section 2 can be made more specific or more primitive. First, the set of rather primitive conditions of Lee (2004a) for the $\sqrt{n/h_n}$ -consistency of the QMLE $\hat{\lambda}_n$ are essential and are summarized below in Assumptions 1-6.

Assumption 1: *The true λ_0 is in the interior of a compact set Λ .*

Assumption 2: *The innovations $\{u_{n,i}\}$ are iid with mean zero and variance σ^2 . $E|u_{n,i}|^{4+\gamma}$ exists for some $\gamma > 0$.*

Assumption 3: *The elements $w_{n,ij}$ of W_n are at most of order h_n^{-1} uniformly for all i and j , where the rate sequence $\{h_n\}$ can be bounded or divergent but satisfying $h_n^{1+\epsilon}/n \rightarrow 0$ for some $\epsilon > 0$ as $n \rightarrow \infty$. As a normalization, $w_{n,ii} = 0$ for all i . Furthermore, the matrices $\{W_n\}$ are uniformly bounded in both row and column sums.*

Assumption 4: *The matrix A_n is nonsingular, $\{A_n^{-1}\}$ are uniformly bounded in both row and column sums, and $\{A_n^{-1}(\lambda)\}$ are uniformly bounded in either row or column sums, uniformly in $\lambda \in \Lambda$.*

Assumption 5: *The elements of the $n \times p$ matrix X_n are uniformly bounded for all n , and $\lim_{n \rightarrow \infty} \frac{1}{n} X_n' X_n$ exists and is nonsingular.*

Assumption 6: *The elements of $M_n \eta_n$ are $O(h_n^{-\frac{1}{2}})$ uniformly, and $\lim_{n \rightarrow \infty} \frac{h_n}{n} \eta_n' M_n \eta_n = c$, where either $c > 0$, or $c = 0$ but $\lim_{n \rightarrow \infty} \frac{h_n}{n} (\ln |\sigma_0^2 A_n^{-1} A_n'^{-1}| - \ln |\sigma_n^2(\lambda) A_n^{-1}(\lambda) A_n'^{-1}(\lambda)|) \neq 0$, whenever $\lambda \neq \lambda_0$, where $\sigma_n^2(\lambda) = \frac{\sigma_0^2}{n} \text{tr}[A_n'^{-1} A_n(\lambda) A_n(\lambda) A_n^{-1}]$.*

The Assumptions 1-6 listed above are the Assumptions 1, 2, 3', 4-7, and 10 of Lee (2004a). Under these assumptions, the QMLE $\hat{\lambda}_n$ is a $\sqrt{n/h_n}$ -consistent estimator of λ_0 . In the regular case where h_n is bounded, i.e., the degree of spatial dependence does not grow with the sample size, $\hat{\lambda}_n$ becomes \sqrt{n} -consistent. These assumptions lead to the first-order asymptotics for $\hat{\theta}_n$, which are shown to be essential as well for the higher-order stochastic expansions for $\hat{\lambda}_n$, and the higher-order expansions for the bias, MSE, and variance of $\hat{\lambda}_n$. Some further conditions are needed for ensuring proper orders of R_{1n} and R_{2n} , which are crucial for the proper behaviors of the derivatives $H_{rn}, r = 1, 2, 3$. Denote $\hat{\sigma}_{n0}^2 \equiv \hat{\sigma}_n^2(\lambda_0)$.

Assumption 7: *(i) $E[\frac{h_n}{n} (Y_n' A_n' M_n W_n Y_n) (\bar{\sigma}_{n0}^{-4} - \sigma_0^{-4}) (\hat{\sigma}_{n0}^2 - \sigma_0^2)] = O((\frac{h_n}{n})^{\frac{1}{2}})$; and (ii) $E[\frac{h_n}{n} (Y_n' W_n' M_n W_n Y_n) (\bar{\sigma}_{n0}^{-4} - \sigma_0^{-4}) (\hat{\sigma}_{n0}^2 - \sigma_0^2)] = O((\frac{h_n}{n})^{\frac{1}{2}})$, for $\bar{\sigma}_{n0}^2$ between σ_0^2 and $\hat{\sigma}_{n0}^2$.*

These conditions are weak as under the earlier assumptions $\frac{h_n}{n} (Y_n' A_n' M_n W_n Y_n) = O_p(1)$, $\frac{h_n}{n} (Y_n' W_n' M_n W_n Y_n) = O_p(1)$, $\hat{\sigma}_{n0}^2 - \sigma_0^2 = O_p(n^{-\frac{1}{2}})$, and $\bar{\sigma}_{n0}^{-4} - \sigma_0^{-4} = O_p(n^{-\frac{1}{2}})$; see Appendix B for their proofs. Thus, the two quantities inside the expectation sign are both $O_p(n^{-1})$, suggesting that the conditions are met as long as the expectations of these two quantities do not 'explode' beyond the order $O((\frac{h_n}{n})^{\frac{1}{2}})$. To ensure the proper stochastic behavior of $H_{rn}, r = 1, 2, 3$, the following conditions are needed.

Assumption 8: (i) $h_n^s \mathbb{E}[(R_{1n} - \mathbb{E}R_{1n})^s] = O((\frac{h_n}{n})^{\frac{1}{2}})$, $s = 2, 3, 4$; (ii) $h_n^2 \mathbb{E}[(R_{2n} - \mathbb{E}R_{2n})^2] = O((\frac{h_n}{n})^{\frac{1}{2}})$; and (iii) $h_n^{s+1} \mathbb{E}[(R_{1n} - \mathbb{E}R_{1n})^s (R_{2n} - \mathbb{E}R_{2n})] = O((\frac{h_n}{n})^{\frac{1}{2}})$, $s = 1, 2$.

The three conditions in Assumption 8 are in fact rather weak since, following the results of Lemma 3.1 below, all the random quantities inside the expectation sign are of order $O_p(\frac{h_n}{n})$ or lower, which suggests the conditions are met as long as their expectations do not ‘explode’ beyond the order $O((\frac{h_n}{n})^{\frac{1}{2}})$. We have the following important lemma.

Lemma 3.1: *Under Assumptions 1-7, we have*

- (i) $h_n R_{1n} = O_p(1)$, $\mathbb{E}(h_n R_{1n}) = O(1)$, and $h_n R_{1n} = \mathbb{E}(h_n R_{1n}) + O_p((\frac{h_n}{n})^{\frac{1}{2}})$;
(ii) $h_n R_{2n} = O_p(1)$, $\mathbb{E}(h_n R_{2n}) = O(1)$, and $h_n R_{2n} = \mathbb{E}(h_n R_{2n}) + O_p((\frac{h_n}{n})^{\frac{1}{2}})$.

With this lemma and under the conditions listed above, we are able to prove the following theorem and corollaries. These theorem and corollaries parallel the general theorem and corollaries given in Section 2 with the order of magnitude of each term adjusted to allow for the possibility that h_n increases with n . All proofs are given in Appendix B.

Theorem 3.1: *Under Assumptions 1-8, we have a third-order stochastic expansion:*

$$\hat{\lambda}_n - \lambda_0 = a_{-1/2} + a_{-1} + a_{-3/2} + O_p((n/h_n)^{-2}), \quad (18)$$

where $a_{-1/2} = \Omega_n \tilde{\psi}_n$, $a_{-1} = \Omega_n \tilde{\psi}_n + \Omega_n^2 H_{1n} \tilde{\psi}_n + \frac{1}{2} \Omega_n^3 \mathbb{E}(H_{2n}) \tilde{\psi}_n^2$, and

$$a_{-3/2} = \Omega_n \tilde{\psi}_n + 2\Omega_n^2 H_{1n} \tilde{\psi}_n + \Omega_n^3 \mathbb{E}(H_{2n}) \tilde{\psi}_n^2 + \Omega_n^3 H_{1n}^2 \tilde{\psi}_n + \frac{1}{2} \Omega_n^3 H_{2n} \tilde{\psi}_n^2 \\ + \frac{3}{2} \Omega_n^4 \mathbb{E}(H_{2n}) H_{1n} \tilde{\psi}_n^2 + \frac{1}{2} \Omega_n^5 \mathbb{E}(H_{2n})^2 \tilde{\psi}_n^3 + \frac{1}{6} \Omega_n^4 \mathbb{E}(H_{3n}) \tilde{\psi}_n^3,$$

having stochastic orders $a_{-s/2} = O_p((n/h_n)^{-s/2})$, $s = 1, 2, 3$; and $\Omega_n = -1/\mathbb{E}(H_{1n})$.

Note that $\{a_{-s/2}\}$ have the same expressions as those in Theorem 2.1. The difference is in their stochastic orders. To simplify the presentations and to facilitate the practical implementations of our results, define $c_{1n} = \{\Omega_n, 0'_{6 \times 1}\}'$, $c_{2n} = \{\Omega_n, \Omega_n^2, \frac{1}{2} \Omega_n^3 \mathbb{E}(H_{2n}), 0'_{4 \times 1}\}'$, $c_{3n} = \{\Omega_n, 2\Omega_n^2, \Omega_n^3 \mathbb{E}(H_{2n}), \Omega_n^3, \frac{1}{2} \Omega_n^3, \frac{3}{2} \Omega_n^4 \mathbb{E}(H_{2n}), \frac{1}{2} \Omega_n^5 \mathbb{E}(H_{2n})^2 + \frac{1}{6} \Omega_n^4 \mathbb{E}(H_{3n})\}'$, $C_{2n} = c_{1n} + c_{2n}$ and $C_{3n} = c_{1n} + c_{2n} + c_{3n}$. Let

$$\zeta_n = \{\tilde{\psi}_n, H_{1n} \tilde{\psi}_n, \tilde{\psi}_n^2, H_{1n}^2 \tilde{\psi}_n, H_{2n} \tilde{\psi}_n^2, H_{1n} \tilde{\psi}_n^2, \psi_n^3\}'. \quad (19)$$

Then, $a_{-1/2} = c'_{1n} \zeta_n$, $a_{-1} = c'_{2n} \zeta_n$, $a_{-3/2} = c'_{3n} \zeta_n$, and $\hat{\lambda}_n - \lambda_0 = C'_{3n} \zeta_n + O_p((n/h_n)^{-2})$.

Corollary 3.1: *Under Assumptions 1-8, assume further that a quantity bounded in probability has a finite expectation. We have a third-order expansion for the bias of $\hat{\lambda}_n$:*

$$\text{Bias}(\hat{\lambda}_n) = b_{-1} + b_{-3/2} + O((n/h_n)^{-2}), \quad (20)$$

where $b_{-1} = C'_{2n} \mathbb{E}(\zeta_n)$ and $b_{-3/2} = c'_{3n} \mathbb{E}(\zeta_n)$; and 2nd- and 3rd-order bias-corrected QMLEs:

$$\hat{\lambda}_n^{\text{bc}2} = \hat{\lambda}_n - \hat{C}'_{2n} \hat{\mathbb{E}}(\zeta_n) \quad \text{and} \quad \hat{\lambda}_n^{\text{bc}3} = \hat{\lambda}_n - \hat{C}'_{3n} \hat{\mathbb{E}}(\zeta_n). \quad (21)$$

where the quantity with an $\hat{\cdot}$ denotes the estimate of the corresponding quantity.

Comparing with the second-order bias formula of Bao and Ullah (2007a) for the pure SAR model, we see that our second-order bias formula b_{-1} contains an extra term $2\Omega_n \mathbf{E}(\tilde{\psi}_n)$, which is of order $O((n/h_n)^{-1})$ in general but vanishes under the pure SAR model with normal errors. If the second-order bias-correction is not enough, our approach provides an easy way for the third-order correction. In contrast, the analytical approach involves tremendous amount of extra algebraic work even with a simple model of normal errors.

Corollary 3.2. *Under Assumptions 1-8, assume further that a quantity bounded in probability has a finite expectation. We have a third-order expansion for the MSE of $\hat{\lambda}_n$:*

$$\text{MSE}(\hat{\lambda}_n) = m_{-1} + m_{-3/2} + m_{-2} + O((n/h_n)^{-5/2}), \quad (22)$$

for $m_{-1} = c'_{1n} \mathbf{E}(\zeta_n \zeta'_n) c_{1n}$, $m_{-3/2} = 2c'_{1n} \mathbf{E}(\zeta_n \zeta'_n) c_{2n}$, $m_{-2} = c'_{2n} \mathbf{E}(\zeta_n \zeta'_n) c_{2n} + 2c'_{1n} \mathbf{E}(\zeta_n \zeta'_n) c_{3n}$.

It is seen that introducing the notation c_{in} and ζ_n greatly simplifies the expression for the MSE expansion. Simplification is even greater for the variance expansions as seen from the following corollaries, which makes the practical implementations much easier.

Corollary 3.3. *Under Assumptions 1-8, assume further that a quantity bounded in probability has a finite expectation. We have a third-order expansion for the variance of $\hat{\lambda}_n$:*

$$\text{Var}(\hat{\lambda}_n) = v_{-1} + v_{-3/2} + v_{-2} + O((n/h_n)^{-2}); \quad (23)$$

where $v_{-1} = c'_{1n} \text{Var}(\zeta_n) c_{1n}$, $v_{-1} + v_{-3/2} = C'_{2n} \text{Var}(\zeta_n) C_{2n}$, and $v_{-1} + v_{-3/2} + v_{-2} = C'_{3n} \text{Var}(\zeta_n) C_{3n}$.

Note that $v_{-1} = \text{Var}(a_{1/2})$, $v_{-1} + v_{-3/2} = \text{Var}(a_{-1/2} + a_{-1})$, and $v_{-1} + v_{-3/2} + v_{-2} = \text{Var}(a_{-1/2} + a_{-1} + a_{-3/2})$. Certain terms such as $\text{Var}(a_{-1})$ for second-order expansion, and $\text{Cov}(a_{-1}, a_{-3/2})$ and $\text{Var}(a_{-3/2})$ for the third-order expansion may be dropped. These lead to asymptotically equivalent but simpler variance expansions $\text{Var}(\hat{\lambda}_n) = \text{MSE}(\hat{\lambda}_n) + O((\frac{n}{h_n})^{-2})$, and $\text{Var}(\hat{\lambda}_n) = \text{MSE}(\hat{\lambda}_n) + b_{-1}^2 + O((\frac{n}{h_n})^{-5/2})$. However, as pointed out in Section 2, these simplifications may not guarantee the positiveness of the variance estimates, thus it is recommended that the results in (23) be followed in the practical applications.

One important issue left is the variances of the bias-corrected estimators, i.e., $\text{Var}(\hat{\lambda}_n^{\text{bc}2})$ and $\text{Var}(\hat{\lambda}_n^{\text{bc}3})$, which are important and more relevant in improving the finite sample inference for λ . Let $\{b'_{\beta}(\theta_0), b_{\lambda}(\theta_0), b_{\sigma^2}(\theta_0)\}' \equiv b_{\theta}(\theta_0) = \frac{\partial}{\partial \theta_0} b_{-1}$. We have the following result.

Corollary 3.4. *Under Assumptions 1-8, assume further that (i) a quantity bounded in probability has a finite expectation, (ii) $b_{\theta}(\theta_0) = O(\frac{h_n}{n})$, (iii) $b_{\theta}(\hat{\theta}_n) - b_{\theta}(\theta_0) = O_p((\frac{h_n}{n})^{3/2})$, and (iv) $\frac{\partial}{\partial \theta_n} \hat{b}_{-3/2} = O_p((\frac{h_n}{n})^{3/2})$. We have the third-order expansions:*

$$\begin{aligned} \text{Var}(\hat{\lambda}_n^{\text{bc}3}) &= C'_{3n} \text{Var}(\zeta_n) C_{3n} (1 - 2b_{\lambda}(\theta_0)) - 2b'_{\beta}(\theta_0) \text{ACov}(\hat{\beta}_n, \hat{\lambda}_n) \\ &\quad - 2b_{\sigma^2}(\theta_0) \text{ACov}(\hat{\sigma}_n^2, \hat{\lambda}_n) + O((n/h_n)^{-5/2}), \end{aligned} \quad (24)$$

and $\text{Var}(\hat{\lambda}_n^{\text{bc}2}) = \text{Var}(\hat{\lambda}_n^{\text{bc}3}) + O((\frac{n}{h_n})^{-5/2})$, where ‘ACov’ denotes the asymptotic covariance.

It turns out that the variance of $\hat{\lambda}_n$ agrees with the variances of $\hat{\lambda}_n^{\text{bc}2}$ and $\hat{\lambda}_n^{\text{bc}3}$ to the second-order $O(n^{-3/2})$ but differ on the third-order $O(n^{-2})$. The analytical expressions for the asymptotic covariances are available from Lee (2004a). The partial derivatives $b'_\beta(\theta_0)$, $b_\lambda(\theta_0)$, and $b_{\sigma^2}(\theta_0)$ can be estimated by numerical derivatives (see Section 4 for details).

With the results of Corollaries 3.1-3.4, various t -ratios for inference for λ can be formed:

$$t_{ij} = \frac{\hat{\lambda}_n^{\text{bc}i} - \lambda_0}{\sqrt{\hat{V}_j(\hat{\lambda}_n^{\text{bc}i})}}, \quad (25)$$

where $\hat{\lambda}_n^{\text{bc}i}$ denotes the i th-order bias-corrected estimator, and $\hat{V}_j(\cdot)$ the estimated j th-order corrected variance, $i, j = 1, 2, 3$. Clearly, $i = 1$ and $j = 1$ correspond to the original QMLE $\hat{\lambda}_n$ and its asymptotic variance. Monte Carlo results given in Section 4 show that the usual t -ratio, t_{11} , leads to the worst results, whereas the fully third-order corrected t -ratio, t_{33} , leads to the best inferences for λ .

3.2 The bootstrap method for practical implementation

While working with the concentrated estimating equation greatly simplifies various expansions, it does not overcome the difficulty in analytically evaluating the expectations of various quantities in the expansions. From the expressions given in (13) to (24) we see that, in order to calculate various expectations in the bias, MSE, and variance expansions, all we need is to find the expectations of R_{1n} and R_{2n} (the ratios of linear-quadratic and quadratic forms), their powers, and the cross-products of powers, and then evaluate them at the estimated parameter values. In other words we need to derive analytically,

$$E(R_{1n}^k), k = 1, \dots, 10; \quad E(R_{2n}^k), k = 1, \dots, 4; \quad \text{and} \quad E(R_{1n}^k R_{2n}^m), k = 1, \dots, 6, m = 1, 2,$$

and then replace θ_0 by $\hat{\theta}_n$, for up to third-order expansions. However, this is either too cumbersome if the errors are normal, or too difficult if the errors are nonnormal, unless for the special case of a pure SAR model with normal errors.⁶ For the general SAR model with normal errors, we managed to derive only a second-order bias formula (requiring only $E(R_{1n}^k), k = 1, 2, 3$, $E(R_{2n})$ and $E(R_{1n}R_{2n})$) by extending the results of Smith (1993). However, the results are too tedious to be fit into the current paper. In the case of the general SAR model with nonnormal errors, we failed to obtain any analytical results.

⁶In this case, the high-order bias and variance formulas involve $E\{(u'_n A_1 u_n)^i (u'_n A_2 u_n)^j / (u'_n u_n)^{i+j}\}$ for symmetric matrices A_1 and A_2 , and $i, j = 0, 1, 2, \dots$, which can be found using the results of Smith (1993). For this simple model, the resulted analytical expressions (Bao and Ullah, 2007a) are complicated but manageable. Either dropping normality or adding exogenous regressors or both invalidate these results. Working with the full likelihood function makes it possible for an analytical solution but at the expense of more tedious expressions; see Bao (2010) for a second-order bias formula for $\hat{\theta}_n$. As we are primarily interested in the finite sample properties of $\hat{\lambda}_n$, we thus work with the concentrated likelihood function.

Thus, for the higher-order results presented above to be practically feasible for a general SAR model, it is highly desirable to have an alternative way to evaluate these expectations. Clearly, it is when the errors are non-normal and the model contains regressors that gives a practical attraction. To solve this puzzle, the bootstrap procedure outlined in Section 2 is made explicit below. Note that the two key ratios can be written as:

$$R_{1n} \equiv R_{1n}(u_n, \theta_0) = \frac{u_n' M_n G_n u_n + u_n' M_n \eta_n}{u_n' M_n u_n},$$

$$R_{2n} \equiv R_{2n}(u_n, \theta_0) = \frac{u_n' G_n' M_n G_n u_n + 2u_n' G_n' M_n \eta_n + \eta_n' M_n \eta_n}{u_n' M_n u_n},$$

where $\eta_n = G_n X_n \beta_0$. These show that $\tilde{\psi}_n = \tilde{\psi}_n(u_n, \theta_0)$ and $H_n = \{H_{1n}, H_{2n}, H_{3n}\} = H_n(u_n, \theta_0)$. Hence, $\zeta_n = \zeta_n(u_n, \theta_0)$. In other words, all the random quantities in the bias, MSE, and variance formulas can be expressed in terms of u_n and θ_0 . This leads naturally to a bootstrap procedure for estimating the expected values of these random quantities (see, e.g., Efron, 1979; Amemiya, 1985, p. 135). The suggested bootstrap procedure is:

1. Compute the QMLEs $\hat{\theta}_n = (\hat{\beta}_n', \hat{\sigma}_n^2, \hat{\lambda}_n)'$ based on the original data,
2. Compute QML residuals $\hat{u}_n = \hat{A}_n Y_n - X_n \hat{\beta}_n$, where $\hat{A}_n = I_n - \hat{\lambda}_n W_n$,
3. Resample \hat{u}_n in the usual way, and denote the resampled vector by $u_{n,b}^*$,
4. Compute $R_{1n}(u_{n,b}^*, \hat{\theta}_n)$ and $R_{2n}(u_{n,b}^*, \hat{\theta}_n)$, and thus $H_n(u_{n,b}^*, \hat{\theta}_n)$, and $\zeta_n(u_{n,b}^*, \hat{\theta}_n)$,
5. Repeat steps (3)-(4) B times to give sequences of bootstrapped values for H_n and ζ_n .

The bootstrap estimates of various expectations thus follow. For example, the bootstrap estimates of the mean and variance of $\tilde{\psi}_n^2$ (the third element of ζ_n) are, respectively, $\widehat{E}(\tilde{\psi}_n^2) = \frac{1}{B} \sum_{b=1}^B \tilde{\psi}_n^2(u_{n,b}^*, \hat{\theta}_n)$, and $\widehat{\text{Var}}(\tilde{\psi}_n^2) = \frac{1}{B} \sum_{b=1}^B \tilde{\psi}_n^4(u_{n,b}^*, \hat{\theta}_n) - [\widehat{E}(\tilde{\psi}_n^2)]^2$. With the quantities $c_{in}, i = 1, 2, 3$, and ζ_n introduced below Theorem 3.1, the practical implementations of our bootstrap-based bias and variance corrections (whether 2nd-order or 3rd-order) can be made much simpler – all that are needed are the following bootstrap estimates:

$$\widehat{E}(H_n) = \frac{1}{B} \sum_{b=1}^B H_n(u_{n,b}^*, \hat{\theta}_n),$$

$$\widehat{E}(\zeta_n) = \frac{1}{B} \sum_{b=1}^B \zeta_n(u_{n,b}^*, \hat{\theta}_n), \quad \text{and}$$

$$\widehat{E}(\zeta_n \zeta_n') = \frac{1}{B} \sum_{b=1}^B \zeta_n(u_{n,b}^*, \hat{\theta}_n) \zeta_n'(u_{n,b}^*, \hat{\theta}_n).$$

We now provide some formal arguments for the validity of the proposed bootstrap procedure. As all the quantities involved in the expansions are smooth functions of either R_{1n} only or both R_{1n} and R_{2n} . It suffices to argue that the above bootstrap procedure leads to valid estimates of $E(R_{1n})$ and $E(R_{2n})$. Let \mathcal{F} be the CDF of u_{ni} , and $\hat{\mathcal{F}}_n$ be the empirical distribution function (EDF) of \hat{u}_n . Consistency of $\hat{\theta}_n$ ensures the consistency of

using $\hat{\mathcal{F}}_n$ to estimate \mathcal{F} . In the real world, if one knew \mathcal{F} and θ_0 , one may approximate $E(R_{1n})$ and $E(R_{2n})$ to an arbitrary accuracy by

$$\frac{1}{M} \sum_{m=1}^M R_{1n}(u_{n,m}, \theta_0) \quad \text{and} \quad \frac{1}{M} \sum_{m=1}^M R_{2n}(u_{n,m}, \theta_0),$$

where $u_{n,m}$ is a random n -vector drawn from \mathcal{F} and M is a large positive integer. If one knew \mathcal{F} but not θ_0 , one may estimate $E(R_{1n})$ and $E(R_{2n})$ by

$$\begin{aligned} \bar{E}(R_{1n}) &= \left(\frac{1}{M} \sum_{m=1}^M R_{1n}(u_{n,m}, \theta_0) \right) \Big|_{\theta_0 = \hat{\theta}_n} = \frac{1}{M} \sum_{m=1}^M R_{1n}(u_{n,m}, \hat{\theta}_n), \\ \bar{E}(R_{2n}) &= \left(\frac{1}{M} \sum_{m=1}^M R_{2n}(u_{n,m}, \theta_0) \right) \Big|_{\theta_0 = \hat{\theta}_n} = \frac{1}{M} \sum_{m=1}^M R_{2n}(u_{n,m}, \hat{\theta}_n). \end{aligned}$$

In reality, however, one knows neither θ_0 nor \mathcal{F} . Under the bootstrap world, these unknown quantities are made ‘known’ to be their estimates, and the bootstrap DGP that mimics the real world DGP is

$$Y_{n,b}^* = \hat{\lambda}_n W_n Y_{n,b}^* + X_n \hat{\beta}_n + u_{n,b}^*,$$

where $u_{n,b}^* \stackrel{iid}{\sim} \hat{\mathcal{F}}_n$, and $(\hat{\beta}_n, \hat{\lambda}_n)$ are the estimates of (β_0, λ_0) based on the original data. Based on the generated bootstrap data $\{Y_{n,b}^*, X_n, W_n\}$ and the bootstrap parameters $(\hat{\beta}_n, \hat{\lambda}_n)$, one computes the bootstrap analogue of R_{1n} and R_{2n} defined in (17) as

$$R_{1n,b}^* = \frac{Y_{n,b}^{*'} \hat{A}'_n M_n W_n Y_{n,b}^*}{Y_{n,b}^{*'} \hat{A}'_n M_n \hat{A}_n Y_{n,b}^*} \quad \text{and} \quad R_{2n,b}^* = \frac{Y_{n,b}^{*'} W_n M_n W_n Y_{n,b}^*}{Y_{n,b}^{*'} \hat{A}'_n M_n \hat{A}_n Y_{n,b}^*},$$

which after simplifications become $R_{1n}(u_{n,b}^*, \hat{\theta}_n)$ and $R_{2n}(u_{n,b}^*, \hat{\theta}_n)$, the bootstrap analogue of $R_{1n}(u_n, \theta_0)$ and $R_{2n}(u_n, \theta_0)$ given above. These give the bootstrap estimates of $E(R_{1n})$ and $E(R_{2n})$ as $E^*[R_{1n}(u_{n,b}^*, \hat{\theta}_n)]$ and $E^*[R_{2n}(u_{n,b}^*, \hat{\theta}_n)]$, respectively, where E^* denotes expectation with respect to the EDF $\hat{\mathcal{F}}_n$. In practical applications, feasible bootstrap estimates can be obtained by taking a large number of bootstrap samples and averaging:

$$\hat{E}(R_{1n}) = \frac{1}{B} \sum_{b=1}^B R_{1n}(u_{n,b}^*, \hat{\theta}_n) \quad \text{and} \quad \hat{E}(R_{2n}) = \frac{1}{B} \sum_{b=1}^B R_{2n}(u_{n,b}^*, \hat{\theta}_n).$$

Thus, the bootstrap estimates $\hat{E}(R_{1n})$ and $\hat{E}(R_{2n})$ have identical structure as the Monte Carlo estimates $\bar{E}(R_{1n})$ and $\bar{E}(R_{2n})$ assuming a known distribution. The difference is between $u_{n,m}$ and $u_{n,b}^*$ or between \mathcal{F} and $\hat{\mathcal{F}}_n$. As argued above, if $\hat{\theta}_n$ is \sqrt{n} -consistent for θ_0 , it is typical (under regularity conditions) that $\hat{\mathcal{F}}_n$ is \sqrt{n} -consistent for \mathcal{F} . This suggests that as estimates of $E(R_{1n})$, the errors resulted from using $\bar{E}(R_{1n})$ and $\hat{E}(R_{1n})$ are of the same order of magnitude. Finally, the existence of various expectations involving R_{1n} and R_{2n} is guaranteed by Lemma 3.1 and Assumption 8 (see the proof of Theorem 3.1). From these, we conclude that our bootstrap method provides valid estimates of various quantities involved in the expansions for the bias of $\hat{\lambda}_n$ and for the variances of $\hat{\lambda}_n$ and bias-corrected $\hat{\lambda}_n$. See, e.g., Efron (1979) and Lahiri (2003) for details on the general bootstrap principles.

4 Monte Carlo Simulation

Extensive Monte Carlo experiments are carried out to investigate (i) the finite sample performance of the QMLE $\hat{\lambda}_n$ and the bias-corrected QMLEs $\hat{\lambda}_n^{\text{bc}2}$ and $\hat{\lambda}_n^{\text{bc}3}$ of the spatial lag parameter λ , (ii) the finite sample performance of the corrected standard errors (se), and (iii) the impact of the bias and se corrections on the subsequent inferences for λ . Also, a comparison is made with the analytical approach of Bao and Ullah (2007a).

For (i)-(iii), Monte Carlo experiments is carried out based on the following SAR model:

$$Y_n = \lambda W_n Y_n + \beta_0 1_n + X_{n1} \beta_1 + X_{n2} \beta_2 + u_n,$$

where 1_n is an n -vector of ones. For all the Monte Carlo experiments, $\beta' = \{\beta_0, \beta_1, \beta_2\}$ is set at $\{5, 1, 1\}$ or $\{.5, .1, .1\}$, σ at 1 or 2, λ takes values $\{.5, .25, 0, -.25, -.5\}$, and n takes values $\{50, 100, 200, 500\}$.⁷ Several ways of generating W_n , (X_{n1}, X_{n2}) , and u_n are considered. First, the values $\{x_{1i}\}$ or $\{x_{1,ir}\}$ of X_{n1} , and the values $\{x_{2i}\}$ or $\{x_{2,ir}\}$ of X_{n2} are,

$$\begin{aligned} \text{MRSAR-A: } \{x_{1i}\} &\stackrel{iid}{\sim} N(0, 1)/\sqrt{2}, \text{ and } \{x_{2i}\} \stackrel{iid}{\sim} N(0, 1)/\sqrt{2}, \text{ or} \\ \text{MRSAR-B: } \{x_{1,ir}\} &= (2z_r + z_{ir})/\sqrt{7}, \text{ and } \{x_{2,ir}\} = (v_r + v_{ir})/\sqrt{7}, \end{aligned}$$

where in MRSAR-B, $\{z_r, z_{ir}, v_r, v_{ir}\} \stackrel{iid}{\sim} N(0, 1)$, across all i and r . Apparently, MRSAR-A gives iid X values, and MRSAR-B gives non-iid X values, or different group means under group interaction, see Lee (2004a) and below for details. The two schemes give signal-to-noise ratios 1 when $\beta_1 = \beta_2 = \sigma = 1$. Partial results with $\beta = \{5, 1, 1\}'$ and $\sigma = 1$ are reported. More extensive results are available at <http://www.mysmu.edu/faculty/zlyang/>.

Spatial layouts. Three general spatial layouts are considered in the Monte Carlo experiments. The first is based on Rook contiguity, the second is based on Queen contiguity and the third is based on the notion of group interactions. The methods used in generating these three spatial layouts are similar to those used in Yang (2010b).

The detail for generating the W_n matrix under rook contiguity is as follows: (i) index the n spatial units by $\{1, 2, \dots, n\}$, randomly permute these indices and then allocate them into a lattice of $k \times m (\geq n)$ squares, (ii) let $W_{ij} = 1$ if the index j is in a square which is on immediate left, or right, or above, or below the square which contains the index i , otherwise $W_{ij} = 0, i, j = 1, \dots, n$, to form an $n \times n$ matrix, and (iii) divide each element of this matrix by its row sum to give W_n . Similarly, one generates the W_n matrix under Queen contiguity with additional neighbors sharing a common vertex with the unit of interest.

⁷As in Lee (2007a), the maximization of $\ell_n^c(\lambda)$ is performed globally without imposing a restricted lower bound on λ . This is important when the true λ value is negative and big, because QMLE is downward biased and a restricted lower bound, -0.9999 say, would result in the searching process to hit the lower bound quite often, thus failing to reach the true maximum point. This would in turn give a wrong impression that the QMLE can be upward biased and the bias-correction may not work in certain cases. This is believed to be the reason for the incoherent Monte Carlo results of Bao and Ullah (2007a). See Anselin (1988, p. 78-79) for a theoretical discussion on the parameter space of λ in relation to the eigenvalues of W_n .

To generate the W_n matrix according to the group interaction scheme, suppose we have k groups of sizes m_1, m_2, \dots, m_k . Define $W_n = \text{diag}\{W_j/(m_j - 1), j = 1, \dots, k\}$, a matrix formed by placing the submatrices W_j along the diagonal direction, where W_j is an $m_j \times m_j$ matrix with ones on the off-diagonal positions and zeros on the diagonal positions. The group sizes $\{m_j\}$ can be the same or different, and independent or dependent on n , allowing for a full range of spatial scenarios considered in Lee (2004a). The details are as follows: (i) calculate the number of groups according to $k = K(n)$, and the approximate average group size $m = n/k$, (ii) generate the group sizes (m_1, m_2, \dots, m_k) according to a discrete distribution centered at m , and (iii) adjust the group sizes so that $\sum_{j=1}^k m_j = n$.⁸

In our Monte Carlo experiments, we use $K(n) = \text{Round}(n^\epsilon)$ with $\epsilon = 0.35, 0.50$, and 0.75 , representing respectively the situations where (a) there are few groups of many spatial units in each, (b) the number of groups and the sizes of the groups are of the same magnitude, and (c) there are many groups of few elements in each. Clearly, $h_n = O(n^{1-\epsilon})$. The group sizes are drawn from a discrete uniform distribution from $0.5m$ to $1.5m$.

Error distributions. To generate $u_n = \sigma e_n$, three distributions are considered: **dgp1**: the elements $\{e_i\}$ of e_n are iid standard normal, **dgp2**: $\{e_i\}$ are iid standardized normal mixture, and **dgp3**: $\{e_i\}$ are iid standardized log-normal. Specifically, for **dgp2**,

$$e_i = ((1 - \xi_i)Z_i + \xi_i\tau Z_i)/(1 - \pi + \pi * \sigma^2)^{0.5}, i = 1, \dots, n,$$

where $\{\xi_i\} \stackrel{iid}{\sim} \text{Bernoulli}(\pi)$, and $\{Z_i\} \stackrel{iid}{\sim} N(0, 1)$ independent of $\{\xi_i\}$. The parameter π represents the proportion of mixing the two normal populations. In our experiments, we choose $\pi = 0.1$, meaning that 90% of the random variates are from standard normal and the remaining 10% are from another normal population with standard deviation τ . We choose $\tau = 4$ to simulate the situation where there are gross errors in the data. For **dgp3**,

$$e_i = [\exp(Z_i) - \exp(0.5)]/[\exp(2) - \exp(1)]^{0.5}, i = 1, \dots, n,$$

which gives an error distribution that is both skewed and leptokurtic. The normal mixture gives an error distribution that is still symmetric like normal but leptokurtic. Other non-normal distributions, such as normal-gamma mixture and chi-square, are also considered and the results (available from the author upon request) exhibit a similar pattern.

Finite sample performance of bias and se corrections. We report the Monte Carlo means, rmses and sds of $\hat{\lambda}_n$, $\hat{\lambda}_n^{\text{bc2}}$ and $\hat{\lambda}_n^{\text{bc3}}$ under various combinations of the values for (n, λ, σ) , the error distributions, and the spatial layouts. We also report the averages

⁸Clearly, this design covers the scenario considered in Case (1991). Typical forms of $K(n)$ include $K(n) = n/m$ where m is a prespecified constant independent of n and $K(n) = \text{Round}(n^\epsilon)$. Lee (2007b) shows that the group size variation plays an important role in the identification and estimation of econometric models with group interactions, contextual factors and fixed effects. Yang (2010b) shows that it also plays an important role in the robustness of the LM test of spatial error components.

(over Monte Carlo samples) of the 1st-, 2nd- and 3rd-order ses: $\widehat{V}_1(\hat{\lambda}_n)^{\frac{1}{2}}$, $\widehat{V}_2(\hat{\lambda}_n)^{\frac{1}{2}}$, $\widehat{V}_3(\hat{\lambda}_n)^{\frac{1}{2}}$ and $\widehat{V}_3(\hat{\lambda}_n^{\text{bc3}})^{\frac{1}{2}}$, defined in (25) and calculated based on the proposed bootstrap method.⁹ Each set of results is based on 10,000 Monte Carlo samples, and $999 + \text{floor}(n^{0.75})$ bootstrap samples for each Monte Carlo sample. Table 1 summarizes partial results with $\beta = \{5, 1, 1\}'$ and $\sigma = 1$. From the results (also unreported), some general observations are in order:

- (i) the bias-corrected QMLEs $\hat{\lambda}_n^{\text{bc2}}$ and $\hat{\lambda}_n^{\text{bc3}}$ are in general nearly unbiased and clearly outperform the original QMLE $\hat{\lambda}_n$;
- (ii) $\hat{\lambda}_n$ is always downward biased and the biasness can be very serious depending on the spatial layout, the sample size and the error standard deviation;
- (iii) $\hat{\lambda}_n^{\text{bc3}}$ improves over $\hat{\lambda}_n^{\text{bc2}}$, but using $\hat{\lambda}_n^{\text{bc2}}$ seems to be sufficient under most of the situations as far as bias-correction is concerned;
- (iv) spatial layouts can have a huge impact on the finite sample performance of $\hat{\lambda}_n$ – the stronger the spatial dependence the worse $\hat{\lambda}_n$ performs;
- (v) the values of σ and the slope parameters also have a big impact – the bigger the σ is, or the smaller the $|\beta_1|$ and $|\beta_2|$ are, the bigger are the biases, rmses and sds of $\hat{\lambda}_n$;
- (vi) the value of λ and the way the regressors being generated affect the finite sample performance of $\hat{\lambda}_n$ – as λ decreases, the bias of $\hat{\lambda}_n$ decreases under iid regressors but increases under non-iid regressors, whereas the [rmse](se) of $\hat{\lambda}_n$ always increases as λ decreases, with a sharper amount for the case of non-iid regressors;
- (vii) the error distribution does not affect much on the general performance of the three estimators, showing the robustness of the proposed approach.
- (viii) The empirical sd of $\hat{\lambda}_n$ can be slightly different from that of $\hat{\lambda}_n^{\text{bc3}}$ when sample size is small, suggesting that the variances of $\hat{\lambda}_n$ and $\hat{\lambda}_n^{\text{bc3}}$ may differ on higher-order term (see panels (a)-(c), Table 1). The results in the last four columns of Table 1 show that $\widehat{V}_3(\hat{\lambda}_n^{\text{bc3}})$ provides the best approximation to the variance of $\hat{\lambda}_n^{\text{bc3}}$. The empirical sds of $\hat{\lambda}_n^{\text{bc2}}$ and $\hat{\lambda}_n^{\text{bc3}}$ agree closely, suggesting that the finite sample variances of $\hat{\lambda}_n^{\text{bc2}}$ and $\hat{\lambda}_n^{\text{bc3}}$ are about the same. These are consistent with the result of Corollary 3.4.

In summary, the proposed bias-correction procedure works excellently in general, it is simple and widely applicable, and thus should be recommended for the practitioners.

The performance of t -ratios. The finite sample behavior of the t -ratios t_{ij} for testing $H_0 : \lambda = 0$, defined in (25) are investigated. Partial Monte Carlo results in terms of means, sds, and tail probabilities are reported in Table 2. From the results, the following conclusions can be drawn: (i) the asymptotic t -ratio t_{11} can perform quite badly with severe distortions

⁹The partial derivatives of $b_{-1} \equiv b_{-1}(\theta_0)$ needed in $\widehat{V}_3(\hat{\lambda}_n^{\text{bc3}})^{\frac{1}{2}}$ are estimated by numerical derivatives: $(b_{-1}(\hat{\theta}_n + \varepsilon \iota_i) - b_{-1}(\hat{\theta}_n))/\varepsilon, i = 1, \dots, k + 2$, where ι_i is a $k + 2$ column vector with 1 on its i th position and zero elsewhere, and ε is taken to be 0.0001 in our experiments.

on mean and sizes; (ii) use of second-order bias-corrected estimator only (t_{21}) immediately improves; (iii) use of the second-order bias-corrected estimator and its second-order variance gives further improvements; and (iv) use of the third-order bias-corrected estimator and its third-order variance estimate gives the best results.

A comparison with Bao and Ullah (2007a). The Monte Carlo experiment conducted by Bao and Ullah (2007a) is replicated and extended for two purposes: (i) to compare our bootstrap-based bias correction with their analytical bias correction, and (ii) to investigate the cause of incoherent Monte Carlo results of Bao and Ullah (2007a) for the case of $\lambda = -0.9$ and $J = 10$, where J is the number of neighbors each spatial unit has when the ‘circular-world’ design is used.

The empirical means and rmses for the three estimators: the MLE $\hat{\lambda}_n$, the second-order bias-corrected MLE $\hat{\lambda}_n^{\text{bc2}}$, and the analytically bias-corrected MLE $\hat{\lambda}_n^{\text{BU}}$ of Bao and Ullah (2007a), are summarized in Table 3. The results show that the second-order bias correction works excellently in general: (i) in case of normal errors, the bootstrap method and analytical method produce almost identical results, confirming the validity of the proposed bootstrap method, and (ii) in case of non-normal errors, bootstrap method produces slightly better results, showing that the proposed method is more robust against the error distribution than the analytical method that is based on the normality assumption. The results also show a very coherent behavior of the MLE and the bias-corrected MLEs in that the MLE is almost unbiased when J is small and is downward biased when J is big. In cases where the MLE is biased, the bias-corrected estimators are always able to correct the bias.

However, the Monte Carlo results of Bao and Ullah (2007a) show a different picture for the cases where J is large and λ is large and negative: the MLE is upward biased and the bias-correction does not work. We believe the reason for this is that a lower bound $c(> -1)$ was given when searching for $\hat{\lambda}_n$ numerically, which ‘raises’ the average value of $\hat{\lambda}_n$ in the Monte Carlo simulation (see also Footnote 7). However, the underlining theories are sound.

5 Conclusions and Discussions

To address the biasness issue in a model containing nonlinear, location as well as scale parameters, one can focus on the estimation of the nonlinear parameter and use the concentrated estimating equation to obtain higher-order expansions to achieve bias and variance corrections. This turns a multidimensional problem to a single dimension and greatly simplifies the higher-order expansions. It is argued that for these abstract formulas to be practically useful, it is necessary to have a feasible method for estimating the various expectations in the formulas. Thus, a simple bootstrap procedure is introduced. These ideas and methods are explored in a full detail in the context of a spatial autoregressive model. Monte Carlo results show that this approach is very effective in that it almost eliminates the bias of

the QMLE, which can be quite large when spatial dependence is strong. Method for third-order correction on the variance of a bias-corrected QMLE is provided as well. With both bias and variance corrections, inferences for the spatial parameter can be greatly improved. In the special case of a pure SAR model where second-order analytical bias-correction is available (Bao and Ullah, 2007a), the proposed approach provides almost identical results when errors are normal, and slightly better results when errors are nonnormal.

We have emphasized through out the paper the tow important aspects of the proposed method: *simplicity* and *generality*. That is, in cases where the analytical approach is feasible, the proposed approach provides a simpler solution that works equally well; in cases where the analytical approach is infeasible, the proposed approach again provides a simple and satisfactory solution. The latter aspect is further demonstrated using the well-known Box-Cox regression model in the introduction. Another example where the analytical approach is infeasible but the proposed one is may be the dynamic panel regression with short time periods and endogenous initial observations (see Hsiao, et. al., 2002, or Hsiao, 2003). The advantage of the proposed approach can be further seen by extending the SAR model to panel with fixed individual effects: our results can be easily extended over, but the analytical results of Bao (2010) cannot be easily done due to the lack of independence in the transformed disturbances. In summary, the approach proposed in this paper offers a general solution for a class of problems that cannot be solved, or cannot be easily solved by the analytical approach, and thus should be recommended to the practitioners.

A problem of immediate interest may be the one involving a vector λ , i.e., there are two or more nonlinear parameters that are the main source of bias in model estimation and the main cause of difficulty in bias-correction. While the dimension is still small and the use of concentrated estimating equation still simplifies the problem in a great deal, it does bring a scalar problem back to a multidimensional one. An immediate example of this is the SAR model where the disturbances also follow a SAR process. Another example is the spatial panel model with fixed effects of Lee and Yu (2010). Our methods can in principle be further generalized to allow for asymptotic (first-order) bias. Typical models of both features are the panel models (dynamic or nonlinear) with fixed effects, and in these cases, it would be interesting to extend our methods to give higher-order bias correction to the problems considered in Hahn and Kuersteiner (2002) and to offer an alternative to the jackknife and analytical bias reduction method of Hahn and Newey (2004) which is based on an iid data set-up. We plan to pursue both issues in future research.

A referee has raised two intriguing issues which we are unable to address in this paper: one is to provide theoretical interpretation for the downward biasness of the QMLE of λ , and the other is to compare the high-order analytical bias derived from the CEE with the one derived from the JEE considered in Bao (2010). As a rigorous study on either issue can be quite involved, we plan to pursue these two issues in future research.

Appendix A: Proofs of the Results in Section 2

Proof of Theorem 2.1: Assumption B allows the Taylor expansions of $\tilde{\psi}_n(\hat{\lambda}_n) = 0$ around λ_0 to be carried out up to third-order, and Assumptions A and D guarantee that the errors in the Taylor approximations are of order $O_p(n^{-1})$, $O_p(n^{-3/2})$, and $O_p(n^{-2})$, respectively, for the 1st-, 2nd- and 3rd-order Taylor expansions. We thus have,

$$\begin{aligned} 0 &= \tilde{\psi}_n + H_{1n}(\hat{\lambda}_n - \lambda_0) + O_p(n^{-1}), \\ 0 &= \tilde{\psi}_n + H_{1n}(\hat{\lambda}_n - \lambda_0) + \frac{1}{2}H_{2n}(\hat{\lambda}_n - \lambda_0)^2 + O_p(n^{-3/2}), \\ 0 &= \tilde{\psi}_n + H_{1n}(\hat{\lambda}_n - \lambda_0) + \frac{1}{2}H_{2n}(\hat{\lambda}_n - \lambda_0)^2 + \frac{1}{6}H_{3n}(\hat{\lambda}_n - \lambda_0)^3 + O_p(n^{-2}), \end{aligned}$$

which give, as $-H_{1n}^{-1} = O_p(1)$ from Assumption C,

$$\hat{\lambda}_n - \lambda_0 = -H_{1n}^{-1}\tilde{\psi}_n + O_p(n^{-1}), \quad (\text{A-1})$$

$$\hat{\lambda}_n - \lambda_0 = -H_{1n}^{-1}\tilde{\psi}_n - \frac{1}{2}H_{1n}^{-1}H_{2n}(\hat{\lambda}_n - \lambda_0)^2 + O_p(n^{-3/2}), \quad (\text{A-2})$$

$$\hat{\lambda}_n - \lambda_0 = -H_{1n}^{-1}\tilde{\psi}_n - \frac{1}{2}H_{1n}^{-1}H_{2n}(\hat{\lambda}_n - \lambda_0)^2 - \frac{1}{6}H_{1n}^{-1}H_{3n}(\hat{\lambda}_n - \lambda_0)^3 + O_p(n^{-2}) \quad (\text{A-3})$$

Under Assumptions B and C, $\Omega_n = -E(H_{1n})^{-1} = O(1)$, $H_{1n}^{-1} = O_p(1)$, and $H_{1n}^\circ = H_{1n} - E(H_{1n}) = O_p(n^{1/2})$. These conditions lead to the following result

$$-H_{1n}^{-1} = (\Omega_n^{-1} - H_{1n}^\circ)^{-1} = (1 - \Omega_n H_{1n}^\circ)^{-1} \Omega_n = \Omega_n + \Omega_n^2 H_{1n}^\circ + \Omega_n^3 H_{1n}^{\circ 2} + O_p(n^{-3/2}),$$

which reduces to $-H_{1n}^{-1} = \Omega_n + \Omega_n^2 H_{1n}^\circ + O_p(n^{-1})$, or $= \Omega_n + O_p(n^{-1/2})$. Substituting $-H_{1n}^{-1} = \Omega_n + O_p(n^{-1/2})$ into (A-1) gives a first-order stochastic expansion for $\hat{\lambda}_n$,

$$\hat{\lambda}_n - \lambda_0 = \Omega_n \tilde{\psi}_n + O_p(n^{-1}) = a_{-1/2} + O_p(n^{-1}). \quad (\text{A-4})$$

Substituting (A-4) into (A-2) for $\hat{\lambda}_n - \lambda_0$, and replacing $-H_{1n}^{-1}$ in the first two terms of (A-2) respectively by $\Omega_n + \Omega_n^2 H_{1n}^\circ + O_p(n^{-1})$ and $\Omega_n + O_p(n^{-1/2})$, we obtain,

$$\hat{\lambda}_n - \lambda_0 = a_{-1/2} + a_{-1} + O_p(n^{-3/2}), \quad (\text{A-5})$$

where $a_{-1} = \Omega_n H_{1n}^\circ a_{-1/2} + \frac{1}{2} \Omega_n E(H_{2n})(a_{-1/2}^2)$. Finally, substituting (A-4) and (A-5) into (A-3) for $\hat{\lambda}_n - \lambda_0$ in the 3rd and 2nd terms, respectively, and replacing $-H_{1n}^{-1}$ in the first three terms of (A-3) respectively by $\Omega_n + \Omega_n^2 H_{1n}^\circ + \Omega_n^3 H_{1n}^{\circ 2} + O_p(n^{-3/2})$, $\Omega_n + \Omega_n^2 H_{1n}^\circ + O_p(n^{-1})$, and $\Omega_n + O_p(n^{-1/2})$, we obtain a third-order stochastic expansion for $\hat{\lambda}_n$,

$$\hat{\lambda}_n - \lambda_0 = a_{-1/2} + a_{-1} + a_{-3/2} + O_p(n^{-2}), \quad (\text{A-6})$$

where $a_{-3/2} = \Omega_n H_{1n}^\circ a_{-1} + \frac{1}{2} \Omega_n H_{2n}^\circ (a_{-1/2}^2) + \Omega_n E(H_{2n})(a_{-1/2} a_{-1}) + \frac{1}{6} \Omega_n E(H_{3n})(a_{-1/2}^3)$.

Proof of Corollary 2.1: We have $\text{MSE}(\hat{\lambda}_n) = E[(a_{-1/2} + a_{-1} + a_{-3/2} + O_p(n^{-2}))^2]$, which simplifies to $\text{MSE}(\hat{\lambda}_n) = m_{-1} + m_{-3/2} + m_{-2} + O(n^{-5/2})$, where $m_{-1} = E(a_{-1/2}^2)$, $m_{-3/2} = 2E(a_{-1/2}a_{-1})$, and $m_{-2} = E(2a_{-1/2}a_{-3/2} + a_{-1}^2)$, following the assumption in the corollary: a quantity bounded in probability has a finite expectation.

Proof of Corollary 2.2: We have $b_{-1} = E(a_{-1/2}) + E(a_{-1})$, and $b_{-3/2} = E(a_{-3/2})$. The result follows as expected error term is $O(n^{-2})$ by the assumption in the corollary.

Proof of Corollary 2.3: Straightforward from the proofs of Corollaries 2.1 and 2.2.

Proof of Corollary 2.4: The additional assumptions stated in the Corollary ensure that $\text{Var}(\hat{\lambda}_n^{\text{bc3}}) = \text{Var}(\hat{\lambda}_n - \hat{b}_{-1} - \hat{b}_{-3/2}) = \text{Var}(\hat{\lambda}_n) - 2\text{Cov}(\hat{\lambda}_n, \hat{b}_{-1}) + O_p(n^{-5/2})$. The mean value theorem gives $\hat{b}_{-1} = b(\hat{\theta}_n) = b(\theta_0) + b_\theta(\theta_0)(\hat{\theta}_n - \theta_0) + [b_\theta(\bar{\theta}_n) - b_\theta(\theta_0)](\hat{\theta}_n - \theta_0)$, where $\bar{\theta}_n$ lies between $\hat{\theta}_n$ and θ_0 . The rest is straightforward.

Appendix B: Proofs of the Results in Section 3

To prove the results of Section 3, we need the following lemmas.

Lemma B.1 (Kelejian and Prucha, 1999; Lee, 2002): *Let $\{A_n\}$ and $\{B_n\}$ be two sequences of $n \times n$ matrices that are uniformly bounded in both row and column sums. Let C_n be a sequence of conformable matrices whose elements are uniformly $O(h_n^{-1})$. Then*

- (i) *the sequence $\{A_n B_n\}$ are uniformly bounded in both row and column sums,*
- (ii) *the elements of A_n are uniformly bounded and $\text{tr}(A_n) = O(n)$, and*
- (iii) *the elements of $A_n C_n$ and $C_n A_n$ are uniformly $O(h_n^{-1})$.*

Lemma B.2 (Lee, 2004a, p.1918): *Let X_n be an $n \times p$ matrix such that (i) its elements are uniformly bounded; and (ii) $\lim_{n \rightarrow \infty} \frac{1}{n} X_n' X_n$ exists and is nonsingular. Then the projectors $P_n = X_n (X_n' X_n)^{-1} X_n'$ and $M_n = I_n - X_n (X_n' X_n)^{-1} X_n'$ are uniformly bounded in both row and column sums.*

Lemma B.3 (Lemma A.9, Lee, 2004b): *Let $\{A_n\}$ be a sequence of $n \times n$ matrices that are uniformly bounded in both row and column sums. For M_n defined in Lemma B.2,*

- (i) $\text{tr}(M_n A_n) = \text{tr}(A_n) + O(1)$,
- (ii) $\text{tr}(A_n' M_n A_n) = \text{tr}(A_n' A_n) + O(1)$,
- (iii) $\text{tr}[(M_n A_n)^2] = \text{tr}(A_n^2) + O(1)$, and (iv) $\text{tr}[(A_n' M_n A_n)^2] = \text{tr}[(A_n' A_n)^2] + O(1)$.

Furthermore, if the elements $a_{n,ij}$ of A_n are $O(h_n^{-1})$ uniformly in all i and j , then,

- (v) $\text{tr}^2(M_n A_n) = \text{tr}^2(A_n) + O(\frac{n}{h_n})$,
- (vi) $\sum_{i=1}^n ((M_n A_n)_{ii})^2 = \sum_{i=1}^n a_{ii}^2 + O(h_n^{-1})$,

where $(M_n A_n)_{ii}$ is the i th diagonal element of $M_n A_n$.

Lemma B.4 (Lemma A.12, Lee, 2004b, extended): *Let $\{A_n\}$ be a sequence of $n \times n$ matrices that are uniformly bounded in either row or column sums. Suppose that the*

elements $a_{n,ij}$ of A_n are $O(h_n^{-1})$ uniformly in all i and j . Let u_n be a random n -vector of iid elements with finite 4th moment, and b_n be a constant n -vector of which the elements are of uniform order $O(h_n^{-1/2})$. Then (i) $E(u_n' A_n u_n) = O(\frac{n}{h_n})$, (ii) $\text{Var}(u_n' A_n u_n) = O(\frac{n}{h_n})$, (iii) $u_n' A_n u_n = O_p(\frac{n}{h_n})$, (iv) $u_n' A_n u_n - E(u_n' A_n u_n) = O_p((\frac{n}{h_n})^{\frac{1}{2}})$, and (v) $u_n' A_n b_n = O_p((\frac{n}{h_n})^{\frac{1}{2}})$.

Note that the results (iv) and (v) in Lemma B.4 extend Lemma A.12 of Lee (2004b), where (iv) follows directly from the generalized Chebyshev inequality and the result (ii): $P(\sqrt{\frac{h_n}{n}} |u_n' A_n u_n - E(u_n' A_n u_n)| \geq M) \leq \frac{1}{M^2} \frac{h_n}{n} \text{Var}(u_n' A_n u_n) = \frac{1}{M^2} O(1)$; and (v) follows from the generalized Chebyshev inequality: $P(\sqrt{\frac{h_n}{n}} |u_n' A_n b_n| \geq M) \leq \frac{1}{M^2} \frac{h_n}{n} \text{Var}(u_n' A_n b_n) = \frac{1}{M^2} \frac{h_n}{n} b_n' A_n' A_n b_n = \frac{1}{M^2} O(1)$.

Lemma B.5 (Kelejian and Prucha, 2001, p.227, extended): *Let $\{A_n\}$ be an $n \times n$ matrix of elements $\{a_{n,ij}\}$, b_n be an $n \times 1$ vector of elements $\{b_{n,i}\}$, and u_n an $n \times 1$ random vector of iid elements, having mean zero, variance σ_0^2 , skewness γ , and excess kurtosis κ . Let $Q_n = u_n' A_n u_n + b_n' u_n$. Then, (i) $E(Q_n) = \sigma_0^2 \text{tr}(A_n)$,*

$$(ii) \text{Var}(Q_n) = \sigma_0^4 \text{tr}(A_n A_n' + A_n^2) + \sigma_0^4 \kappa \sum_{i=1}^n a_{n,ii}^2 + \sigma_0^2 \sum_{i=1}^n b_{n,i}^2 + 2\sigma_0^3 \gamma \sum_{i=1}^n b_{n,i} a_{n,ii}.$$

Furthermore, if $\{a_{n,ij}\}$ are of uniform order $O_p(h_n^{-1})$, $\{b_{n,i}\}$ are of uniform order $O_p(h_n^{-\frac{1}{2}})$, and $\{A_n\}$ are uniformly bounded in either row or column sums, then

$$(iii) E(Q_n) = O(\frac{n}{h_n}), \text{ and (iv) } \text{Var}(Q_n) = O(\frac{n}{h_n}).$$

Note that (iii) and (iv) extend Kelejian and Prucha (2001) to account for $h_n \rightarrow \infty$.

Proof of Lemma 3.1. Denote $\hat{\sigma}_{n0}^2 = \hat{\sigma}_{n0}^2(\lambda_0)$. By the mean value theorem,

$$\hat{\sigma}_{n0}^{-2} = \sigma_0^{-2} - \sigma_0^{-4}(\hat{\sigma}_{n0}^2 - \sigma_0^2) - (\bar{\sigma}_{n0}^{-4} - \sigma_0^{-4})(\hat{\sigma}_{n0}^2 - \sigma_0^2), \quad (\text{B-1})$$

where $\bar{\sigma}_{n0}^2$ lies between $\hat{\sigma}_{n0}^2$ and σ_0^2 . We need to show $\hat{\sigma}_{n0}^2 - \sigma_0^2 = O_p(n^{-\frac{1}{2}})$ and $\bar{\sigma}_{n0}^{-4} - \sigma_0^{-4} = O_p(n^{-\frac{1}{2}})$, so that $\hat{\sigma}_{n0}^{-2} - \sigma_0^{-2} = O_p(n^{-\frac{1}{2}})$. As $\hat{\sigma}_{n0}^2 = \frac{1}{n} Y_n' A_n' M_n A_n Y_n = \frac{1}{n} u_n' M_n u_n$, by Assumptions 2 and 5, and Lemmas B.2, B.1(ii) and B.5(ii), $\text{Var}(\hat{\sigma}_{n0}^2) = O(n^{-1})$. By the generalized Chebyshev inequality: $P(\sqrt{n} |\hat{\sigma}_{n0}^2 - \sigma_0^2| \geq M) \leq \frac{1}{M^2} n \text{Var}(\hat{\sigma}_{n0}^2) = \frac{1}{M^2} O(1)$. It follows that $\hat{\sigma}_{n0}^2 - \sigma_0^2 = O_p(n^{-\frac{1}{2}})$, and hence $\bar{\sigma}_{n0}^2 - \sigma_0^2 = O_p(n^{-\frac{1}{2}})$, $\bar{\sigma}_{n0}^4 = (\sigma_0^2 + O_p(n^{-\frac{1}{2}}))^2 = \sigma_0^4 + O_p(n^{-\frac{1}{2}})$, and finally $\bar{\sigma}_{n0}^{-4} = (\sigma_0^4 + O_p(n^{-\frac{1}{2}}))^{-1} = \sigma_0^{-4} (1 + O_p(n^{-\frac{1}{2}}))^{-1} = \sigma_0^{-4} + O_p(n^{-\frac{1}{2}})$.

For Lemma 3.1(i), $h_n R_{1n} = \frac{h_n}{n} \hat{\sigma}_{n0}^{-2} Y_n' A_n' M_n W_n Y_n = \frac{h_n}{n} \hat{\sigma}_{n0}^{-2} (u_n' M_n G_n u_n + u_n' M_n \eta_n)$. Lemma B.4(iii) implies $\frac{h_n}{n} u_n' M_n G_n u_n = O_p(1)$ and Lemma B.4(v) implies $\frac{h_n}{n} u_n' M_n \eta_n = O_p((\frac{h_n}{n})^{\frac{1}{2}})$. As $\hat{\sigma}_{n0}^{-2} = \sigma_0^{-2} + O_p(n^{-\frac{1}{2}})$, it follows that

$$h_n R_{1n} = \frac{h_n}{n \sigma_0^2} u_n' M_n G_n u_n + O_p((h_n/n)^{\frac{1}{2}}) = O_p(1). \quad (\text{B-2})$$

Now, by (B-1), $E(h_n R_{1n}) = \sigma_0^{-2} E[\frac{h_n}{n} Y_n' A_n' M_n W_n Y_n] - \sigma_0^{-4} E[\frac{h_n}{n} Y_n' A_n' M_n W_n Y_n] (\hat{\sigma}_{n0}^2 - \sigma_0^2) - E[\frac{h_n}{n} Y_n' A_n' M_n W_n Y_n] (\bar{\sigma}_{n0}^{-4} - \sigma_0^{-4}) (\hat{\sigma}_{n0}^2 - \sigma_0^2)$, where the 1st term equals $\frac{h_n}{n} \text{tr}(M_n G_n) =$

$O(1)$ and the 3rd term is $O((\frac{h_n}{n})^{\frac{1}{2}})$ by Assumption 7. For the 2nd term, Cauchy-Schwarz inequality and Lemma B.5 lead to,

$$\begin{aligned} & |\mathbb{E}[(Y_n' A_n' M_n W_n Y_n)(\hat{\sigma}_{n0}^2 - \sigma_0^2)]| \\ & \leq |\mathbb{E}[(Y_n' A_n' M_n W_n Y_n - \sigma_0^2 \text{tr}(M_n G_n))(\hat{\sigma}_{n0}^2 - \sigma_0^2)]| + \sigma_0^2 |\mathbb{E}[\text{tr}(M_n G_n)(\hat{\sigma}_{n0}^2 - \sigma_0^2)]| \\ & \leq \frac{1}{n} \{ \text{Var}(u_n' M_n G_n u_n + u_n' M_n \eta_n) \text{Var}(u_n' M_n u_n - n\sigma_0^2) \}^{\frac{1}{2}} + O(h_n^{-1}) \\ & = \frac{1}{n} \{ O(\frac{n}{h_n}) O(n) \}^{\frac{1}{2}} + O(h_n^{-1}) = O(h_n^{-\frac{1}{2}}), \end{aligned}$$

where we note $\mathbb{E}(\hat{\sigma}_{n0}^2) = \sigma_0^2 + O(n^{-1})$ and $\mathbb{E}[\text{tr}(M_n G_n)] = O(\frac{n}{h_n})$. These lead to,

$$\mathbb{E}(h_n R_{1n}) = \frac{h_n}{n} \text{tr}(M_n G_n) + O(h_n^{1/2}/n) + O((h_n/n)^{\frac{1}{2}}) = O(1). \quad (\text{B-3})$$

Taking difference between (B-2) and (B-3) and using Lemma B.4(iv), we obtain $h_n R_{1n} - \mathbb{E}(h_n R_{1n}) = \frac{h_n}{\sigma_0^2 n} u_n' M_n G_n u_n - \frac{h_n}{n} \text{tr}(M_n G_n) + O_p((\frac{h_n}{n})^{\frac{1}{2}}) - O(\frac{h_n^{1/2}}{n}) = O_p((\frac{h_n}{n})^{\frac{1}{2}})$.

For Lemma 3.1(ii), note that $h_n R_{2n} = \frac{h_n}{n} \hat{\sigma}_{n0}^{-2} Y_n' W_n' M_n W_n Y_n = \frac{h_n}{n} \hat{\sigma}_{n0}^{-2} (u_n' G_n' M_n G_n u_n + 2u_n' G_n' M_n \eta_n + \eta_n' M_n \eta_n)$. Similar arguments as for Lemma 3.1(i) lead to

$$h_n R_{2n} = \frac{h_n}{n\sigma_0^2} (u_n' G_n' M_n G_n u_n + \eta_n' M_n \eta_n) + O_p((h_n/n)^{\frac{1}{2}}) = O_p(1), \quad (\text{B-4})$$

$$\mathbb{E}(h_n R_{2n}) = \frac{h_n}{n} \text{tr}(G_n' M_n G_n) + \frac{h_n}{n\sigma_0^2} \eta_n' M_n \eta_n + O(h_n^{1/2}/n) = O(1). \quad (\text{B-5})$$

The results follow by differencing (B-4) and (B-5) and applying Lemma B.4(iv). *Q.E.D.*

Proof of Theorem 3.1. Clearly, the $\tilde{\psi}(\lambda)$ function given in (13) is differentiable for λ in a neighborhood of λ_0 with its first three derivatives $H_{rn}(\lambda)$, $r = 1, 2$, and 3, given in (14)-(16). These allow us to implement the following third-order Taylor expansion:

$$\begin{aligned} 0 = \tilde{\psi}_n(\hat{\lambda}_n) &= \tilde{\psi}_n + H_{1n}(\hat{\lambda}_n - \lambda_0) + \frac{1}{2} H_{2n}(\hat{\lambda}_n - \lambda_0)^2 + \frac{1}{6} H_{3n}(\hat{\lambda}_n - \lambda_0)^3 \\ &\quad + \frac{1}{6} [H_{3n}(\bar{\lambda}) - H_{3n}](\hat{\lambda}_n - \lambda_0)^3, \end{aligned}$$

where $\bar{\lambda}$ lies between $\hat{\lambda}_n$ and λ_0 . Under Assumptions 1-6, $\hat{\lambda}_n$ is $\sqrt{n/h_n}$ -consistent. Incorporating h_n and following the arguments leading to the result of Theorem 2.1, the result of Theorem 3.1 follows if the following results hold:

- (a) $\tilde{\psi}_n = O_p((\frac{h_n}{n})^{\frac{1}{2}})$ and $\mathbb{E}(\tilde{\psi}_n) = O(\frac{h_n}{n})$;
- (b) $\mathbb{E}(H_{rn}) = O(1)$ and $H_{rn}^\circ = O_p((\frac{h_n}{n})^{\frac{1}{2}})$, $r = 1, 2, 3$;
- (c) $H_{1n}^{-1} = O_p(1)$ and $\mathbb{E}(H_{1n})^{-1} = O(1)$; and
- (d) $H_{3n}(\bar{\lambda}) - H_{3n} = O_p((\frac{h_n}{n})^{\frac{1}{2}})$.

First, Lemma B.1 and Assumptions 3 and 4 give $h_n T_{rn} = O(1)$, $r = 1, 2, 3$.

For (a), by (B-2), $\tilde{\psi}_n = -h_n T_{0n} + h_n R_{1n} = -h_n T_{0n} + \frac{h_n}{\sigma_0^2 n} u_n' M_n G_n u_n + O_p((\frac{h_n}{n})^{\frac{1}{2}})$. As $u_n' M_n G_n u_n - \sigma_0^2 \text{tr}(M_n G_n) = O_p((\frac{h_n}{n})^{\frac{1}{2}})$ by Lemma B.4(iv) and $\text{tr}(M_n G_n) = \text{tr}(G_n) + O(1) = nT_{0n} + O(1)$ by Lemma B.3(i), $\tilde{\psi}_n = O_p((\frac{h_n}{n})^{\frac{1}{2}})$. By (B-3), $E(\tilde{\psi}_n) = -h_n T_{0n} + \frac{h_n}{n} \text{tr}(M_n G_n) + O(\frac{h_n^{1/2}}{n})$. By Lemma B.3(i), $\text{tr}(M_n G_n) = \text{tr}(G_n) + O(1)$. It follows that $E(\tilde{\psi}_n) = O(\frac{h_n}{n})$.

For (b), Lemma 3.1 implies $(h_n R_{1n})^s = (E(h_n R_{1n}))^s + O_p((\frac{h_n}{n})^{\frac{1}{2}})$, for $s = 2, 3, 4$; $(h_n R_{2n})^2 = (E(h_n R_{2n}))^2 + O_p((\frac{h_n}{n})^{\frac{1}{2}})$; and $(h_n R_{1n})^s (h_n R_{2n}) = (E(h_n R_{1n}))^s (E(h_n R_{2n})) + O_p((\frac{h_n}{n})^{\frac{1}{2}})$, for $s = 1, 2$. These and Assumption 8 give $E((h_n R_{1n})^s) = (E(h_n R_{1n}))^s + O((\frac{h_n}{n})^{\frac{1}{2}})$, $s = 2, 3, 4$; $E((h_n R_{2n})^2) = (E(h_n R_{2n}))^2 + O((\frac{h_n}{n})^{\frac{1}{2}})$; and $E((h_n R_{1n})^s (h_n R_{2n})) = (E(h_n R_{1n}))^s E(h_n R_{2n}) + O((\frac{h_n}{n})^{\frac{1}{2}})$, $s = 1, 2$. Finally, as $h_n T_{rn} = O(1)$, $r = 1, 2, 3$, the above results lead immediately to

$$E(H_{rn}) = O(1), r = 1, 2, 3, \quad \text{and} \quad H_{rn}^c = O_p((\frac{h_n}{n})^{\frac{1}{2}}), r = 1, 2, 3.$$

For (c), by (B-3) and (B-5), and the result $E((h_n R_{1n})^2) = (E(h_n R_{1n}))^2 + O((\frac{h_n}{n})^{\frac{1}{2}})$,

$$\begin{aligned} E(H_{1n}) &= -h_n T_{1n} - E(h_n R_{2n}) + \frac{2}{h_n} E((h_n R_{1n})^2) \\ &= -h_n T_{1n} - \frac{h_n}{n} \text{tr}(G_n' M_n G_n) - \frac{h_n}{\sigma_0^2 n} \eta_n' M_n \eta_n - O(\frac{h_n^{1/2}}{n}) \\ &\quad + \frac{2}{h_n} (\frac{h_n}{n} \text{tr}(M_n G_n) + O(\frac{h_n^{1/2}}{n}))^2 \\ &= -h_n T_{1n} - \frac{h_n}{n} \text{tr}(G_n' M_n G_n) - \frac{h_n}{\sigma_0^2 n} \eta_n' M_n \eta_n + \frac{2}{h_n} (\frac{h_n}{n} \text{tr}(M_n G_n))^2 + O(\frac{h_n^{1/2}}{n}) \\ &= -\frac{h_n}{n} \text{tr}(G_n^2) - \frac{h_n}{n} \text{tr}(G_n' G_n) - \frac{h_n}{\sigma_0^2 n} \eta_n' M_n \eta_n + \frac{2}{h_n} (\frac{h_n}{n} \text{tr}(G_n))^2 + O(\frac{h_n}{n}) \\ &= -\frac{h_n}{n} [\text{tr}(G_n - T_{0n} I_n)^2 + \text{tr}(G_n - T_{0n} I_n)' (G_n - T_{0n} I_n) + \frac{1}{\sigma_0^2} \eta_n' M_n \eta_n] + O(\frac{h_n}{n}). \end{aligned}$$

This shows that $E(H_{1n}) < 0$ for n sufficiently large and thus $E(H_{1n})^{-1} = O(1)$. As $H_{1n} = E(H_{1n}) + O_p((\frac{h_n}{n})^{\frac{1}{2}})$, we have $H_{1n}^{-1} = O_p(1)$.

For (d), as $\hat{\sigma}_n^2(\bar{\lambda}) = \frac{1}{n} Y_n' A_n'(\bar{\lambda}) M_n A_n(\bar{\lambda}) Y_n$, we have

$$\begin{aligned} \hat{\sigma}_n^2(\bar{\lambda}) &= \frac{1}{n} Y_n' A_n' M_n A_n Y_n - 2(\bar{\lambda} - \lambda_0) \frac{1}{n} Y_n' A_n' M_n W_n Y_n + (\bar{\lambda} - \lambda_0)^2 \frac{1}{n} Y_n' W_n' M_n W_n Y_n \\ &= \hat{\sigma}_{n0}^2 - 2(\bar{\lambda} - \lambda_0) O_p(h_n^{-1}) + (\bar{\lambda} - \lambda_0)^2 O_p(h_n^{-1}) = \hat{\sigma}_{n0}^2 + O_p((h_n n)^{-\frac{1}{2}}), \end{aligned}$$

leading to $\hat{\sigma}_n^{-2}(\bar{\lambda}) = (\hat{\sigma}_{n0}^2 + O_p((h_n n)^{-\frac{1}{2}}))^{-1} = \hat{\sigma}_{n0}^{-2} (1 + O_p((h_n n)^{-\frac{1}{2}}))^{-1} = \hat{\sigma}_{n0}^{-2} + O_p((h_n n)^{-\frac{1}{2}})$.

Now, as $h_n R_{1n}(\bar{\lambda}) = \hat{\sigma}_n^{-2}(\bar{\lambda}) \frac{h_n}{n} Y_n' A_n'(\bar{\lambda}) M_n W_n Y_n$, we have,

$$\begin{aligned} h_n R_{1n}(\bar{\lambda}) &= \hat{\sigma}_n^{-2}(\bar{\lambda}) \frac{h_n}{n} Y_n' A_n' M_n W_n Y_n - \hat{\sigma}_n^{-2}(\bar{\lambda}) (\bar{\lambda} - \lambda_0) \frac{h_n}{n} Y_n' W_n' M_n W_n Y_n \\ &= (h_n R_{1n} + O_p((h_n n)^{\frac{1}{2}})) - O_p((\frac{h_n}{n})^{\frac{1}{2}}) = h_n R_{1n} + O_p((\frac{h_n}{n})^{\frac{1}{2}}). \end{aligned}$$

Similarly, one shows that $h_n R_{2n}(\bar{\lambda}) = h_n R_{2n} + O_p((\frac{h_n}{n})^{\frac{1}{2}})$. By the mean value theorem, $h_n T_{3n}(\bar{\lambda}) = \frac{h_n}{n} \text{tr}(G_n^4(\bar{\lambda})) = \frac{h_n}{n} \text{tr}(G_n^4) + 4 \frac{h_n}{n} \text{tr}(G_n^3(\check{\lambda})) (\bar{\lambda} - \lambda_0)$, where $\check{\lambda}$ lies between $\bar{\lambda}$ and λ_0 . By Assumption 4 and Lemma B.1, $\frac{h_n}{n} \text{tr}(G_n^3(\check{\lambda})) = O(1)$. It follows that $h_n T_{3n}(\bar{\lambda}) - h_n T_{3n} = O_p((\frac{h_n}{n})^{\frac{1}{2}})$. These lead to $H_{3n}(\bar{\lambda}) - H_{3n} = O_p((\frac{h_n}{n})^{\frac{1}{2}})$. Q.E.D.

Proof of Corollary 3.1. Straightforward.

Proof of Corollary 3.2. Straightforward.

Proof of Corollary 3.3. Straightforward.

Proof of Corollary 3.4. Follow the proof of Corollary 2.4 and replace n by n/h_n .

References

- Amemiya, T., 1985. *Advanced Econometrics*. Cambridge, Massachusetts: Harvard University Press.
- Anselin L., 1988. *Spatial Econometrics: Methods and Models*. The Netherlands: Kluwer Academic Publishers.
- Anselin L., 2001. Rao's score test in spatial econometrics. *Journal of Statistical Planning and Inferences* 97, 113-139.
- Anselin L., Bera, A. K., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: *Handbook of Applied Economic Statistics*, Edited by Aman Ullah and David E. A. Giles. New York: Marcel Dekker.
- Bao, Y., 2010. Finite sample bias of QMLE in spatial autoregressive models. Paper presented in the *IVth World Conference of the Spatial Econometrics Association*, June 9-12, 2010, Chicago (http://www.agecon.purdue.edu/sea_2010/).
- Bao, Y., Ullah, A., 2007a. Finite sample properties of maximum likelihood estimator in spatial models. *Journal of Econometrics* 137, 396-413.
- Bao, Y., Ullah, A., 2007b. The second-order bias and mean squared error of estimators in time-series models. *Journal of Econometrics* 140, 650-669.
- Besley, T., Case, A. C., 1995. Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition. *The American Economic Review* 85, 25-45.
- Bell, K. P., Bockstael, N. E., 2000. Applying the generalized-moment estimation approach to spatial problems involving microlevel data. *Review of Economics and Statistics* 82, 72-82.
- Bertrand, M., Luttmer, E. F. P., Mullainathan, S., 2000. Network effects and welfare cultures. *Quarterly Journal of Economics*, 115, 1019-1055.
- Box, G. E. P., Cox, D. R., 1964. An analysis of transformations (with discussion). *Journal of the Royal Statistical Society B* 26, 211-246.
- Brueckner, J. K., 1998. Testing for Strategic Interaction Among Local Governments: The Case of Growth Controls. *Journal of Urban Economics* 44, 438-467.
- Bun, M. J. G., Carree, M. A. 2005. Bias-corrected estimation in dynamic panel data models. *Journal of Business and Economic Statistics* 23, 200-210.

- Case, A. C., 1991. Spatial Patterns in Household Demand. *Econometrica* 59, 953-965.
- Case, A. C., Rosen, H. S. and Hines, J. R., 1993. Budget Spillovers and Fiscal Policy Interdependence: Evidence from the States. *Journal of Public Economics* 52, 285-307.
- Cliff, A. D., Ord, J. K., 1973. *Spatial Autocorrelation*. London: Pion.
- Cliff, A. D., Ord, J. K., 1981. *Spatial Process, Models and Applications*. London: Pion.
- Efron, B., 1979. Bootstrap method: another look at the jackknife. *The Annals of Statistics* 7, 1-26.
- Hahn J., Kuersteiner, G., 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70, 1639-1657.
- Hahn J., Newey, W., 2004. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72, 1295-1319.
- Hsiao, C., Pesaran, M. H., Tahmiscioglu, A. K., 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short panels. *Journal of Econometrics* 109, 107-150.
- Hsiao, C., (2003). *Analysis of Panel Data, 2nd ed.* Cambridge University Press.
- Huber, P. J., 1964. Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35, 73-101.
- Kelejian H. H., Prucha, I. R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance. *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., Prucha, I. R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104, 219-257.
- Kiviet, J. F., 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of Econometrics* 68, 53-78.
- Laskar, M. R., King, M. L., 1998. Modified likelihood and related methods for handling nuisance parameters in the linear regression model. *Monash Econometrics and Business Statistics Working Papers, no. 5/98*.
- Laskar, M. R., King, M. L., 1998. Modified likelihood and related methods for handling nuisance parameters in the linear regression model. *Monash Econometrics and Business Statistics Working Papers, no. 5/98*.
- Lee, L. F., 2002. Consistency and efficiency of least squares estimation for mixed regressive spatial autoregressive models. *Econometric Theory* 18, 252-277.
- Lee, L. F., 2003. Best spatial two-stage least squares estimators for a spatial autoregressive models with autoregressive disturbances. *Econometric Reviews* 22, 307-335.

- Lee, L. F., 2004a. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899-1925.
- Lee, L. F., 2004b. A supplement to 'Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models'. Working paper, Department of Economics, Ohio State University.
- Lee, L. F., 2007a. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137, 489-514.
- Lee, L. F., 2007b. Identification and estimation of econometric models with group interaction, contextual factors and fixed effects. *Journal of Econometrics* 140, 333-374.
- Lee, L. F., Yu, J. 2010. Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154, 165-185.
- Lahiri, S. N., 2003. *Resampling Methods for Dependent Data*. New York: Springer.
- Mynbaev, K. T., Ullah, A., 2008. Asymptotic distribution of the OLS estimator for a purely autoregressive spatial model. *Journal of Multivariate Analysis* 99, 245-277.
- Ord, J., 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, 369-395.
- Robinson, P. M., 2010. Efficient estimation of the semiparametric spatial autoregressive model. *Journal of Econometrics* 157, 6-17.
- 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.
- Su, L. J., Jin, S. N., 2010. Profile quasi-maximum likelihood estimation of spatial autoregressive models. *Journal of Econometrics* 157, 18-33.
- Su, L. J., 2012. Semiparametric GMM estimation of spatial autoregressive models. *Journal of Econometrics* 167, 543-560.
- Topa, G., 2001. Social Interaction, local spillovers and unemployment. *Review of Economic Studies*. 68, 261-295.
- White, H., 1994. *Estimation, Inference and Specification Analysis*. New York: Cambridge University Press.
- Ullah, A., 2004. *Finite Sample Econometrics*. New York: Oxford University Press.
- Yang, Z. L., 2010a. Bias-corrected estimation for spatial autocorrelation. Paper presented in the *IVth World Conference of the Spatial Econometrics Association*, June 9-12, 2010, Chicago (http://www.agecon.purdue.edu/sea_2010/).
- Yang, Z. L., 2010b. A robust LM test for spatial error components. *Regional Science and Urban Economics* 40, 299-310.

Table 1. Empirical Mean[rmse](sd) of Estimators of λ , and Averaged Bootstrap SEs

λ	n	$\hat{\lambda}_n$	$\hat{\lambda}_n^{bc2}$	$\hat{\lambda}_n^{bc3}$	\overline{se}_1	\overline{se}_2	\overline{se}_3	\overline{se}_3^c
(a) Queen Contiguity, Normal Errors, MRSAR-A								
.50	50	.411 [.195](.174)	.492 .175	.497 .175	.159	.171	.179	.179
	100	.459 [.123](.116)	.498 .117	.500 .117	.113	.116	.120	.120
	200	.480 [.078](.076)	.499 .075	.499 .075	.073	.074	.076	.075
	500	.493 [.049](.048)	.501 .048	.501 .048	.048	.048	.049	.049
.25	50	.163 [.222](.204)	.242 .209	.246 .210	.190	.203	.202	.207
	100	.212 [.146](.140)	.248 .142	.250 .143	.136	.139	.139	.141
	200	.231 [.094](.092)	.250 .093	.250 .093	.090	.092	.092	.092
	500	.242 .060	.250 .060	.250 .060	.060	.060	.060	.060
.00	50	-.078 [.229](.216)	-.006 .224	-.003 .226	.210	.222	.215	.223
	100	-.034 [.157](.153)	-.002 .156	-.001 .157	.151	.154	.151	.155
	200	-.018 [.106](.104)	-.000 .105	.000 .105	.103	.104	.104	.104
	500	-.008 [.068](.067)	-.000 .068	-.000 .068	.068	.068	.068	.068
-.25	50	-.317 [.233](.223)	-.255 .236	-.254 .237	.221	.232	.220	.231
	100	-.279 [.164](.161)	-.253 .166	-.253 .166	.158	.161	.156	.161
	200	-.266 [.112](.111)	-.252 .112	-.251 .112	.110	.112	.110	.112
	500	-.256 [.073](.072)	-.250 .073	-.250 .073	.072	.073	.072	.072
-.50	50	-.552 [.228](.222)	-.504 .236	-.504 .237	.223	.232	.217	.230
	100	-.519 [.162](.161)	-.501 .166	-.501 .166	.159	.161	.155	.160
	200	-.514 .113	-.502 .114	-.502 .114	.113	.114	.112	.114
	500	-.505 .073	-.500 .073	-.500 .073	.074	.074	.073	.074
(b) Queen Contiguity, Normal Mixture Errors, MRSAR-A								
.50	50	.420 [.182](.164)	.494 .165	.498 .165	.149	.160	.167	.167
	100	.462 [.120](.114)	.499 .114	.500 .114	.108	.111	.115	.115
	200	.482 [.076](.074)	.500 .074	.500 .074	.071	.072	.074	.074
	500	.492 [.049](.048)	.500 .048	.500 .048	.048	.048	.048	.048
.25	50	.169 [.207](.190)	.241 .195	.244 .195	.179	.190	.191	.195
	100	.213 [.140](.135)	.248 .136	.249 .137	.130	.133	.134	.136
	200	.230 [.092](.090)	.249 .090	.249 .090	.088	.089	.090	.090
	500	.242 .060	.250 .060	.250 .060	.059	.059	.060	.060
.00	50	-.070 [.217](.206)	-.004 .213	-.002 .214	.197	.207	.204	.211
	100	-.032 [.150](.147)	-.002 .150	-.001 .150	.145	.148	.146	.149
	200	-.018 [.104](.103)	-.001 .103	-.001 .103	.100	.101	.101	.102
	500	-.008 [.068](.067)	-.001 .067	-.001 .067	.067	.067	.067	.067
-.25	50	-.314 [.223](.213)	-.258 .224	-.257 .225	.208	.216	.209	.219
	100	-.275 [.155](.153)	-.251 .157	-.250 .157	.152	.154	.151	.155
	200	-.263 [.111](.110)	-.249 .112	-.249 .112	.108	.109	.108	.109
	500	-.257 .072	-.251 .072	-.251 .072	.071	.072	.071	.072
-.50	50	-.550 [.218](.212)	-.506 .224	-.505 .225	.210	.216	.207	.218
	100	-.520 [.155](.153)	-.503 .158	-.503 .158	.152	.154	.150	.155
	200	-.513 [.112](.111)	-.502 .113	-.502 .113	.111	.111	.110	.112
	500	-.505 [.074](.073)	-.500 .074	-.500 .074	.073	.073	.073	.073

Note: $\overline{se}_1 = \text{mean}(\widehat{V}_1(\hat{\lambda}_n)^{\frac{1}{2}})$, $\overline{se}_2 = \text{mean}(\widehat{V}_2(\hat{\lambda}_n)^{\frac{1}{2}})$, $\overline{se}_3 = \text{mean}(\widehat{V}_3(\hat{\lambda}_n)^{\frac{1}{2}})$ and $\overline{se}_3^c = \text{mean}(\widehat{V}_3(\hat{\lambda}_n^{bc3})^{\frac{1}{2}})$.

Table 1 (cont'd). Empirical Mean[rmse](sd) of Estimators of λ , and Averaged Bootstrap SEs

λ	n	$\hat{\lambda}_n$	$\hat{\lambda}_n^{bc2}$	$\hat{\lambda}_n^{bc3}$	\overline{se}_1	\overline{se}_2	\overline{se}_3	\overline{se}_3^c
(c) Queen Contiguity, Lognormal Errors, MRSAR-A								
.50	50	.426 [.163](.146)	.491 .146	.493 .146	.138	.144	.154	.154
	100	.465 [.110](.105)	.498 .105	.498 .105	.102	.103	.106	.106
	200	.482 [.072](.069)	.499 .069	.499 .069	.067	.067	.069	.069
	500	.491 [.047](.046)	.499 .046	.499 .046	.046	.046	.046	.046
.25	50	.179 [.185](.171)	.241 .174	.244 .174	.163	.170	.176	.179
	100	.216 [.128](.124)	.247 [.126](.125)	.248 [.126](.125)	.123	.123	.124	.126
	200	.232 [.087](.085)	.249 .085	.249 .085	.084	.083	.085	.085
	500	.242 [.058](.057)	.249 .057	.249 .057	.057	.057	.057	.057
.00	50	-.067 [.198](.186)	-.011 [.192](.191)	-.008 .192	.180	.186	.190	.196
	100	-.029 [.139](.136)	-.003 .138	-.002 .138	.135	.136	.136	.138
	200	-.017 [.099](.097)	-.002 .098	-.001 .098	.095	.095	.096	.096
	500	-.007 [.065](.064)	-.000 .065	.000 .065	.065	.064	.065	.065
-.25	50	-.307 [.199](.191)	-.258 .198	-.256 .199	.189	.194	.197	.205
	100	-.272 [.142](.140)	-.252 .144	-.251 .144	.141	.142	.141	.145
	200	-.264 [.105](.104)	-.251 .105	-.250 .105	.102	.102	.102	.103
	500	-.256 .070	-.250 .070	-.250 .070	.069	.069	.069	.069
-.50	50	-.548 [.196](.190)	-.509 [.200](.199)	-.507 .200	.191	.195	.196	.206
	100	-.514 [.145](.144)	-.500 .148	-.499 .148	.141	.141	.141	.145
	200	-.511 .106	-.501 .107	-.501 .107	.105	.105	.105	.106
	500	-.505 .070	-.501 .070	-.500 .070	.071	.070	.070	.071
(d) Group Interaction with $k = n^{0.5}$, Normal Errors, MRSAR-B								
.50	50	.426 [.145](.124)	.495 .122	.499 .122	.111	.123	.128	.121
	100	.449 [.112](.099)	.498 .097	.500 .097	.092	.097	.102	.098
	200	.474 [.073](.068)	.499 .067	.500 .067	.065	.067	.069	.068
	500	.491 [.042](.041)	.500 .041	.500 .041	.040	.041	.041	.041
.25	50	.142 [.209](.179)	.239 .178	.244 [.179](.178)	.159	.175	.180	.174
	100	.177 [.160](.143)	.248 .141	.250 .141	.133	.140	.146	.142
	200	.212 [.108](.102)	.249 .101	.249 .101	.096	.098	.101	.100
	500	.236 [.062](.060)	.250 .061	.250 .061	.060	.060	.060	.060
.00	50	-.134 [.261](.224)	-.013 .226	-.007 .227	.202	.220	.223	.219
	100	-.094 [.206](.183)	-.005 .182	-.003 .182	.172	.180	.186	.183
	200	-.047 [.137](.128)	.001 .128	.001 .128	.125	.128	.131	.130
	500	-.019 [.081](.079)	-.000 .079	-.000 .079	.078	.079	.079	.080
-.25	50	-.404 [.303](.260)	-.265 .266	-.259 .267	.240	.258	.259	.259
	100	-.358 [.243](.218)	-.253 .219	-.251 .219	.207	.215	.221	.219
	200	-.308 [.170](.160)	-.251 .160	-.250 .160	.153	.156	.159	.159
	500	-.274 [.101](.098)	-.252 .098	-.252 .098	.097	.098	.098	.098
-.50	50	-.670 [.337](.291)	-.516 [.302](.301)	-.511 [.303](.302)	.273	.290	.288	.293
	100	-.620 [.281](.254)	-.501 .257	-.499 .257	.238	.246	.252	.252
	200	-.568 [.198](.186)	-.502 .187	-.501 .187	.179	.183	.186	.186
	500	-.527 [.118](.115)	-.500 .116	-.500 .116	.115	.116	.116	.117

Note: $\overline{se}_1 = \text{mean}(\widehat{V}_1(\hat{\lambda}_n)^{\frac{1}{2}})$, $\overline{se}_2 = \text{mean}(\widehat{V}_2(\hat{\lambda}_n)^{\frac{1}{2}})$, $\overline{se}_3 = \text{mean}(\widehat{V}_3(\hat{\lambda}_n)^{\frac{1}{2}})$ and $\overline{se}_3^c = \text{mean}(\widehat{V}_3(\hat{\lambda}_n^{bc3})^{\frac{1}{2}})$.

Table 1 (cont'd). Empirical Mean[rmse](sd) of Estimators of λ , and Averaged Bootstrap SEs

λ	n	$\hat{\lambda}_n$	$\hat{\lambda}_n^{bc2}$	$\hat{\lambda}_n^{bc3}$	\overline{se}_1	\overline{se}_2	\overline{se}_3	\overline{se}_3^c
(e) Group Interaction with $k = n^{0.5}$, Normal Mixture Errors, MRSAR-B								
.50	50	.427 [.144](.124)	.489 [.121](.120)	.493 .120	.104	.114	.120	.115
	100	.449 [.111](.098)	.495 .096	.497 .096	.088	.093	.098	.095
	200	.474 [.073](.068)	.498 .068	.499 .068	.064	.065	.067	.067
	500	.490 [.042](.041)	.500 .041	.500 .041	.040	.040	.041	.041
.25	50	.147 [.203](.175)	.234 [.172](.171)	.239 [.172](.171)	.149	.163	.171	.166
	100	.179 [.157](.140)	.245 [.138](.137)	.248 .138	.128	.133	.140	.137
	200	.214 [.104](.097)	.249 .097	.249 .097	.093	.095	.099	.098
	500	.236 [.062](.060)	.250 .060	.250 .060	.059	.059	.060	.060
.00	50	-.126 [.250](.215)	-.017 [.214](.213)	-.011 .214	.191	.207	.215	.212
	100	-.089 [.203](.182)	-.006 .180	-.004 .181	.164	.171	.179	.176
	200	-.047 [.137](.129)	-.002 [.129](.128)	-.001 .129	.122	.124	.128	.128
	500	-.020 [.082](.080)	-.002 .080	-.002 .080	.078	.078	.079	.079
-.25	50	-.393 [.294](.256)	-.267 .259	-.261 [.260](.259)	.227	.244	.251	.252
	100	-.358 [.241](.216)	-.259 .216	-.257 .216	.199	.205	.213	.212
	200	-.306 [.165](.156)	-.251 .155	-.250 .155	.149	.151	.156	.156
	500	-.274 [.101](.098)	-.251 .099	-.251 .099	.096	.096	.097	.098
-.50	50	-.658 [.330](.289)	-.519 [.298](.297)	-.513 .298	.260	.277	.282	.287
	100	-.618 [.270](.243)	-.507 .245	-.505 .245	.230	.237	.245	.246
	200	-.566 [.197](.185)	-.503 .186	-.502 .186	.175	.178	.183	.183
	500	-.528 [.120](.117)	-.501 .117	-.501 .117	.113	.114	.115	.116
(f) Group Interaction with $k = n^{0.5}$, Lognormal Errors, MRSAR-B								
.50	50	.437 [.127](.110)	.488 [.107](.106)	.492 .106	.094	.102	.117	.113
	100	.452 [.104](.092)	.492 .090	.494 .090	.083	.085	.092	.090
	200	.475 [.071](.067)	.497 .066	.497 .066	.060	.060	.065	.064
	500	.490 [.041](.039)	.499 .040	.499 .040	.038	.037	.039	.039
.25	50	.165 [.180](.158)	.236 [.156](.155)	.241 .155	.135	.146	.166	.162
	100	.184 [.146](.131)	.242 .128	.244 .128	.118	.120	.130	.128
	200	.215 [.101](.094)	.247 [.094](.093)	.247 .093	.088	.088	.094	.094
	500	.235 [.060](.059)	.248 .059	.248 .059	.056	.056	.058	.058
.00	50	-.109 [.231](.203)	-.020 [.202](.201)	-.014 .201	.174	.187	.210	.207
	100	-.081 [.179](.160)	-.008 .159	-.006 [.159](.158)	.151	.152	.164	.163
	200	-.046 [.131](.123)	-.005 .122	-.004 .122	.114	.114	.122	.122
	500	-.018 [.078](.076)	-.001 .077	-.001 [.077](.076)	.074	.073	.077	.077
-.25	50	-.377 [.274](.243)	-.272 [.245](.244)	-.265 .244	.210	.225	.253	.252
	100	-.346 [.216](.193)	-.261 .194	-.259 [.194](.193)	.181	.181	.194	.195
	200	-.307 [.159](.148)	-.258 .148	-.257 .148	.140	.139	.148	.148
	500	-.273 [.099](.096)	-.252 .097	-.252 .097	.092	.091	.095	.095
-.50	50	-.639 [.312](.279)	-.524 [.286](.285)	-.516 .285	.241	.258	.285	.288
	100	-.610 [.245](.219)	-.516 [.222](.221)	-.514 [.222](.221)	.209	.208	.220	.224
	200	-.565 [.186](.174)	-.508 .174	-.507 .174	.165	.164	.173	.174
	500	-.527 [.117](.114)	-.503 .114	-.502 .114	.110	.109	.113	.114

Note: $\overline{se}_1 = \text{mean}(\widehat{V}_1(\hat{\lambda}_n)^{\frac{1}{2}})$, $\overline{se}_2 = \text{mean}(\widehat{V}_2(\hat{\lambda}_n)^{\frac{1}{2}})$, $\overline{se}_3 = \text{mean}(\widehat{V}_3(\hat{\lambda}_n)^{\frac{1}{2}})$ and $\overline{se}_3^c = \text{mean}(\widehat{V}_3(\hat{\lambda}_n^{bc3})^{\frac{1}{2}})$.

Table 2a. Null Behavior of t -Ratios for Testing $H_0 : \lambda = 0$: Group Interaction with $k = n^{0.35}$

n	Test	dgp	Mean	SD	Empirical Tail Probabilities: L=left, R=right					
					L-1%	L-2.5%	L-5%	R-5%	R-2.5%	R-1%
Nominal Values			0.0000	1.0000	0.0100	0.0250	0.0500	0.0500	0.0250	0.0100
50	t_{11}	1	-0.5904	1.0572	0.0470	0.0965	0.1553	0.0210	0.0113	0.0051
		2	-0.6080	1.0801	0.0554	0.0974	0.1585	0.0209	0.0109	0.0042
		3	-0.5607	1.1100	0.0622	0.1072	0.1590	0.0159	0.0069	0.0030
	t_{21}	1	0.0088	1.1571	0.0193	0.0404	0.0729	0.0796	0.0472	0.0265
		2	-0.0526	1.1665	0.0246	0.0505	0.0794	0.0712	0.0436	0.0244
		3	-0.0655	1.1712	0.0339	0.0601	0.0926	0.0610	0.0340	0.0151
	t_{22}	1	-0.0085	1.0573	0.0143	0.0315	0.0591	0.0574	0.0323	0.0156
		2	-0.0644	1.0830	0.0201	0.0417	0.0705	0.0563	0.0323	0.0152
		3	-0.0739	1.0780	0.0280	0.0487	0.0809	0.0425	0.0208	0.0077
	t_{33}	1	0.0106	1.0583	0.0141	0.0309	0.0564	0.0588	0.0343	0.0157
		2	-0.0254	1.0350	0.0153	0.0322	0.0580	0.0512	0.0295	0.0141
		3	-0.0258	0.9535	0.0185	0.0332	0.0533	0.0327	0.0145	0.0055
100	t_{11}	1	-0.5341	1.0220	0.0383	0.0771	0.1376	0.0185	0.0084	0.0038
		2	-0.5089	1.0464	0.0387	0.0828	0.1385	0.0202	0.0101	0.0035
		3	-0.5296	1.0904	0.0518	0.0959	0.1508	0.0241	0.0111	0.0047
	t_{21}	1	0.0300	1.0906	0.0138	0.0315	0.0590	0.0687	0.0403	0.0203
		2	0.0339	1.1103	0.0169	0.0351	0.0657	0.0745	0.0423	0.0200
		3	-0.0398	1.1400	0.0205	0.0468	0.0812	0.0722	0.0407	0.0199
	t_{22}	1	0.0189	1.0320	0.0111	0.0274	0.0529	0.0574	0.0304	0.0137
		2	0.0216	1.0671	0.0153	0.0326	0.0612	0.0647	0.0339	0.0151
		3	-0.0479	1.1219	0.0209	0.0470	0.0791	0.0664	0.0365	0.0161
	t_{33}	1	0.0293	1.0090	0.0095	0.0245	0.0485	0.0534	0.0272	0.0119
		2	0.0401	1.0104	0.0105	0.0250	0.0466	0.0552	0.0270	0.0122
		3	-0.0091	0.9954	0.0099	0.0252	0.0505	0.0502	0.0257	0.0109
200	t_{11}	1	-0.3593	1.0045	0.0254	0.0539	0.0978	0.0236	0.0125	0.0059
		2	-0.3578	1.0367	0.0283	0.0581	0.1062	0.0295	0.0148	0.0063
		3	-0.3633	1.0686	0.0326	0.0628	0.1104	0.0328	0.0163	0.0053
	t_{21}	1	0.0483	1.0508	0.0114	0.0277	0.0523	0.0628	0.0346	0.0173
		2	0.0393	1.0823	0.0137	0.0302	0.0555	0.0690	0.0414	0.0199
		3	0.0054	1.1089	0.0159	0.0363	0.0626	0.0701	0.0433	0.0206
	t_{22}	1	0.0445	1.0266	0.0105	0.0255	0.0498	0.0578	0.0313	0.0146
		2	0.0346	1.0680	0.0130	0.0295	0.0550	0.0648	0.0378	0.0176
		3	0.0003	1.1114	0.0176	0.0372	0.0648	0.0703	0.0419	0.0199
	t_{33}	1	0.0408	0.9954	0.0095	0.0236	0.0462	0.0500	0.0262	0.0120
		2	0.0352	1.0140	0.0105	0.0250	0.0477	0.0575	0.0308	0.0121
		3	0.0161	1.0048	0.0104	0.0232	0.0460	0.0548	0.0292	0.0111

Note: (1) X_1 and X_2 are generated from MRSAR-A schme, $\sigma = 1$, and $\beta = (5, 1, 1)'$;
(2) dgp: 1=normal, 2=normal mixture($\tau = 4, p = .1$), 3=lognormal;
(3) t_{ij} : t -ratio with i th-order corrected estimator and j th-order corrected variance of it.

Table 2b. Null Behavior of t -Ratios for Testing $H_0 : \lambda = 0$: Group Interaction with $k = n^{0.5}$

n	stat	dgp	Mean	SD	Empirical Tail Probabilities: L=left, R=right					
					L-1%	L-2.5%	L-5%	R-5%	R-2.5%	R-1%
Nominal Values			0.0000	1.0000	0.0100	0.0250	0.0500	0.0500	0.0250	0.0100
50	t_{11}	1	-0.5396	1.0523	0.0430	0.0833	0.1395	0.0220	0.0118	0.0055
		2	-0.5609	1.0624	0.0468	0.0875	0.1460	0.0199	0.0096	0.0054
		3	-0.5135	1.0627	0.0427	0.0841	0.1398	0.0227	0.0111	0.0049
	t_{21}	1	0.0607	1.1225	0.0167	0.0342	0.0627	0.0767	0.0467	0.0245
		2	0.0036	1.1299	0.0218	0.0410	0.0675	0.0706	0.0413	0.0209
		3	-0.0106	1.1193	0.0201	0.0393	0.0696	0.0660	0.0386	0.0186
	t_{22}	1	0.0412	1.0279	0.0120	0.0269	0.0517	0.0569	0.0317	0.0139
		2	-0.0115	1.0438	0.0171	0.0327	0.0572	0.0530	0.0277	0.0117
		3	-0.0205	1.0448	0.0162	0.0333	0.0589	0.0530	0.0281	0.0121
	t_{33}	1	0.0700	1.0439	0.0126	0.0261	0.0497	0.0639	0.0363	0.0172
		2	0.0227	1.0307	0.0137	0.0284	0.0504	0.0547	0.0293	0.0135
		3	0.0183	0.9889	0.0121	0.0267	0.0456	0.0498	0.0256	0.0126
100	t_{11}	1	-0.3930	1.0200	0.0292	0.0619	0.1088	0.0219	0.0126	0.0049
		2	-0.3850	1.0367	0.0288	0.0632	0.1112	0.0266	0.0131	0.0059
		3	-0.3872	1.0523	0.0332	0.0677	0.1129	0.0271	0.0134	0.0055
	t_{21}	1	0.0542	1.0577	0.0121	0.0289	0.0543	0.0625	0.0351	0.0171
		2	0.0470	1.0737	0.0128	0.0292	0.0576	0.0710	0.0405	0.0185
		3	0.0103	1.0824	0.0167	0.0355	0.0643	0.0638	0.0364	0.0162
	t_{22}	1	0.0472	1.0117	0.0100	0.0245	0.0496	0.0533	0.0274	0.0129
		2	0.0391	1.0373	0.0108	0.0268	0.0524	0.0629	0.0335	0.0148
		3	0.0053	1.0654	0.0161	0.0347	0.0620	0.0597	0.0332	0.0146
	t_{33}	1	0.0570	1.0058	0.0097	0.0235	0.0483	0.0530	0.0273	0.0126
		2	0.0520	1.0107	0.0090	0.0231	0.0456	0.0601	0.0312	0.0130
		3	0.0240	0.9940	0.0105	0.0243	0.0474	0.0520	0.0260	0.0108
200	t_{11}	1	-0.3265	1.0085	0.0213	0.0499	0.0939	0.0265	0.0124	0.0050
		2	-0.3182	1.0250	0.0239	0.0524	0.0979	0.0288	0.0133	0.0055
		3	-0.3165	1.0360	0.0237	0.0552	0.0972	0.0322	0.0173	0.0072
	t_{21}	1	0.0418	1.0376	0.0094	0.0251	0.0492	0.0640	0.0343	0.0141
		2	0.0433	1.0531	0.0121	0.0280	0.0521	0.0663	0.0353	0.0149
		3	0.0217	1.0610	0.0125	0.0287	0.0575	0.0649	0.0377	0.0182
	t_{22}	1	0.0377	1.0101	0.0087	0.0220	0.0463	0.0575	0.0296	0.0119
		2	0.0386	1.0330	0.0116	0.0268	0.0503	0.0602	0.0316	0.0134
		3	0.0179	1.0592	0.0129	0.0290	0.0575	0.0634	0.0358	0.0175
	t_{33}	1	0.0396	0.9997	0.0085	0.0215	0.0446	0.0557	0.0278	0.0114
		2	0.0435	1.0076	0.0102	0.0232	0.0467	0.0557	0.0287	0.0113
		3	0.0310	0.9997	0.0081	0.0193	0.0466	0.0551	0.0301	0.0132

Note: (1) X_1 and X_2 are generated from MRSAR-A schme, $\sigma = 1$, and $\beta = (5, 1, 1)'$;
(2) dgp: 1=normal, 2=normal mixture($\tau = 4, p = .1$), 3=lognormal;
(3) t_{ij} : t -ratio with i th-order corrected estimator and j th-order corrected variance of it.

Table 3. Empirical Means [sds] of the MLE and Second-Order Bias-Corrected MLEs of λ : Comparison with Analytical Approach of Bao and Ullah (2007a), Pure SAR Model

J	λ	$\hat{\lambda}_n$	$\hat{\lambda}_n^{bc2}$	$\hat{\lambda}_n^{BU}$	$\hat{\lambda}_n$	$\hat{\lambda}_n^{bc2}$	$\hat{\lambda}_n^{BU}$
		$n = 30, u_n \sim N(0, I_n)$			$n = 100, u_n \sim N(0, I_n)$		
2	0.9	0.881 [.057]	0.899 [.054]	0.899 [.054]	0.894 [.027]	0.900 [.026]	0.900 [.026]
	0.4	0.380 [.158]	0.397 [.164]	0.398 [.164]	0.395 [.086]	0.401 [.087]	0.401 [.087]
	0.2	0.189 [.176]	0.198 [.185]	0.198 [.184]	0.199 [.096]	0.202 [.098]	0.202 [.098]
	0.0	0.001 [.181]	0.000 [.190]	0.001 [.190]	-0.001 [.099]	-0.001 [.100]	-0.001 [.100]
	-0.2	-0.194 [.174]	-0.204 [.182]	-0.204 [.182]	-0.196 [.095]	-0.199 [.097]	-0.199 [.097]
	-0.4	-0.384 [.160]	-0.402 [.166]	-0.401 [.166]	-0.394 [.086]	-0.400 [.087]	-0.400 [.087]
	-0.9	-0.882 [.057]	-0.900 [.054]	-0.900 [.054]	-0.895 [.027]	-0.900 [.026]	-0.900 [.026]
6	0.9	0.853 [.109]	0.898 [.097]	0.897 [.097]	0.887 [.041]	0.899 [.038]	0.899 [.038]
	0.4	0.322 [.267]	0.397 [.271]	0.397 [.270]	0.376 [.135]	0.400 [.135]	0.400 [.135]
	0.2	0.126 [.300]	0.198 [.312]	0.198 [.312]	0.178 [.159]	0.201 [.161]	0.201 [.160]
	0.0	-0.072 [.324]	-0.008 [.344]	-0.007 [.343]	-0.025 [.174]	-0.004 [.177]	-0.004 [.177]
	-0.2	-0.259 [.341]	-0.207 [.366]	-0.206 [.366]	-0.216 [.188]	-0.198 [.192]	-0.198 [.192]
	-0.4	-0.442 [.349]	-0.403 [.378]	-0.402 [.378]	-0.414 [.197]	-0.401 [.202]	-0.401 [.202]
	-0.9	-0.898 [.346]	-0.900 [.377]	-0.899 [.377]	-0.900 [.193]	-0.900 [.198]	-0.900 [.198]
10	0.9	0.826 [.165]	0.899 [.139]	0.897 [.139]	0.880 [.055]	0.899 [.051]	0.899 [.051]
	0.4	0.255 [.369]	0.388 [.374]	0.387 [.373]	0.353 [.178]	0.395 [.176]	0.395 [.176]
	0.2	0.054 [.411]	0.186 [.431]	0.185 [.430]	0.154 [.208]	0.197 [.208]	0.197 [.208]
	0.0	-0.143 [.441]	-0.018 [.474]	-0.018 [.473]	-0.046 [.235]	-0.004 [.238]	-0.004 [.238]
	-0.2	-0.327 [.457]	-0.215 [.501]	-0.214 [.500]	-0.239 [.249]	-0.199 [.255]	-0.200 [.255]
	-0.4	-0.497 [.475]	-0.401 [.530]	-0.401 [.529]	-0.437 [.266]	-0.402 [.274]	-0.402 [.274]
	-0.9	-0.942 [.484]	-0.901 [.556]	-0.900 [.555]	-0.917 [.283]	-0.901 [.296]	-0.901 [.296]
6	0.4	$n = 100, u_n \sim LN(0, I_n)$			$n = 200, u_n \sim LN(0, I_n)$		
		0.386 [.121]	0.402 [.122]	0.409 [.120]	0.393 [.085]	0.401 [.085]	0.405 [.084]
		0.187 [.143]	0.202 [.145]	0.211 [.144]	0.192 [.100]	0.201 [.101]	0.204 [.101]
		-0.008 [.159]	0.004 [.162]	0.013 [.161]	-0.006 [.113]	0.001 [.115]	0.005 [.114]
		-0.207 [.173]	-0.199 [.177]	-0.189 [.176]	-0.206 [.122]	-0.201 [.123]	-0.197 [.123]
-0.4	-0.405 [.182]	-0.401 [.187]	-0.391 [.187]	-0.403 [.130]	-0.400 [.132]	-0.396 [.132]	
10	0.4	0.370 [.159]	0.404 [.159]	0.412 [.156]	0.386 [.106]	0.403 [.106]	0.407 [.105]
		0.172 [.189]	0.206 [.190]	0.216 [.188]	0.183 [.132]	0.201 [.132]	0.206 [.131]
		-0.026 [.216]	0.005 [.220]	0.017 [.218]	-0.016 [.147]	0.001 [.149]	0.006 [.148]
		-0.224 [.233]	-0.197 [.241]	-0.184 [.238]	-0.214 [.163]	-0.200 [.165]	-0.194 [.165]
		-0.420 [.249]	-0.400 [.259]	-0.385 [.257]	-0.411 [.174]	-0.399 [.177]	-0.393 [.177]
14	0.4	0.357 [.193]	0.412 [.192]	0.417 [.187]	0.376 [.131]	0.402 [.130]	0.406 [.129]
		0.155 [.235]	0.210 [.237]	0.219 [.233]	0.177 [.158]	0.203 [.158]	0.209 [.157]
		-0.039 [.258]	0.013 [.264]	0.025 [.260]	-0.023 [.179]	0.003 [.180]	0.010 [.179]
		-0.234 [.284]	-0.187 [.294]	-0.172 [.290]	-0.223 [.199]	-0.199 [.202]	-0.192 [.201]
		-0.433 [.305]	-0.393 [.318]	-0.377 [.315]	-0.422 [.213]	-0.400 [.217]	-0.392 [.217]

Note: $\hat{\lambda}_n^{BU}$: 2nd-Order Analytically Bias-Corrected MLE of Bao and Ullah (2007a).