# 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*

email: zlyang@smu.edu.sg

June 26, 2014

**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. The validity of the bootstrap procedure is formally established. The proposed method is then extended to the case of more than one nonlinear estimator.

**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 the spatial autoregressive model, spatial

*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.

panel model with fixed effects, dynamic regression model, dynamic panel model with fixed effects, Box-Cox regression, Weibull duration model, etc.. The bias 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, 2013). Another important issue, the high-order correction on the variance of a bias-corrected estimator, has not been formally addressed.

Stochastic expansion (Rilstone 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, 2009; Kundhi and Rilstone, 2008; Bao, 2013). However, in high-order bias and variance corrections: (*i*) it involves high dimension matrix manipulations and (*ii*) it requires closed form 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:

$$Y_n = \lambda W_n Y_n + X_n\beta + \varepsilon_n, \ \ \varepsilon_n = \sigma u_n, \tag{1}$$

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

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

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

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 $\theta$ if the errors are exactly normal, otherwise the quasi-MLE (QMLE). Given $\lambda$, the constrained QMLEs of $\beta$ and $\sigma^2$ are

$$\hat\beta_n(\lambda) = (X_n'X_n)^{-1}X_n'A_n(\lambda)Y_n \ \text{ and } \ \hat\sigma_n^2(\lambda) = n^{-1}Y_n'A_n'(\lambda)M_nA_n(\lambda)Y_n, \tag{3}$$

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

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

Maximizing $\ell_n^c(\lambda)$ gives the unconstrained QMLE $\hat\lambda_n$ of $\lambda$. The unconstrained QMLEs of $\beta$ and $\sigma^2$ 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 $\hat\theta_n = (\hat\lambda_n, \hat\beta_n', \hat\sigma_n^2)'$.

[1] For 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, 2013, for a second-order bias formula). We note that if $\lambda$ 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 $\lambda$. Lee (2007a) made a similar remark based on his Monte Carlo results. Further, given $\lambda$ 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 $\lambda$. 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): $h(Y_n, \lambda) = X_n\beta + \sigma u_n$, 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 $\lambda$, applied to $Y_n$ elementwise. The concentrated log-likelihood of $\lambda$ 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 $h_y(Y_{n,i}, \lambda) = \partial h(Y_{n,i}, \lambda)/\partial Y_{n,i}$. 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 $\theta$ 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 (2013), 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 validity of the proposed bootstrap procedure is formally established, in general and under the SAR model. Finally, the method is extended to the models of more than one nonlinear parameter.

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, followed by Monte Carlo results for the finite sample performance of the proposed method under the SAR model, where the effectiveness of the proposed method in correcting bias, variance, and hence in improving inference is fully demonstrated. Section 4 extends the proposed method to models of more than one nonlinear parameter. Section 5 concludes the paper.

# 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 introduce the bootstrap method for estimating quantities in the bias and variance formulas and prove its validity.

## 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

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

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 $\theta$, of the same dimension as $\theta$, and normalized to have order $O_p(n^{-1/2})$.[2] They obtained a third-order stochastic expansion for $\hat{\theta}_n$, and a

[2]This 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 $\mathrm{E}\psi_n(\theta_0) = 0$, where $\theta_0$ denotes the true value of the parameter vector $\theta$.

We note that the condition, $\mathrm{E}\psi_n(\theta_0) = 0$, 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 $\sigma^2$ 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 $\lambda$ is the scalar nonlinear parameter of which the estimation incurs bias that is difficult to correct, and given $\lambda$ the estimation of the parameter vector $\alpha$ has an analytical solution, either unbiased or easily bias-correctable. Let $\hat{\alpha}_n(\lambda)$ be the constrained estimator of $\alpha$ for a given $\lambda$ value. Let $\theta_0 = (\lambda_0, \alpha_0')'$. 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 $\lambda$ would typically be

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

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. We first focus on the cases where $\lambda$ is a scalar. The CEE looks identical to the JEE when $\theta$ 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 $\mathrm{E}\psi_n(\theta_0) = 0$. If $\hat{\alpha}_n(\lambda)$ is $\sqrt{n}$-consistent, it is typical that $\mathrm{E}[\tilde{\psi}_n(\lambda_0)] = O(n^{-1})$, i.e., the expectation goes to zero at an $n$-rate. If this is true, then $\mathrm{E}[\tilde{\psi}_n(\lambda_0)]$ constitutes an important term in the bias correction. In this case, the bias formulas need to be modified. As a consequence, the higher-order approximations to the variance need 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 $H_{rn}^{\circ} = H_{rn} - \mathrm{E}(H_{rn}), r = 1, 2, 3$. Define $\Omega_n = -\mathrm{E}(H_{1n})^{-1}$. Note that here and hereafter the expectation operator 'E' corresponds to the true model or the true parameter values $\theta_0$.

[3]Making 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.

Let $\Lambda$ be the parameter space of $\lambda$. 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:** *$\Lambda$ is compact with $\lambda_0$ being an interior point.* $\mathrm{E}(\tilde{\psi}_n) = O(n^{-1})$*, and $\hat{\lambda}_n$, as a solution of $\tilde{\psi}_n(\lambda) = 0$, is a $\sqrt{n}$-consistent estimator of $\lambda_0$.*

**Assumption B:** *$\tilde{\psi}_n(\lambda)$ is differentiable up to rth order for $\lambda$ in a neighborhood of $\lambda_0$,* $\mathrm{E}(H_{rn}) = O(1)$*, and $H_{rn}^{\circ} = O_p(n^{-\frac{1}{2}}), r = 1, 2, 3$.*

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

**Assumption D:** *$|H_{rn}(\lambda) - H_{rn}(\lambda_0)| \leq |\lambda - \lambda_0| U_n$ for $\lambda$ in a neighborhood of $\lambda_0, r = 1, 2, 3$, and* $\mathrm{E}(|U_n|) < C < \infty$ *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 $\mathrm{E}(\tilde{\psi}_n) = o(n^{-1/2})$ 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 $\mathrm{E}(\tilde{\psi}_n) = O(n^{-1})$. Further, the $\sqrt{n}$-consistency of $\hat{\lambda}_n$ implies $\tilde{\psi}_n = O_p(n^{-\frac{1}{2}})$. Assumptions B and C are the tightened versions of Assumptions 4 and 5 in Bao and Ullah (2007a). The conditions $\mathrm{E}(H_{rn}) = O(1)$ and $H_{rn}^{\circ} = O_p(n^{-\frac{1}{2}})$ are needed so that $H_{rn}$ in a relevant term can be replaced by $\mathrm{E}(H_{rn})$ 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:*

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

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

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 $\hat{\lambda}_n$:*

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

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

[4] Under a specific model with a specific estimation method, these generic conditions may be replaced by a set of weaker and more primitive conditions. 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 $m_{-1} = \Omega_n^2 \mathrm{E}(\tilde{\psi}_n^2)$ 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 $\mathrm{MSE}(\hat{\lambda}_n)$. 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 $\hat{\lambda}_n$:*

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

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

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 $\mathrm{E}(a_{-1/2}) = \Omega_n \mathrm{E}(\tilde{\psi}_n)$. 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, $2\Omega_n \mathrm{E}(\tilde{\psi}_n)$. 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 $\lambda$ are, respectively,

$$\hat{\lambda}_n^{\mathrm{bc2}} = \hat{\lambda}_n - \widehat{b}_{-1} \quad \text{and} \quad \hat{\lambda}_n^{\mathrm{bc3}} = \hat{\lambda}_n - \widehat{b}_{-1} - \widehat{b}_{-3/2}, \tag{10}$$

where $\widehat{b}_{-1}$ and $\widehat{b}_{-3/2}$ are the estimates of $b_{-1}$ and $b_{-3/2}$, respectively.

**Remark 1:** The practical implementation of $\hat{\lambda}_n^{\mathrm{bc3}}$ 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 $\widehat{b}_{-1}$ for feasible third-order bias correction. This issue is addressed in Corollary 2.5, in conjunction with the issue on the validity of using a bootstrap method to obtain the estimates $\widehat{b}_{-1}$ and $\widehat{b}_{-3/2}$.

While it is important to have higher-order expansions for $\mathrm{MSE}(\hat{\lambda}_n)$ for the purpose of efficiency comparison, it is more important to have higher-order expansions for the variances of $\hat{\lambda}_n$, $\hat{\lambda}_n^{\mathrm{bc2}}$ and $\hat{\lambda}_n^{\mathrm{bc3}}$ 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: $\mathrm{Var}(\hat{\lambda}_n) = m_{-1} + m_{-3/2} + O(n^{-2})$, and $\mathrm{Var}(\hat{\lambda}_n) = m_{-1} + m_{-3/2} + m_{-2} - b_{-1}^2 + O(n^{-\frac{5}{2}})$. Theoretically these are correct, but empirically they do not guarantee positiveness of the variance estimator when $n$ is not large. We thus propose having 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* $\hat{\lambda}_n$*:*

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

*where* $v_{-1} = \mathrm{Var}(a_{-1/2})$, $v_{-3/2} = 2\mathrm{Cov}(a_{-1/2},\ a_{-1})$, *and* $v_{-2} = 2\mathrm{Cov}(a_{-1/2},\ a_{-3/2}) + \mathrm{Var}(a_{-1} + a_{-3/2})$, *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 $\lambda$. However, the bias-corrected estimators, $\hat{\lambda}_n^{\mathrm{bc2}}$ and $\hat{\lambda}_n^{\mathrm{bc3}}$, contain additional random elements due to the estimation of the bias. Therefore, for improving finite sample inference for $\lambda$ it is more relevant to use $\mathrm{Var}(\hat{\lambda}_n^{\mathrm{bc2}})$ or $\mathrm{Var}(\hat{\lambda}_n^{\mathrm{bc3}})$ to calculate the standard errors of $\hat{\lambda}_n^{\mathrm{bc2}}$ or $\hat{\lambda}_n^{\mathrm{bc3}}$. We have the following corollary.

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

$$\mathrm{Var}(\hat{\lambda}_n^{\mathrm{bc3}}) = v_{-1} + v_{-3/2} + v_{-2} - 2\mathrm{ACov}(\hat{\lambda}_n, \hat{b}_{-1}) + O(n^{-\frac{5}{2}}), \tag{12}$$

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

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

## 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 $\mathrm{E}(H_{rn})$, $\mathrm{E}(\tilde{\psi}_n^2)$, $\mathrm{E}(H_{1n}\tilde{\psi}_n^2)$, etc. The analytical approach is to first find the closed form expressions for these expectations and then replace $\theta_0$ in the resulted expressions by its consistent estimator $\hat{\theta}_n$. In case that the error distribution is not fully specified, these expectations may involve higher order moments of the errors which have to be estimated as well. However, finding these expectations analytically is either very cumbersome or impossible (see Section 4 for more cases, and Section 5 for more discussions). Thus, alternative methods are highly desirable. We now introduce a simple bootstrap method for estimating these quantities. Consider a general model of the form

$$g(Z_n, \theta_0) = u_n, \tag{13}$$

where $u_n$ is the standardized disturbance vector of iid (not necessarily normal) components with zero mean, unit variance, and cumulative distribution function (CDF) $\mathcal{F}_0$. Clearly, the SAR model given in (1) can be written in this form: $\sigma_0^{-1}[A_n(\lambda_0)Y_n - X_n\beta_0] = u_n$.

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 original data, and $\hat{\mathcal{F}}_n$ be the empirical distribution function (EDF) of $\hat{u}_n$ (centered). The bootstrap estimates of the quantities in the bias and variance corrections are thus;

$$\widehat{\mathrm{E}}(\tilde{\psi}_n^i H_{rn}^j) = \mathrm{E}^*[\tilde{\psi}_n^i(\hat{u}_n^*, \hat{\theta}_n) H_{rn}^j(\hat{u}_n^*, \hat{\theta}_n)], \quad i, j = 0, 1, 2, \ldots, \quad r = 1, 2, 3, \tag{14}$$

where $\mathrm{E}^*$ denotes the expectation with respect to $\hat{\mathcal{F}}_n$, and $\hat{u}_n^*$ is a vector of $n$ random draws from $\hat{\mathcal{F}}_n$. To make (14) practically feasible, we propose the following bootstrap procedure:

1. *Compute* $\hat{\theta}_n$ *defined by* (5), $\hat{u}_n = g(Z_n, \hat{\theta}_n)$ *by* (13), *and EDF* $\hat{\mathcal{F}}_n$ *of centered* $\hat{u}_n$*;*
2. *Draw a random sample of size* $n$ *from* $\hat{\mathcal{F}}_n$*, and denote the resampled vector by* $\hat{u}_{n,b}^*$*,*
3. *Compute* $\tilde{\psi}_n(\hat{u}_{n,b}^*, \hat{\theta}_n)$ *and* $H_{rn}(\hat{u}_{n,b}^*, \hat{\theta}_n), r = 1, 2, 3$*;*
4. *Repeat steps 2.-3. for* $B$ *times, to give approximate bootstrap estimates as,*

$$\mathrm{E}^*[\tilde{\psi}_n^i(\hat{u}_n^*, \hat{\theta}_n) H_{rn}^j(\hat{u}_n^*, \hat{\theta}_n)] \doteq \tfrac{1}{B}\textstyle\sum_{b=1}^B \tilde{\psi}_n^i(\hat{u}_{n,b}^*, \hat{\theta}_n) H_{rn}^j(\hat{u}_{n,b}^*, \hat{\theta}_n),$$

*for* $i, j = 0, 1, 2, \ldots, \quad r = 1, 2, 3,$

where the approximations in the last step can be made arbitrarily accurate by choosing an arbitrarily large $B$, leading directly to the bootstrap estimates of $b_{-s/2}$, $m_{-s/2}$ and $v_{-s/2}$ appeared in Corollaries 2.1-2.4, except $\mathrm{ACov}(\hat{\lambda}_n, \hat{b}_{-1})$ in Corollary 2.4, which will be addressed in the next subsection under the 'specified' model (13). 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 $\theta$ is thus avoided, which makes this bootstrap procedure time-efficient.

## 2.3 Validity of the bootstrap method

As in Remark 2, the validity of the bootstrap procedure for third-order bias correction centers on the validity of replacing $b_{-1}$ by its bootstrap estimate $\widehat{b}_{-1}$. Clearly, $b_{-s/2}$ depends on $\theta_0$. It may also depend on $\mathcal{F}_0$ through some of its higher order moments $\mu_0 \equiv \mu(\mathcal{F}_0)$. In general, denote $b_{-s/2} \equiv b_{-s/2}(\vartheta_0)$ where $\vartheta_0 = (\theta_0', \mu_0')'$, $s = 2, 3$. Then, the bootstrap estimates of $b_{-s/2}$ must be such that $\widehat{b}_{-s/2} \equiv b_{-s/2}(\hat{\vartheta}_n)$ where $\hat{\vartheta}_n = (\hat{\theta}_n', \hat{\mu}_n')'$ and $\hat{\mu}_n \equiv \mu(\hat{\mathcal{F}}_n)$, the corresponding moments of $\hat{\mathcal{F}}_n$.[5] Let $\mathcal{N}_{\vartheta_0}$ be a neighborhood of $\vartheta_0$.

**Assumption E:** *(i)* $\hat{\vartheta}_n$ *is* $\sqrt{n}$*-consistent for* $\vartheta_0$*, (ii)* $\mathrm{Bias}(\hat{\vartheta}_n) = O(n^{-1})$*, and (iii)* $b_{-1}(\vartheta)$ *is continuously differentiable and* $b_{-3/2}(\vartheta)$ *is differentiable in* $\vartheta \in \mathcal{N}_{\vartheta_0}$*.*

Assumptions E($i$) and E($ii$) are satisfied by the nonlinear estimator $\hat{\lambda}_n$, and thus are expected to hold for $\hat{\vartheta}_n$ (see Lemma B.6). Assumption E($iii$) is satisfied if the $\tilde{\psi}_n$ function is smooth enough. The validity of bootstrap bias corrections is established below.

[5] Introducing $\mu_0$ and $\hat{\mu}_n$ is only for theoretical purpose, as practical implementation of the bootstrap method does not require $\hat{\mu}_n$. This stands in contrast to the analytical approach.

**Corollary 2.5:** *Under the model specified by* (13)*, Assumptions A-E, and the assumption that a quantity bounded in probability has a finite expectation, we have:*

$$\mathrm{E}(\widehat{b}_{-1}) = b_{-1} + O(n^{-2}) \ \ \textit{and} \ \ \mathrm{E}(\widehat{b}_{-3/2}) = b_{-3/2} + O(n^{-2}). \tag{15}$$

*It follows that* $\mathrm{Bias}(\hat{\lambda}_n^{\mathrm{bc2}}) = O(n^{-\frac{3}{2}})$ *and* $\mathrm{Bias}(\hat{\lambda}_n^{\mathrm{bc3}}) = O(n^{-2})$.

The results of Corollary 2.5 show that the bias-corrected estimators defined in (10) are valid in the sense that replacing $b_{-1}$ by its bootstrap estimate $\widehat{b}_{-1}$ does not induce additional bias of order lower than $O(n^{-2})$. It is evident that replacing the third-order term $b_{-3/2}$ by its bootstrap estimate $\widehat{b}_{-3/2}$ only induces an additional bias of order $O(n^{-2})$ or higher.

The validity of the bootstrap implementation of the third-order variance correction given in (12) rests on the validity of replacing $v_{-1}$ by its estimate. Similarly to the case of bias correction, write $v_{-s/2} = v_{-s/2}(\vartheta_0)$ and let $\widehat{v}_{-s/2} \equiv v_{-s/2}(\hat{\vartheta}_n)$ be the bootstrap estimates of $v_{-s/2}$, $s = 2, 3, 4$. Unlike the case of third-order bias correction where $\mathrm{Bias}(\widehat{b}_{-1})$ has the same order as the error term, $\mathrm{Bias}(\widehat{v}_{-1})$ may not have the desired order for a third-order variance correction. Thus, a further correction on $\widehat{v}_{-1}$ may be necessary. Note that $v_{-1}$ is the first-order variance of $\hat{\lambda}_n$ based on CEE, which relates to $\nu_{-1}$, the first-order variance of $\hat{\lambda}_n$ based on JEE, through $v_{-1} = \nu_{-1} + \nu_{-3/2} + \nu_{-2} + O(n^{-\frac{5}{2}})$ where $\nu_{-s/2} = O(n^{-s/2}), s = 2, 3, 4$. Thus, to bias-correct $\widehat{v}_{-1}$ it is sufficient to bias-correct $\widehat{\nu}_{-1}$. The higher-order terms $\nu_{-3/2}$ and $\nu_{-2}$ are not needed as seen from the proof of Corollary 2.6. As $\nu_{-1}$ is typically an explicit and smooth function of $\vartheta_0$, and $\hat{\mu}_n$ is an explicit and smooth function of $\hat{\theta}_n$, a Taylor series expansion of $v_{-1}(\hat{\vartheta}_n)$ and then of $\hat{\mu}_n \equiv \mu_n(\hat{\theta}_n)$ leads to a second-order bias-corrected $\widehat{v}_{-1}$:

$$\widehat{v}_{-1}^{\mathrm{bc2}} = \widehat{v}_{-1} - \widehat{\nu}'_{-1,\theta_0}\widehat{\mathbf{b}}_{-1} - \widehat{\nu}'_{-1,\mu_0}\widehat{\mathbf{c}}_{-1} - \widehat{\mathbf{d}}_{-2}, \tag{16}$$

where $\mathbf{b}_{-1}$ is the second-order bias of $\hat{\theta}_n$ which can be obtained from the second-order stochastic expansion of $\hat{\theta}_n$ based on the JEE defined in (5) or simply using $b_{-1}$ and then further bias-correcting $\hat{\alpha}(\hat{\lambda}_n^{\mathrm{bc2}})$; $\mathbf{c}_{-1}$ is the second-order bias of $\hat{\mu}_n$ which relates to $\mathbf{b}_{-1}$ through a Taylor expansion of $\mu_n(\hat{\theta}_n)$; $\mathbf{d}_{-2} = \frac{1}{2}\mathrm{tr}[\nu_{-1,\vartheta_0\vartheta_0'}\mathrm{AVar}(\hat{\vartheta}_n)]$; and $\nu_{-1,\vartheta} = \frac{\partial}{\partial\vartheta}\nu_{-1}(\vartheta)$ and $\nu_{-1,\vartheta\vartheta'} = \frac{\partial^2}{\partial\vartheta\partial\vartheta'}\nu_{-1}(\vartheta)$, partitioned according to $\theta$ and $\mu$. To ease the presentation, the detailed derivation of (16) is given in Appendix A along the proof of Corollary 2.6, and the exact expressions of $\mathbf{b}_{-1}$, $\mathbf{c}_{-1}$ and $\mathbf{d}_{-2}$ are given in (A-12), (A-14) and (A-15).

Let $b_{-1,\vartheta} = \frac{\partial}{\partial\vartheta}b_{-1}(\vartheta) = (b'_{-1,\theta}, b'_{-1,\mu})'$. Let $\mathcal{F}_n$ be the EDF of $u_n \equiv u_n(\theta_0)$, $\mu_{n,r}$ the $r$th moment of $\mathcal{F}_n$, $\dot{\mu}_{n,r} = \frac{\partial}{\partial\theta_0}\mu_{n,r}$ and $\ddot{\mu}_{n,r} = \frac{\partial}{\partial\theta_0'}\dot{\mu}_{n,r}$. Put $\mu_n = (\mu_{n,r}, r = 3, 4, \ldots)'$. For the validity of (16) and our third-order variance correction, the following conditions are needed.

**Assumption F:** *(i)* $v_{-1}(\vartheta)$ *is twice differentiable,* $v_{-3/2}(\vartheta)$ *is continuously differentiable, and* $v_{-2}(\vartheta)$ *is differentiable in* $\vartheta \in \mathcal{N}_{\vartheta_0}$*, (ii)* $b'_{-1,\mu_0}\mathrm{Cov}(\tilde{\psi}_n, \mu_n) = O(n^{-\frac{5}{2}})$*, and (iii)* $\dot{\mu}_{n,r} = \mathrm{E}(\dot{\mu}_{n,r}) + O_p(n^{-\frac{1}{2}})$ *and* $\ddot{\mu}_{n,r} = \mathrm{E}(\ddot{\mu}_{n,r}) + O_p(n^{-\frac{1}{2}})$, $r = 3, 4, \ldots$.

The smoothness conditions in F($i$) ensure the proper order of the quantities in certain expansions. Assumption F($ii$) says that either the dependence of $b_{-1}$ on $\mu_0$ is weak or the correlation between $\tilde{\psi}_n$ and $\mu_n$ is weak. This is reasonable considering the fact that $\tilde{\psi}_n$ is the concentrated estimating function of $\lambda_0$ and $\mu_n$ represents higher-order moments of $\mathcal{F}_n$. Assumption F($iii$) follows from some smoothness conditions on the $g$ function in (13).

**Corollary 2.6:** *Under Model* (13)*, Assumptions A-F, and the assumption that a quantity bounded in probability has a finite expectation, the variance expansion* (12) *becomes:*

$$\mathrm{Var}(\hat{\lambda}_n^{\mathrm{bc3}}) = v_{-1} + v_{-3/2} + v_{-2} - 2\mathrm{E}(\tilde{b}'_{-1,\theta_0})\mathrm{ACov}(\hat{\lambda}_n, \hat{\theta}_n) + O(n^{-\frac{5}{2}}), \tag{17}$$

*where* $\tilde{b}_{-1,\theta_0} = \frac{\partial}{\partial \hat{\theta}_n} b_{-1}(\hat{\theta}_n, \mu_n(\hat{\theta}_n))|_{\hat{\theta}=\theta_0}$*, and* $\mathrm{E}(\widehat{v}_{-1}^{\mathrm{bc2}} + \widehat{v}_{-3/2} + \widehat{v}_{-2}) = v_{-1} + v_{-3/2} + v_{-2} + O(n^{-\frac{5}{2}})$*. Thus, the bootstrap estimate of* $\mathrm{Var}(\hat{\lambda}_n^{\mathrm{bc3}})$ *with* $\widehat{v}_{-1}^{\mathrm{bc2}}$ *has a bias of order* $O(n^{-\frac{5}{2}})$*.*

Note that $\mathrm{ACov}(\hat{\theta}_n, \hat{\lambda}_n)$ often has a closed-form expression and hence can be estimated by the plug-in method. As $\tilde{b}_{-1,\theta_0} - \mathrm{E}(\tilde{b}_{-1,\theta_0}) = O_p((n^{-\frac{3}{2}}))$, $\mathrm{E}(\tilde{b}_{-1,\theta_0})$ in (17) can be replaced by $\tilde{b}_{-1,\theta_0}$ of which the bootstrap estimate can be obtained simply by numerical differentiation. When the underlining JEE involves only the linear and quadratic forms of $u_n$, as in many applications, $\nu_{-1}$ typically depends on the 3rd and 4th moments of $\mathcal{F}_0$ and in a linear manner. See Sections 3 and 5 for more discussions. 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 established therein.

## 3 Bias and Variance Corrections for SAR Model

We now consider the estimation of spatial lag parameter $\lambda$ 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 $\lambda$ enters 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 $\lambda$, 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 $\lambda$ and $\beta$ may not be $\sqrt{n}$-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 bias 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 Rilstone et al. (1996). Their results, though limited to the pure SAR model with normal errors, shed much light on a general solution to the bias problem of the general SAR model (or a class of similar models). Bao (2013) 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 $\sigma^2$ unknown in the pure SAR model, Bao and Ullah (2007a, p.400) advocate the use of concentrated likelihood function of $\lambda$ 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, $\lambda$ is the parameter of primary interest as, given $\lambda$, the model reduces to a linear regression, and the constrained QMLEs $\hat{\beta}_n(\lambda)$ and $\hat{\sigma}_n^2(\lambda)$ are either unbiased or can be made unbiased. More importantly, as the general result in Section 2 suggests, use of CEE captures the additional second-order bias inherited from the estimation of linear or scale parameters.

### 3.1 The main results

The QMLE $\hat{\lambda}_n$ of the spatial parameter $\lambda$, 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,

$$\tilde{\psi}_n(\lambda) = -h_n T_{0n}(\lambda) + h_n R_{1n}(\lambda), \tag{18}$$

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

$$h_n^{-1}H_{1n}(\lambda) = -T_{1n}(\lambda) - R_{2n}(\lambda) + 2R_{1n}^2(\lambda), \tag{19}$$

$$h_n^{-1}H_{2n}(\lambda) = -2T_{2n}(\lambda) - 6R_{1n}(\lambda)R_{2n}(\lambda) + 8R_{1n}^3(\lambda), \tag{20}$$

$$h_n^{-1}H_{3n}(\lambda) = -6T_{3n}(\lambda) + 6R_{2n}^2(\lambda) - 48R_{1n}^2(\lambda)R_{2n}(\lambda) + 48R_{1n}^4(\lambda), \tag{21}$$

where $T_{rn}(\lambda) = n^{-1}\mathrm{tr}(G_n^{r+1}(\lambda)), r = 0, 1, 2, 3$, $G_n(\lambda) = W_n A_n^{-1}(\lambda)$,[6]

$$R_{1n}(\lambda) = \frac{Y_n' A_n'(\lambda) M_n W_n Y_n}{Y_n' A_n'(\lambda) M_n A_n(\lambda) Y_n} \quad \text{and} \quad R_{2n}(\lambda) = \frac{Y_n' W_n' M_n W_n Y_n}{Y_n' A_n'(\lambda) M_n A_n(\lambda) Y_n}. \tag{22}$$

Clearly the function $\tilde{\psi}_n(\lambda)$ defined in (18) leads to a concentrated estimating equation, and fits into the general framework described in Section 2. The difference is that the quantity

[6] The 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 $\hat{\lambda}_n^{\mathrm{bc3}}$ performs consistently better than $\hat{\lambda}_n^{\mathrm{bc2}}$. See Yang (2010a).

$h_n$ 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 the possibility that $h_n$ is unbounded.

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 $\eta_n = \frac{1}{\sigma_0} G_n X_n \beta_0$. 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 $\lambda_0$ is in the interior of a compact set $\Lambda$.*

**Assumption 2:** *The innovations $\{u_{n,i}\}$ are iid with mean zero, variance one, and CDF $\mathcal{F}_0$. $\mathrm{E}|u_{n,i}|^{4+\gamma}$ 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 $h_n^{1+\epsilon}/n \to 0$ for some $\epsilon > 0$ as $n \to \infty$. 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 $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ exists and is nonsingular.*

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

Assumptions 1-6 listed above are 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 $\lambda_0$. 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 behavior of the derivatives $H_{rn}, r = 1, 2, 3$. Denote $\hat{\sigma}_{n0}^2 \equiv \hat{\sigma}_n^2(\lambda_0)$.

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

7 are both $O_p(n^{-1})$, suggesting that their expectations would be of order $O(n^{-1})$, which is smaller than the desired 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) $h_n^s \mathrm{E}[(R_{1n} - \mathrm{E}R_{1n})^s] = O((\frac{h_n}{n})^{\frac{1}{2}}), s = 2, 3, 4$; (ii) $h_n^2 \mathrm{E}[(R_{2n} - \mathrm{E}R_{2n})^2] = O((\frac{h_n}{n})^{\frac{1}{2}})$; and (iii) $h_n^{s+1} \mathrm{E}[(R_{1n} - \mathrm{E}R_{1n})^s (R_{2n} - \mathrm{E}R_{2n})] = O((\frac{h_n}{n})^{\frac{1}{2}}), s = 1, 2$.

The three conditions in Assumption 8 are in fact rather weak since, following the results of Lemma 3.1, 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)* $h_n R_{1n} = O_p(1)$, $\mathrm{E}(h_n R_{1n}) = O(1)$, *and* $h_n R_{1n} = \mathrm{E}(h_n R_{1n}) + O_p((\frac{h_n}{n})^{\frac{1}{2}})$;
*(ii)* $h_n R_{2n} = O_p(1)$, $\mathrm{E}(h_n R_{2n}) = O(1)$, *and* $h_n R_{2n} = \mathrm{E}(h_n R_{2n}) + O_p((\frac{h_n}{n})^{\frac{1}{2}})$.

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:*

$$\hat{\lambda}_n - \lambda_0 = a_{-1/2} + a_{-1} + a_{-3/2} + O_p((\tfrac{h_n}{n})^2), \tag{23}$$

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

$$\begin{aligned} a_{-3/2} = &\ \Omega_n \tilde{\psi}_n + 2\Omega_n^2 H_{1n} \tilde{\psi}_n + \Omega_n^3 \mathrm{E}(H_{2n}) \tilde{\psi}_n^2 + \Omega_n^3 H_{1n}^2 \tilde{\psi}_n + \tfrac{1}{2}\Omega_n^3 H_{2n} \tilde{\psi}_n^2 \\ &+ \tfrac{3}{2}\Omega_n^4 \mathrm{E}(H_{2n}) H_{1n} \tilde{\psi}_n^2 + \tfrac{1}{2}\Omega_n^5 \mathrm{E}(H_{2n})^2 \tilde{\psi}_n^3 + \tfrac{1}{6}\Omega_n^4 \mathrm{E}(H_{3n}) \tilde{\psi}_n^3, \end{aligned}$$

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

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 $c_{1n} = \{\Omega_n, 0'_{6\times 1}\}'$, $c_{2n} = \{\Omega_n,\ \Omega_n^2,\ \frac{1}{2}\Omega_n^3 \mathrm{E}(H_{2n}), 0'_{4\times 1}\}'$, $c_{3n} = \{\Omega_n,\ 2\Omega_n^2,\ \Omega_n^3 \mathrm{E}(H_{2n}),\ \Omega_n^3, \frac{1}{2}\Omega_n^3,\ \frac{3}{2}\Omega_n^4 \mathrm{E}(H_{2n}),\ \frac{1}{2}\Omega_n^5 \mathrm{E}(H_{2n})^2 + \frac{1}{6}\Omega_n^4 \mathrm{E}(H_{3n})\}'$, $C_{2n} = c_{1n} + c_{2n}$ and $C_{3n} = c_{1n} + c_{2n} + c_{3n}$. Let

$$\zeta_n = \{\tilde{\psi}_n,\ H_{1n}\tilde{\psi}_n,\ \tilde{\psi}_n^2,\ H_{1n}^2 \tilde{\psi}_n,\ H_{2n}\tilde{\psi}_n^2,\ H_{1n}\tilde{\psi}_n^2,\ \tilde{\psi}_n^3\}'. \tag{24}$$

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

**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* $\hat{\lambda}_n$*:*

$$\mathrm{Bias}(\hat{\lambda}_n) = b_{-1} + b_{-3/2} + O((\tfrac{h_n}{n})^2), \tag{25}$$

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

$$\hat{\lambda}_n^{\mathrm{bc2}} = \hat{\lambda}_n - \widehat{C}'_{2n}\widehat{\mathrm{E}}(\zeta_n) \quad \text{and} \quad \hat{\lambda}_n^{\mathrm{bc3}} = \hat{\lambda}_n - \widehat{C}'_{3n}\widehat{\mathrm{E}}(\zeta_n), \tag{26}$$

*where the quantity with an* $\widehat{\ }$ *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 $2\Omega_n \mathrm{E}(\tilde{\psi}_n)$, which is of order $O((\frac{h_n}{n}))$ 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* $\hat{\lambda}_n$*:*

$$\mathrm{MSE}(\hat{\lambda}_n) = m_{-1} + m_{-3/2} + m_{-2} + O((\tfrac{h_n}{n})^{\frac{5}{2}}), \tag{27}$$

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

It is seen that introducing the notation $c_{in}$ and $\zeta_n$ 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* $\hat{\lambda}_n$*:*

$$\mathrm{Var}(\hat{\lambda}_n) = v_{-1} + v_{-3/2} + v_{-2} + O((\tfrac{h_n}{n})^{\frac{5}{2}}); \tag{28}$$

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

Note that $v_{-1} + v_{-3/2} + v_{-2} = \mathrm{Var}(a_{-1/2} + a_{-1} + a_{-3/2}) = C'_{3n}\mathrm{Var}(\zeta_n)C_{3n}$, which guarantees the positiveness of the third-order variance estimate. Similarly, one may use $\mathrm{Var}(a_{-1/2} + a_{-1}) = C'_{2n}\mathrm{Var}(\zeta_n)C_{2n}$ for second-order variance correction. The asymptotically equivalent but simpler variance expansions, $\mathrm{Var}(\hat{\lambda}_n) = m_{-1} + m_{-3/2} + O((\frac{h_n}{n})^2)$ and $\mathrm{Var}(\hat{\lambda}_n) = m_{-1} + m_{-3/2} + m_{-2} - b_{-1}^2 + O((\frac{h_n}{n})^{\frac{5}{2}})$, may be used. 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 (28) be followed in the practical applications.

One important issue left is the variances of the bias-corrected estimators, i.e., $\mathrm{Var}(\hat{\lambda}_n^{\mathrm{bc2}})$ and $\mathrm{Var}(\hat{\lambda}_n^{\mathrm{bc3}})$, which is more relevant in improving the finite sample inferences for $\lambda$.

**Corollary 3.4.** *Under the assumptions of Corollary 3.3, assume further that* $\widehat{b}_{-1} - b_{-1} = O((\frac{h_n}{n})^{\frac{3}{2}})$ *and* $\widehat{b}_{-3/2} - b_{-3/2} = O((\frac{h_n}{n})^2)$*. We have the third-order expansion:*

$$\mathrm{Var}(\hat{\lambda}_n^{\mathrm{bc3}}) = v_{-1} + v_{-3/2} + v_{-2} - 2\mathrm{ACov}(\hat{\lambda}_n, \widehat{b}_{-1}) + O((\tfrac{h_n}{n})^{\frac{5}{2}}) \tag{29}$$

*and* $\mathrm{Var}(\hat{\lambda}_n^{\mathrm{bc2}}) = \mathrm{Var}(\hat{\lambda}_n^{\mathrm{bc3}}) + O((\frac{h_n}{n})^{\frac{5}{2}})$*, where 'ACov' denotes the asymptotic covariance.*

It turns out that the variance of $\hat{\lambda}_n$ agrees with the variances of $\hat{\lambda}_n^{\text{bc2}}$ and $\hat{\lambda}_n^{\text{bc3}}$ to the second-order $O((\frac{h_n}{n})^{\frac{3}{2}})$ but differ on the third-order $O((\frac{h_n}{n})^2)$. With the results of Corollaries 3.1-3.4, various $t$-ratios for inference for $\lambda$ can be formed:

$$t_{ij} = \frac{\hat{\lambda}_n^{\text{bc}i} - \lambda_0}{\sqrt{\widehat{V}_j(\hat{\lambda}_n^{\text{bc}i})}}, \tag{30}$$

where $\hat{\lambda}_n^{\text{bc}i}$ denotes the $i$th-order bias-corrected estimator, and $\widehat{V}_j(\cdot)$ 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 $\lambda$.

It would be interesting to see the impacts of bias correcting $\hat{\lambda}_n$ on the estimation of the regression coefficients $\beta$ and the error variance $\sigma^2$. As discussed in the introduction, in estimating an econometric model with nonlinear, linear and scale parameters, the estimation of the nonlinear parameter is the main source of bias and the main difficulty in bias correction. Now, given the bias corrected QMLE of $\lambda$, $\hat{\lambda}_n^{\text{bc2}}$ say, what are the finite sample properties of $\hat{\beta}_n(\hat{\lambda}_n^{\text{bc2}})$ and $\hat{\sigma}_n^{*2}(\hat{\lambda}_n^{\text{bc2}}) = \frac{n}{n-p}\hat{\sigma}_n^2(\hat{\lambda}_n^{\text{bc2}})$? We have, for $\tilde{\lambda}_n = \hat{\lambda}_n$ or $\hat{\lambda}_n^{\text{bc2}}$:

$$\begin{aligned} \text{Bias}[\hat{\beta}_n(\tilde{\lambda}_n)] &= -b_n \text{Bias}(\tilde{\lambda}_n) - \sigma_0 \Omega_n \text{E}(\xi_n \tilde{\psi}_n) + O(h_n n^{-\frac{3}{2}}), & (31) \\ \text{Bias}[\hat{\sigma}_n^{*2}(\tilde{\lambda}_n)] &= -2\sigma_0^2 T_{0n} \text{Bias}(\tilde{\lambda}_n) + 2\Omega_n \text{E}(\zeta_n \tilde{\psi}_n) + \sigma_0^2 d_n \Omega_n^2 \text{E}(\tilde{\psi}_n^2) + O(h_n n^{-\frac{3}{2}}), & (32) \end{aligned}$$

where $b_n = (X_n' X_n)^{-1} X_n' G_n X_n \beta_0 = O(1)$, $\xi_n = (X_n' X_n)^{-1} X_n' G_n u_n = O_p(n^{-\frac{1}{2}})$, $\zeta_n = \sigma_0^2 T_{0n} - \frac{1}{n} Y_n' A_n' M_n W_n Y_n = O_p((h_n n)^{-\frac{1}{2}})$, and $d_n = T_{1n} + \frac{1}{n}\eta_n'\eta_n = O(h_n^{-1})$.

As $\text{Bias}(\hat{\lambda}_n) = O(\frac{h_n}{n})$ and $\text{Bias}(\hat{\lambda}_n^{\text{bc2}}) = O((\frac{h_n}{n})^{\frac{3}{2}})$, the above results show clearly the impacts of using $\tilde{\lambda}_n = \hat{\lambda}_n$ or $\hat{\lambda}_n^{\text{bc2}}$ on the finite sample performance of $\hat{\beta}_n(\tilde{\lambda}_n)$ and $\hat{\sigma}_n^{*2}(\tilde{\lambda}_n)$. Monte Carlo results (not reported for brevity) show that the biases of $\hat{\beta}_n(\hat{\lambda}_n^{\text{bc2}})$ and $\hat{\sigma}_n^{*2}(\hat{\lambda}_n^{\text{bc2}})$ are very small, and can be significantly smaller than those of $\hat{\beta}_n(\hat{\lambda}_n)$ and $\hat{\sigma}_n^{*2}(\hat{\lambda}_n)$. Although in theory the second term in (31) is $O(\frac{\sqrt{h_n}}{n})$ and the middle two terms in (32) are $O(n^{-1})$, Monte Carlo results do not show that they have significant contributions to bias. Indeed, we can show that when $\lambda = 0$, the second term in (31) becomes $O(\frac{h_n}{n^2})$. These are consistent with our discussion in the introduction. In any case, the above results can be used to further correct $\hat{\beta}_n(\hat{\lambda}_n^{\text{bc2}})$ and $\hat{\sigma}_n^{*2}(\hat{\lambda}_n^{\text{bc2}})$ so that they become *truly* unbiased up to $O(\frac{h_n}{n})$.

### 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 (18) to (21) 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,

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

and then in the resulted closed form expressions replace $\theta_0$ by $\hat{\theta}_n$ and certain moments of $\mathcal{F}_0$ by their estimates, 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.[7] For the general SAR model with normal errors, we managed to derive only a second-order bias formula (requiring only $\mathrm{E}(R_{1n}^k), k=1,2,3$, $\mathrm{E}(R_{2n})$ and $\mathrm{E}(R_{1n}R_{2n})$) 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.

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 $\eta_n = \sigma_0^{-1} G_n X_n \beta_0$. 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 $\theta_0$. 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* $\hat{u}_n = \hat{\sigma}_n^{-1}(\hat{A}_n Y_n - X_n\hat{\beta}_n)$*, where* $\widehat{A}_n = I_n - \hat{\lambda}_n W_n$*,*
3. *Resample* $\hat{u}_n$ *(centered) in the usual way, and denote the resampled vector by* $\hat{u}_{n,b}^*$*,*
4. *Compute* $R_{1n}(\hat{u}_{n,b}^*, \hat{\theta}_n)$ *and* $R_{2n}(\hat{u}_{n,b}^*, \hat{\theta}_n)$*, and thus* $H_n(\hat{u}_{n,b}^*, \hat{\theta}_n)$*, and* $\zeta_n(\hat{u}_{n,b}^*, \hat{\theta}_n)$*,*
5. *Repeat steps 3-4 B times to give sequences of bootstrapped values for* $H_n$ *and* $\zeta_n$*.*

[7]In 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 $i,j = 0,1,2,...$, 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 (2013) 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.

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 $\zeta_n$) are, respectively, $\widehat{\mathrm{E}}(\tilde{\psi}_n^2) = \frac{1}{B}\sum_{b=1}^{B}\tilde{\psi}_n^2(\hat{u}_{n,b}^*, \hat{\theta}_n)$, and $\widehat{\mathrm{Var}}(\tilde{\psi}_n^2) = \frac{1}{B}\sum_{b=1}^{B}\tilde{\psi}_n^4(\hat{u}_{n,b}^*, \hat{\theta}_n) - [\widehat{\mathrm{E}}(\tilde{\psi}_n^2)]^2$. With the quantities $c_{in}, i = 1, 2, 3$, and $\zeta_n$ 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:

$$\begin{aligned} \widehat{\mathrm{E}}(H_n) &= \tfrac{1}{B}\textstyle\sum_{b=1}^{B} H_n(\hat{u}_{n,b}^*, \hat{\theta}_n), \\ \widehat{\mathrm{E}}(\zeta_n) &= \tfrac{1}{B}\textstyle\sum_{b=1}^{B} \zeta_n(\hat{u}_{n,b}^*, \hat{\theta}_n), \quad \text{and} \\ \widehat{\mathrm{E}}(\zeta_n\zeta_n') &= \tfrac{1}{B}\textstyle\sum_{b=1}^{B} \zeta_n(\hat{u}_{n,b}^*, \hat{\theta}_n)\zeta_n'(\hat{u}_{n,b}^*, \hat{\theta}_n). \end{aligned}$$

See, e.g., Efron (1979) and Lahiri (2003) for details on the general bootstrap principles.

### 3.3 Validity of the bootstrap method

We first provide some heuristic arguments for the validity of the proposed bootstrap method. Formal results are given in Corollaries 3.5 and 3.6 and proofs are given in Appendix B. As all the quantities in the expansions are smooth functions of either $R_{1n}$ only or both $R_{1n}$ and $R_{2n}$. It may suffice to argue that the above bootstrap procedure leads to *valid* estimates of $\mathrm{E}(R_{1n})$ and $\mathrm{E}(R_{2n})$. Recall $\mathcal{F}_0$ is the CDF of $u_{ni}$. Let $\mu_0 = \mu(\mathcal{F}_0)$ be the higher-order moments of $\mathcal{F}_0$ that $\mathrm{E}(R_{1n})$ and $\mathrm{E}(R_{2n})$ depend upon. Let $\hat{\mathcal{F}}_n$ be the EDF of $\hat{u}_n$. Consistency of $\hat{\theta}_n$ ensures the consistency of using $\hat{\mu}_n = \mu(\hat{\mathcal{F}}_n)$ to estimate $\mu_0$ (see Lemma B.6, Appendix B). If, in the real world, one knew both $\mathcal{F}_0$ and $\theta_0$, one may approximate $\mathrm{E}(R_{jn}),\ j = 1, 2$, to an arbitrary accuracy by

$$\mathrm{E}(R_{jn}) \doteq \tfrac{1}{M}\textstyle\sum_{m=1}^{M} R_{jn}(u_{n,m}, \theta_0),\ j = 1, 2,$$

where $u_{n,m}$ is a random $n$-vector drawn from $\mathcal{F}_0$ and $M$ is a large positive integer. If, however, one knew $\mathcal{F}_0$ but not $\theta_0$, one may estimate $\mathrm{E}(R_{jn})$ by

$$\overline{\mathrm{E}}(R_{jn}) = \Big(\tfrac{1}{M}\textstyle\sum_{m=1}^{M} R_{jn}(u_{n,m}, \theta_0)\Big)\Big|_{\theta_0=\hat{\theta}_n} = \tfrac{1}{M}\textstyle\sum_{m=1}^{M} R_{jn}(u_{n,m}, \hat{\theta}_n),\ j = 1, 2,$$

where the estimation error comes only from the estimation of the model parameters. If, instead, one knew $\theta_0$ but not $\mathcal{F}_0$, one may estimate $\mathrm{E}(R_{jn})$ using the bootstrap estimates

$$\widetilde{\mathrm{E}}(R_{jn}) = \mathrm{E}^*(R_{jn}(u_{n,b}^*, \theta_0)) \doteq \tfrac{1}{B}\textstyle\sum_{b=1}^{M} R_{jn}(u_{n,b}^*, \theta_0),\ j = 1, 2,$$

for a large $B$, where $u_{n,b}^* \overset{iid}{\sim} \mathcal{F}_n$, $\mathcal{F}_n$ is the EDF of $u_n$, and $\mathrm{E}^*$ denotes expectation with respect to $\mathcal{F}_n$. The error in estimating $\mathrm{E}(R_{jn})$ now comes from the estimation of $\mathcal{F}_0$.

In reality, however, one knows neither $\theta_0$ nor $\mathcal{F}_0$. Under the bootstrap world, these unknown quantities are made 'known' to be their estimates $\hat{\theta}_n$ and $\hat{\mathcal{F}}_n$, and the bootstrap DGP that mimics the real world DGP given in (1) is

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

where $\hat{u}^*_{n,b} \overset{iid}{\sim} \hat{\mathcal{F}}_n$, and $\hat{\theta}_n = (\hat{\lambda}_n, \hat{\beta}'_n, \hat{\sigma}^2_n)'$ are the estimates of $\theta_0 = (\lambda_0, \beta'_0, \sigma^2_0)'$ based on the original data. Based on the generated bootstrap data $\{Y^*_{n,b}, X_n, W_n\}$ and the bootstrap parameters $\hat{\theta}_n$, one computes the bootstrap analogue of $R_{1n}$ and $R_{2n}$ defined in (22) as

$$R^*_{1n,b} = \frac{Y^{*\prime}_{n,b}\hat{A}'_n M_n W_n Y^*_{n,b}}{Y^{*\prime}_{n,b}\hat{A}'_n M_n \hat{A}_n Y^*_{n,b}} \quad \text{and} \quad R^*_{2n,b} = \frac{Y^{*\prime}_{n,b} W_n M_n W_n Y^*_{n,b}}{Y^{*\prime}_{n,b}\hat{A}'_n M_n \hat{A}_n Y^*_{n,b}},$$

which after simplifications become $R_{1n}(\hat{u}^*_{n,b}, \hat{\theta}_n)$ and $R_{2n}(\hat{u}^*_{n,b}, \hat{\theta}_n)$, the bootstrap analogue of $R_{1n}(u_n, \theta_0)$ and $R_{1n}(u_n, \theta_0)$ given above. Thus, the bootstrap estimates of $\mathrm{E}(R_{jn})$ are

$$\widehat{\mathrm{E}}(R_{jn}) = \mathrm{E}^*[R_{jn}(\hat{u}^*_{n,b}, \hat{\theta}_n)] \doteq \tfrac{1}{B}\textstyle\sum_{b=1}^{B} R_{jn}(\hat{u}^*_{n,b}, \hat{\theta}_n), \; j = 1, 2,$$

for a large $B$, where $\mathrm{E}^*$ denotes expectation with respect to the EDF $\hat{\mathcal{F}}_n$. In this case, the error of using $\widehat{\mathrm{E}}(R_{jn})$ to estimate $\mathrm{E}(R_{jn})$ comes from the estimations of both $\theta_0$ and $\mathcal{F}_0$.

The bootstrap estimates $\widehat{\mathrm{E}}(R_{jn})$ are seen to have identical structures as the Monte Carlo estimates $\overline{\mathrm{E}}(R_{jn})$ assuming a known $\mathcal{F}_0$, the simple bootstrap estimates $\widetilde{\mathrm{E}}(R_{jn})$ assuming a known $\theta_0$, and the original estimand $\mathrm{E}(R_{jn})$ assuming both $\theta_0$ and $\mathcal{F}_0$ known. These arguments suggest that if we write $\mathrm{E}(R_{jn}) \equiv \mathcal{R}_{jn}(\theta_0, \mu_0)$, then we have $\mathrm{E}^*[R_{jn}(\hat{u}^*_{n,b}, \hat{\theta}_n)] = \mathcal{R}_{jn}(\hat{\theta}_n, \hat{\mu}_n)$. On the other hand, if the closed form expressions for $\mathcal{R}_{jn}(\theta_0, \mu_0)$ exist, then their analytical or plug-in estimates are $\mathcal{R}_{jn}(\hat{\theta}_n, \hat{\mu}_n)$, the same as the bootstrap estimates. From these, we conclude that *the validity of the proposed bootstrap method follows that of the analytical method*. However, the bootstrap method avoids the analytical derivations of the expectations and the direct estimations of $\mu_0$, which, in particular the former, greatly extends the applicability of the methods for bias and variance corrections.

The above arguments extend directly to the more general functions $b_{-s/2}(\theta_0, \mu_0)$, $s = 2, 3$, and $v_{-s/2}(\theta_0, \mu_0)$, $s = 2, 3, 4$, in that their bootstrap estimates can be written as $\widehat{b}_{-s/2} = b_{-s/2}(\hat{\theta}_n, \hat{\mu}_n)$ and $\widehat{v}_{-s/2} = v_{-s/2}(\hat{\theta}_n, \hat{\mu}_n)$. Evidently, both the $b$ and $v$ functions satisfy the general smoothness conditions set out in Assumptions E and F in Section 2.3. The validity of the bootstrap method for bias correction is established as follows.

**Corollary 3.5:** *Assume the assumptions of Corollary 3.1 hold. For the bias-corrected estimators defined in* (26) *with bias corrections estimated using the proposed bootstrap method, we have:* $\mathrm{Bias}(\hat{\lambda}^{\mathrm{bc2}}_n) = O((\frac{h_n}{n})^{\frac{3}{2}})$ *and* $\mathrm{Bias}(\hat{\lambda}^{\mathrm{bc3}}_n) = O((\frac{h_n}{n})^2)$.

Note that the exact content of $\mu_0$ is not important as $\hat{\mu}_n$ is not required in the bootstrap process. The proof of the results, however, requires that $\hat{\mu}_n$, or in general $\hat{\mathcal{F}}_n$, possess similar properties as $\hat{\theta}_n$, i.e., being $\sqrt{n/h_h}$-consistent and having a bias of order $O(\frac{h_n}{n})$. These properties are shown in Lemma B.6 for $\hat{\mathcal{F}}_n$, and in detail for the case of third moment. The cases of higher order moments can be shown in a similar manner.

The validity of the bootstrap method for third-order variance correction follows closely the arguments used in the proof of Corollary 3.5, except that the estimation of the first-order term $v_{-1}$ needs an additional attention for achieving a third-order variance correction.

Similar to the general case of Section 2, one shows for the SAR model $v_{-1} = \nu_{-1} + \nu_{-3/2} + \nu_{-2} + O((\frac{h_n}{n})^{\frac{5}{2}})$, and by (A-9) the first-order variance of $\hat{\lambda}_n$ based on JEE takes the form:

$$\nu_{-1} = \tau_n^{-1} + \tau_n^{-2}[2\mu_{0,3}\, g_n^{\circ\prime} M_n \eta_n + (\mu_{0,4} - 3) g_n^{\circ\prime} g_n^{\circ}], \tag{33}$$

where $\tau_n = \eta_n' M_n \eta_n + \text{tr}[G_n^{\circ}(G_n^{\circ} + G_n^{\circ\prime})]$, $G_n^{\circ} = G_n - \frac{1}{n}\text{tr}(G_n) I_n$, $g_n^{\circ} = \text{diag}(G_n^{\circ})$ (the vector of diagonal elements of $G_n^{\circ}$), and $\mu_{0,r}$ is the $r$th moment of $u_{ni}$, $r = 3$ and 4. Clearly, $\nu_{-1}$ is an explicit and smooth function of $\vartheta_0$, and it depends on $\mu_0 = (\mu_{0,3}, \mu_{0,4})'$ linearly.

With $\tilde{b}_{-1,\theta_0}$ defined in Corollary 2.6 and $\widehat{v}_{-1}^{\text{bc2}}$ defined by (16) with $n$ replaced by $n/h_n$, the validity of the third-order variance correction is established as follows.

**Corollary 3.6:** *Assume the assumptions of Corollary 3.4 hold. The variance expansion (29) becomes* $\text{Var}(\hat{\lambda}_n^{\text{bc3}}) = v_{-1} + v_{-3/2} + v_{-2} - 2\text{E}(\tilde{b}'_{-1,\theta_0})\text{ACov}(\hat{\lambda}_n, \hat{\theta}_n) + O((\frac{h_n}{n})^{\frac{5}{2}})$*, and further* $\text{E}(\widehat{v}_{-1}^{\text{bc2}} + \widehat{v}_{-3/2} + \widehat{v}_{-2}) = v_{-1} + v_{-3/2} + v_{-2} + O((\frac{h_n}{n})^{\frac{5}{2}})$.

The implication of the result in Corollary 3.6 is that the bootstrap estimate of $\text{Var}(\hat{\lambda}_n^{\text{bc3}})$ with $\widehat{v}_{-1}^{\text{bc2}}$ is valid in that it does not induce additional bias larger than $O((\frac{h_n}{n})^{\frac{5}{2}})$, because the other term $\text{E}(\tilde{b}'_{-1,\theta_0})\text{ACov}(\hat{\lambda}_n, \hat{\theta}_n)$ is of order $O((\frac{h_n}{n})^2)$ and the error from the estimation of it will be $O((\frac{h_n}{n})^{\frac{5}{2}})$. $\text{ACov}(\hat{\lambda}_n, \hat{\theta}_n)$ can be estimated by the plug-in method as an explicit expression of $\text{AVar}(\hat{\theta}_n)$ is available (Lee, 2004a), and $\text{E}(\tilde{b}_{-1,\theta_0})$ can be replaced by $\tilde{b}_{-1,\theta_0}$ which can be estimated numerically (see Sec 3.4 for details). The full expression for $\widehat{v}_{-1}^{\text{bc2}}$ is given in (B-9) and (B-10), and the details for the practical implementation of $\widehat{v}_{-1}^{\text{bc2}}$ are given in Appendix B, after the proof of Corollary 3.6. Some useful remarks are given below.

The $\widehat{v}_{-1}^{\text{bc2}}$ can be simplified greatly when either $h_n$ is unbounded, under which $\frac{h_n}{n} g_n^{\circ\prime} g_n^{\circ} = O(h_n^{-1}) = o(1)$ and $\frac{h_n}{n} g_n^{\circ\prime}\eta_n \leq O(h_n^{-\frac{1}{2}}) = o(1)$, or the variability in the number of neighbors for each unit gets small as $n$ goes large.[8] In these cases, $\nu_{-1}$ essentially equals $\tau_n^{-1}$, and

$$\widehat{v}_{-1}^{\text{bc2}} = \widehat{v}_{-1} - \widehat{\nu}_{-1,\theta_0}\widehat{\mathbf{b}}_{-1} - \tfrac{1}{2}\text{tr}[\widehat{\nu}_{-1,\theta_0\theta_0'}\widehat{\text{AVar}}(\hat{\theta}_n)], \tag{34}$$

where $\nu_{-1,\theta_0} = -\tau_n^{-2}\dot{\tau}_n$ and $\nu_{-1,\theta_0\theta_0'} = 2\tau_n^{-3}\dot{\tau}_n\dot{\tau}_n' - \tau_n^{-2}\ddot{\tau}_n$ with $\dot{\tau}_n$ and $\ddot{\tau}_n$ being the gradient and Hessian of $\tau_n$; $\widehat{\mathbf{b}}_{-1} = \{\widehat{b}_{-1}, (\hat{\beta}_n - \hat{\beta}_n^{\text{bc2}})', \hat{\sigma}_n^2 - \hat{\sigma}_n^{2,\text{bc2}}\}'$ with $\hat{\beta}_n^{\text{bc2}}$ and $\hat{\sigma}_n^{2,\text{bc2}}$ being the second-order bias-corrected estimators of $\beta_0$ and $\sigma_0^2$ obtained from (31) and (32) using $\hat{\lambda}_n^{\text{bc2}}$. Monte Carlo results show that $\widehat{v}_{-1}^{\text{bc2}}$ defined in (34) works very well in that the additional terms in the full expression of $\widehat{v}_{-1}^{\text{bc2}}$ are negligible unless $n$ is small and errors are skewed.

### 3.4 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 $\hat{\lambda}_n^{\text{bc2}}$ and $\hat{\lambda}_n^{\text{bc3}}$ of the spatial

[8] This is seen from ($i$) $\frac{1}{n} g_n^{\circ\prime} g_n^{\circ}$ is the sample variance of the elements of $\text{diag}(G_n)$, which is 0 when $\lambda_0 = 0$; ($ii$) $G_n = W_n + \lambda_0 W_n^2 + \lambda_0^2 W_n^3 + \ldots$, if $|\lambda_0| < 1$ and $w_{n,ij} < 1$; and ($iii$) the diagonal elements of $W_n^r, r \geq 2$ inversely relate to the number of neighbors for each unit, say $\{k_i\}_{i=1}^n$, see Anselin (2003). In fact, when $W_n$ is row-normalized and symmetric, $\text{diag}(W_n^2) = \{k_i^{-1}\}$. That $\text{Var}(k_i) = o(1)$ can be seen to be true for many popular spatial layouts such as Rook, Queen, group interactions, etc, see Yang (2010b). Note $\tau_n = O(\frac{n}{h_n})$.

lag parameter $\lambda$, $(ii)$ the finite sample performance of the corrected standard errors (se), $(iii)$ the impact of the bias and se corrections on the subsequent inferences for $\lambda$, and $(iv)$ the impact of bias corrections for $\hat{\lambda}_n$ on the estimators of $\beta$ and $\sigma^2$.

Monte Carlo experiments are carried out based on the following SAR model:

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

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 $\{.5, .1, .1\}$, $\sigma$ at 1 or 2, $\lambda$ takes values $\{.5, .25, 0, -.25, -.5\}$, and $n$ takes values $\{50, 100, 200, 500\}$.[9] 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

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

where in `MRSAR-B`, $\{z_r, z_{ir}, v_r, v_{ir}\} \overset{iid}{\sim} N(0,1)$, 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 details for generating the $W_n$ matrix under rook contiguity are as follows: $(i)$ index the $n$ spatial units by $\{1, 2, \cdots, n\}$, 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 $W_{ij} = 0, i, j = 1, \cdots, n$, 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.

To generate the $W_n$ matrix according to the group interaction scheme, suppose we have $k$ groups of sizes $m_1, m_2, \cdots, m_k$. Define $W_n = \text{diag}\{W_j/(m_j - 1), j = 1, \cdots, k\}$, 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

[9] As in Lee (2007a), the maximization of $\ell_n^c(\lambda)$ is performed globally without imposing a restricted lower bound on $\lambda$. This is important when the true $\lambda$ 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 $\lambda$ in relation to the eigenvalues of $W_n$.

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 $(m_1, m_2, \cdots, m_k)$ according to a discrete distribution centered at $m$, and $(iii)$ adjust the group sizes so that $\sum_{j=1}^{k} m_j = n$.[10]

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 $u_n$, three distributions are considered: `dgp1:` the elements $\{u_i\}$ of $u_n$ are iid standard normal, `dgp2:` $\{u_i\}$ are iid standardized normal mixture, and `dgp3:` $\{u_i\}$ are iid standardized log-normal. Specifically, for `dgp2`,

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

where $\{\xi_i\} \overset{iid}{\sim} \text{Bernoulli}(\pi)$, and $\{Z_i\} \overset{iid}{\sim} N(0,1)$ independent of $\{\xi_i\}$. The parameter $\pi$ 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 $\tau$. We choose $\tau = 4$ to simulate the situation where there are gross errors in the data. For `dgp3`,

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

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 $(n, \lambda, \sigma)$, the error distributions, and the spatial layouts. We also report the averages (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 (28)-(29) and calculated based on the proposed bootstrap method.[11]

[10]Clearly, 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.

[11]In connection to Corollary 3.6 and discussions thereafter, the partial derivative $\tilde{b}_{-1,\theta_0}$ needed in $\widehat{V}_3(\hat{\lambda}_n^{\text{bc3}})^{\frac{1}{2}}$ is estimated by the bootstrap numerical derivative: $[(\widehat{b}_{-1} \text{ at } \hat{\theta}_n + \varepsilon\iota_i) - (\widehat{b}_{-1} \text{ at } \hat{\theta}_n)]/\varepsilon$, where $\iota_i$ is a $(p+2) \times 1$ vector with 1 on its $i$th position and zero elsewhere, $i = 1, \cdots, p+2$, and $\varepsilon$ is taken to be 0.0001 in our experiments. The $v_{-1}$ in $\widehat{V}_3(\hat{\lambda}_n^{\text{bc3}})^{\frac{1}{2}}$ is estimated by $\widehat{v}_{-1}^{\text{bc2}}$ with details given in and after (B-9).

Each set of results is based on 10,000 Monte Carlo samples, and $999+floor(n^{0.75})$ 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 $\hat{\lambda}_n^{\rm bc2}$ and $\hat{\lambda}_n^{\rm bc3}$ are in general nearly unbiased and clearly outperform the original QMLE $\hat{\lambda}_n$;

(ii) $\hat{\lambda}_n$ is always downward biased and the bias can be very serious depending on the spatial layout, the sample size and the error standard deviation;

(iii) $\hat{\lambda}_n^{\rm bc3}$ improves over $\hat{\lambda}_n^{\rm bc2}$, but using $\hat{\lambda}_n^{\rm bc2}$ seems to be sufficient under most of the situations as far as bias-correction is concerned;[12]

(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 $\sigma$ and the slope parameters also have a big impact – the bigger the $\sigma$ 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 $\lambda$ and the way the regressors being generated affect the finite sample performance of $\hat{\lambda}_n$ – as $\lambda$ 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 $\lambda$ 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 $\hat{\lambda}_n^{\rm bc3}$ when sample size is small, suggesting that the variances of $\hat{\lambda}_n$ and $\hat{\lambda}_n^{\rm bc3}$ may differ on higher-order term (see panels (a)-(c), Table 1). The results in the last four columns of Table 1 show that $\widehat{V}_3(\hat{\lambda}_n^{\rm bc3})$ provides the best approximation to the variance of $\hat{\lambda}_n^{\rm bc3}$. The empirical sds of $\hat{\lambda}_n^{\rm bc2}$ and $\hat{\lambda}_n^{\rm bc3}$ agree closely, suggesting that the finite sample variances of $\hat{\lambda}_n^{\rm bc2}$ and $\hat{\lambda}_n^{\rm bc3}$ 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 (30) 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 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

[12] In many cases, such as AR and MA models, $b_{-3/2}$ is in fact zero as a referee points out. This might help explain the small difference between $\hat{\lambda}_n^{\rm bc2}$ and $\hat{\lambda}_n^{\rm bc3}$ for the SAR model we consider.

gives further improvements; and (*iv*) use of the third-order bias-corrected estimator and its third-order variance estimate gives the best results.

**The performance of $\hat{\beta}_n(\hat{\lambda}_n^{\rm bc2})$ and $\frac{n}{n-p}\hat{\sigma}_n^2(\hat{\lambda}_n^{\rm bc2})$.** Monte Carlo experiments are also carried out to investigate the performance of QMLEs of $\beta$ and $\sigma^2$ following a second-order bias corrected $\hat{\lambda}_n$. Monte Carlo results (not reported for brevity) show that the biases of $\hat{\beta}_n(\hat{\lambda}_n^{\rm bc2})$ and $\frac{n}{n-p}\hat{\sigma}_n^2(\hat{\lambda}_n^{\rm bc2})$ are very small, and can be significantly smaller than those of $\hat{\beta}_n(\hat{\lambda}_n)$ and $\frac{n}{n-p}\hat{\sigma}_n^2(\hat{\lambda}_n)$. Thus, further bias corrections as discussed around (31) and (32) may not be necessary. More detailed study is interesting but beyond the scope of this paper.

# 4 Extension to a Vector of Nonlinear Estimators

In previous sections, we have focused on the models of only one nonlinear parameter. Econometric or spatial econometric models may well contain two or more nonlinear parameters. An immediate example is the so-called SARAR model, which extends the SAR model given in (1) by allowing the disturbance $\varepsilon_n$ to follow a SAR process. Other examples include the spatial heteroskedastic model and the Box-Cox heteroskedastic model. We will discuss these three examples in some detail after the extension of the method.

When there are two or more nonlinear parameters in the model that are the main source of bias in model estimation and the main cause of difficulty in bias-correction, our method can be extended in a straightforward manner. Use of CEE still greatly reduces the dimensionality, and more importantly it captures the additional second-order bias inherited from the estimation of linear and scale parameters.[13] Furthermore, as far as bootstrap method is concerned, there is little difficulty in extending a scalar nonlinear estimator problem to a vector nonlinear estimator problem. The details are as follows.

Using the same notation, under the same set-up (except that $\tilde{\psi}_n$ is now a vector), and with almost identical conditions as in Section 2 (except that $|\cdot|$ in $|H_{rn}(\lambda) - H_{rn}(\lambda_0)|$ and $|\lambda - \lambda_0|$ of Assumption D is replaced by the matrix norm $\|\cdot\|$), we have the third-order stochastic expansion for the vector $\hat{\lambda}_n$ defined by the CEE, $\hat{\lambda}_n = \arg\{\tilde{\psi}_n(\lambda) = 0\}$: $\hat{\lambda}_n - \lambda_0 = a_{-1/2} + a_{-1} + a_{-3/2} + O_p(n^{-2})$, where

$$\begin{aligned} a_{-1/2} &= \Omega_n \tilde{\psi}_n, \quad\quad a_{-1} = \Omega_n H_{1n}^{\circ} a_{-1/2} + \tfrac{1}{2}\Omega_n \mathrm{E}(H_{2n})(a_{-1/2} \otimes a_{-1/2}), \\ a_{-3/2} &= \Omega_n H_{1n}^{\circ} a_{-1} + \tfrac{1}{2}\Omega_n H_{2n}^{\circ}(a_{-1/2} \otimes a_{-1/2}) + \tfrac{1}{2}\Omega_n \mathrm{E}(H_{2n})(a_{-1/2} \otimes a_{-1} \\ &\quad + a_{-1} \otimes a_{-1/2}) + \tfrac{1}{6}\Omega_n \mathrm{E}(H_{3n})(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}), \end{aligned}$$

and $\otimes$ denotes the Kronecker product. This CEE-based expansion takes the same form as that based on JEE of Bao and Ullah (2007a), except that when taking the expectations for

[13] However, if one is interested in the analytical properties of the bias corrections, e.g., the main cause of the bias problem and the factors affecting the magnitude of the bias, it is easier to work with JEE.

deriving the biases, the terms involving $\mathrm{E}(\tilde{\psi}_n)$ do not vanish. The bias terms of $\hat{\lambda}_n$ are:

$$
\begin{aligned}
b_{-1} &= 2\Omega_n \mathrm{E}(\tilde{\psi}_n) + \Omega_n \mathrm{E}(H_{1n}\Omega_n\tilde{\psi}_n) + \tfrac{1}{2}\Omega_n \mathrm{E}(H_{2n})\mathrm{E}[(\Omega_n\tilde{\psi}_n)\otimes(\Omega_n\tilde{\psi}_n)], \text{ and} \\
b_{-3/2} &= \Omega_n \mathrm{E}(H_{1n}^{\circ}a_{-1}) + \tfrac{1}{2}\Omega_n \mathrm{E}[H_{2n}^{\circ}(a_{-1/2}\otimes a_{-1/2})] + \tfrac{1}{2}\Omega_n \mathrm{E}(H_{2n})\mathrm{E}(a_{-1/2}\otimes a_{-1} \\
&\quad + a_{-1}\otimes a_{-1/2}) + \tfrac{1}{6}\Omega_n \mathrm{E}(H_{3n})\mathrm{E}(a_{-1/2}\otimes a_{-1/2}\otimes a_{-1/2}).
\end{aligned}
$$

We have the options of performing second-order bias correction on $\hat{\lambda}_n$: $\hat{\lambda}_n^{\mathrm{bc2}} = \hat{\lambda}_n - \widehat{b}_{-1}$, and the third-order bias correction: $\hat{\lambda}_n^{\mathrm{bc3}} = \hat{\lambda}_n - \widehat{b}_{-1} - \widehat{b}_{-3/2}$, where $\widehat{b}_{-1}$ and $\widehat{b}_{-3/2}$ are the bootstrap estimates. Note that the nonstochastic matrices $\Omega_n$ and $\mathrm{E}(H_{2n})$ appear between the random quantities $H_{1n}$ and $\tilde{\psi}_n$ inside the expectation sign. This makes the evaluation of the expectations more difficult. Using the well-known properties of Kronecker product $A \otimes B)(C \otimes D) = AC \otimes BD$ and $\mathrm{vec}(ACB) = (B' \otimes A)\mathrm{vec}(C)$, where 'vec' vectorizes a matrix by stacking its columns, one can 'pull' these nonstochastic matrices outside the expectation sign (see, e.g., Horn and Johnson, 1985). For example, $b_{-1}$ can be expressed as

$$
b_{-1} = 2\Omega_n \mathrm{E}(\tilde{\psi}_n) + \Omega_n \mathrm{E}(\tilde{\psi}_n' \otimes H_{1n})\mathrm{vec}(\Omega_n) + \tfrac{1}{2}\Omega_n \mathrm{E}(H_{2n})(\Omega_n \otimes \Omega_n)\mathrm{E}(\tilde{\psi}_n \otimes \tilde{\psi}_n),
$$

which simplifies the estimation (plug-in and bootstrap) of various expectations. While the same can be done for $b_{-3/2}$, the final expression becomes lengthy. To avoid this complication, we revise the bootstrap procedure described in Section 2.2 as follows:

1. *Draw $B$ independent random samples, $\{\hat{u}_{n,b}^*,\ b = 1, 2, \ldots, B\}$, from $\hat{\mathcal{F}}_n$,*
2. *Calculate the bootstrap estimates of $E(H_{1n})$ and $E(H_{2n})$,*

$$
\widehat{\mathrm{E}}(H_{1n}) = \tfrac{1}{B}\textstyle\sum_{b=1}^{n} H_{1n}(\hat{u}_{n,b}^*, \hat{\theta}_n) \text{ and } \widehat{\mathrm{E}}(H_{2n}) = \tfrac{1}{B}\textstyle\sum_{b=1}^{n} H_{2n}(\hat{u}_{n,b}^*, \hat{\theta}_n)
$$

3. *Based on the bootstrap estimates $\hat{\Omega}_n = -\widehat{\mathrm{E}}^{-1}(H_{1n})$ and $\widehat{\mathrm{E}}(H_{2n})$, calculate the bootstrap estimate of, e.g., $\mathrm{E}[H_{2n}^{\circ}(a_{-1/2}\otimes a_{-1/2})]$, as*

$$
\frac{1}{B}\sum_{b=1}^{n}\Big\{[H_{2n}(\hat{u}_{n,b}^*, \hat{\theta}_n) - \widehat{\mathrm{E}}(H_{2n})][\hat{\Omega}_n\tilde{\psi}_n(\hat{u}_{n,b}^*, \hat{\theta}_n) \otimes \hat{\Omega}_n\tilde{\psi}_n(\hat{u}_{n,b}^*, \hat{\theta}_n)]\Big\}.
$$

All the other expectations in $b_{-3/2}$ can be handled in the same manner. The computational cost is still low as the bootstrap replications do not involve reestimation of model parameters. The validity of this procedure can be inferred from that in Section 2.2. The results on higher-order variance-covariance matrix of $\hat{\lambda}_n$ can be handled in a similar manner. We outline a few applications, concentrating on bias correction, to demonstrate the extended method. Detailed treatments are beyond the scope of the paper.

**The SARAR Model.** Consider an extension of Model (1) by allowing the disturbance to follow a SAR process (Anselin, 1988): $Y_n = \lambda W_{1n} Y_n + X_n\beta + \varepsilon_n,\ \ \varepsilon_n = \rho W_{2n}\varepsilon_n + \sigma u_n$.

Both spatial parameters enter the model in a nonlinear fashion and thus their estimation is expected to incur bias. Letting $A_n(\lambda) = I_n - \lambda W_{1n}$ and $B_n(\rho) = I_n - \rho W_{2n}$, the concentrated quasi-Gaussian loglikelihood for $(\lambda, \rho)$, assuming $u_n \sim N(0, I_n)$, is

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

where $\hat{\sigma}_n^2(\lambda, \rho) = \frac{1}{n}Y_n'(\lambda)M_n(\rho)Y_n(\lambda)$, $Y_n(\lambda) = A_n(\lambda)Y_n$, and $M_n(\rho)$ is an $n \times n$ matrix involving only the $\rho$ parameter. The concentrated estimating function is

$$\tilde{\psi}(\lambda, \rho) = \begin{cases} -T_{0n}(\lambda) + \dfrac{Y_n'(\lambda)M_n(\rho)W_{1n}Y_n}{Y_n'(\lambda)M_n(\rho)Y_n(\lambda)}, \\ -K_{0n}(\rho) - \dfrac{Y_n'(\lambda)\dot{M}_n(\rho)Y_n(\lambda)}{2Y_n'(\lambda)M_n(\rho)Y_n(\lambda)}, \end{cases}$$

where $T_{0n}(\lambda) = \frac{1}{n}\text{tr}(W_{1n}A_n^{-1}(\lambda))$, $K_{0n}(\rho) = \frac{1}{n}\text{tr}(W_{2n}B_n^{-1}(\rho))$, and $\dot{M}_n(\rho) = \frac{d}{d\rho}M_n(\rho)$. Straightforward but tedious algebra leads to the expressions for the $2 \times 2$ matrix $H_{1n}(\lambda, \rho)$, the $2 \times 4$ matrix $H_{2n}(\lambda, \rho)$, and the $2 \times 8$ matrix $H_{3n}(\lambda, \rho)$. Bias and variance corrections proceed with the above bootstrap method.

**Spatial heteroskedastic regression.** An alternative extension of Model (1) is to let $\varepsilon_n = \sigma\Gamma_n^{\frac{1}{2}}(\gamma)u_n$, where $\Gamma_n(\gamma) = \text{diag}\{\varpi_i(\gamma), i = 1\ldots, n\}$ and $\varpi_i(\gamma)$ are the given skedastic functions with a $q$ vector of unknown heteroskedasticity parameters $\gamma$ (Anselin, 1988). The concentrated quasi-Gaussian loglikelihood for $(\lambda, \gamma)$, assuming $u_n \sim N(0, I_n)$, is

$$\ell_n^c(\lambda, \gamma) = -\frac{n}{2}[\log(2\pi) + 1] - \frac{n}{2}\log[\hat{\sigma}_n^2(\lambda, \gamma)] + \log|A_n(\lambda)| - \sum_{i=1}^{n}\log\varpi_i(\gamma),$$

where $\hat{\sigma}_n^2(\lambda, \gamma) = \frac{1}{n}Y_n'(\lambda)M_n(\gamma)Y_n(\lambda)$, $Y_n(\lambda) = A_n(\lambda)Y_n$, and $M_n(\gamma)$ is a projection matrix involving only the $\gamma$ parameters. The concentrated estimating function is

$$\tilde{\psi}(\lambda, \gamma) = \begin{cases} -T_{0n}(\lambda) + \dfrac{Y_n'(\lambda)M_n(\gamma)W_{1n}Y_n}{Y_n'(\lambda)M_n(\gamma)Y_n(\lambda)}, \\ -\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\dfrac{\varpi_{i\gamma_j}(\gamma)}{\varpi_i(\gamma)} - \dfrac{Y_n'(\lambda)\dot{M}_{nj}(\gamma)Y_n(\lambda)}{2Y_n'(\lambda)M_n(\gamma)Y_n(\lambda)}, \quad j = 1, \ldots, q, \end{cases}$$

where $\varpi_{i\gamma_j}(\gamma) = \frac{\partial}{\partial\gamma_j}\varpi_i(\gamma)$, and $\dot{M}_{nj}(\gamma) = \frac{\partial}{\partial\gamma_j}M_n(\gamma)$, $j = 1, \ldots, q$. Straightforward but more tedious algebra leads to the expressions for the matrices $H_{jn}(\lambda, \gamma), j = 1, 2, 3$. Bias and variance corrections proceed with the above bootstrap method.

**Box-Cox heteroskedastic regression.** Consider a non-spatial model: $h(Y_n, \lambda) = X_n\beta + \sigma\Gamma_n^{\frac{1}{2}}(\gamma)u_n$, where the response $Y_n$ is transformed by the Box-Cox transformation $h(y, \lambda) = \frac{1}{\lambda}(y^\lambda - 1), \lambda \neq 0; \log y, \lambda = 0$, $\lambda$ is an unknown transformation parameter, and the disturbances are heteroskedastic defined as in the spatial heteroskedastic model considered above. See, e.g., Yang and Tse (2006, 2007) for some interesting applications of this model. The concentrated quasi-Gaussian loglikelihood for $(\lambda, \gamma)$, assuming normality, is

$$\ell_n^c(\lambda,\gamma) = -\frac{n}{2}[\log(2\pi)+1] - \frac{n}{2}\log[\hat{\sigma}_n^2(\lambda,\gamma)] + \sum_{i=1}^n \log h_Y(Y_i,\lambda) - \sum_{i=1}^n \log \varpi_i(\gamma),$$

where $\hat{\sigma}_n^2(\lambda,\gamma) = \frac{1}{n}h(Y_n,\lambda)'M_n(\gamma)h(Y_n,\lambda)$, $h_Y(Y_i,\lambda) = \frac{\partial}{\partial Y_i}h(Y_i,\lambda)$ and $M_n(\gamma)$ is as in the spatial heteroskedastic model considered above. The concentrated estimating function is

$$\tilde{\psi}(\lambda,\gamma) = \begin{cases} \frac{1}{n}\sum_{i=1}^n \frac{h_{Y\lambda}(Y_i,\lambda)}{h_Y(Y_i,\lambda)} - \frac{h'_\lambda(Y_n,\lambda)M_n(\gamma)h(Y_n,\lambda)}{h'(Y_n,\lambda)M_n(\gamma)h(Y_n,\lambda)}, \\ -\frac{1}{n}\sum_{i=1}^n \frac{\varpi_{i\gamma_j}(\gamma)}{\varpi_i(\gamma)} - \frac{h'(Y_n,\lambda)\dot{M}_{nj}(\gamma)h(Y_n,\lambda)}{2h'(Y_n,\lambda)M_n(\gamma)h(Y_n,\lambda)}, \quad j=1,\ldots,q, \end{cases}$$

where $h_\lambda(Y_n,\lambda) = \frac{\partial}{\partial\lambda}h(Y_n,\lambda)$ and $h_{Y\lambda}(Y_i,\lambda) = \frac{\partial}{\partial\lambda}h_Y(Y_i,\lambda)$. The expressions for $H_{jn}(\lambda,\gamma)$, $j = 1, 2, 3$, can be derived in a straightforward manner though tedious, but their expectations do not have closed form expressions, even working with the joint likelihood function, as the expectations of the form, $\mathrm{E}[h_{Y\lambda}(Y_i,\lambda_0)]$, $\mathrm{E}\big(\frac{h_{Y\lambda}(Y_i,\lambda)}{h_Y(Y_i,\lambda)}\big)$, etc., do not have closed form expressions. However, the proposed bootstrap method does not require the closed form expressions for these expectations, which stands in contrast to the analytical method.

Each of the three cases presented above merits a detailed study theoretically and empirically, but are quite involved and thus are clearly beyond the scope of the current paper. We will pursue these studies in future research.[14]

## 5 Conclusions and Discussions

To address the bias 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 often 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 full details 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.

[14]There is a general issue on the existence of $\mathrm{E}(a_{-s/2})$ in conducting bootstrap bias corrections, which boils down to the existence of the moments of ratios of two stochastic quantities. If the denominator of a random ratio is bounded away from zero with probability one (indeed it is, as it typically corresponds to the estimator of the error variance), the existence of the moments of the random ratio boils down to the existence of the moments of the numerator. Thus, moment requirements for bias correction based on a CEE are essentially equivalent to the moment requirements for bias correction based on a JEE.

We have emphasized through out the paper the two 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 models. 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 (Lee and Yu, 2010): our results can be easily extended over, but the analytical results of Bao (2013) 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.

An important extension of the proposed method is made to the models involving two or more nonlinear parameters that are the main source of bias in model estimation and the main cause of difficulty in bias-correction. The proposed bootstrap method shows little difficulty in such an extension. Further, besides the dimensionality reduction, it is emphasized that use of CEE captures the additional bias inherited from the estimation of linear and scale parameters, which is typically of second order. This latter point is very interesting and deserves more attention in line with the models discussed in the early section. However, these are beyond the scope of the paper and will be dealt with in future research.

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, as a future work, 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.

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 bias of the QMLE of $\lambda$, 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 (2013). As a rigorous study on either issue can be quite involved, we plan to pursue these two issues in future research. Finally, there is an issue on bias-correcting $\hat{v}_{-1}$ when $\nu_{-1}$ does not possess a closed form expression. An alternative would be the *jackknife-bootstrap* estimate of the bias of $\hat{v}_{-1}$: $(n-1)(\overline{\hat{v}}_{-1} - \hat{v}_{-1})$, where $\overline{\hat{v}}_{-1} = \frac{1}{n}\sum_{i=1}^{n} \hat{v}_{-1,i}$, and $\hat{v}_{-1,i}$ is obtained in the same way as $\hat{v}_{-1}$ except that the $i$th residual is deleted. This leads to a jackknife-bootstrap bias-corrected estimator of $v_{-1}$: $\hat{v}_{-1}^{\text{jbbc}} = n\hat{v}_{-1} - (n-1)\overline{\hat{v}}_{-1}$. This approach is simple but computationally demanding. Clearly, a rigorous study on its properties renders a separate 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 $\lambda_0$ 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,

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

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

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

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

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

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

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

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 $\hat{\lambda}_n$,

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

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 $\Omega_n + O_p(n^{-\frac{1}{2}})$, we obtain,

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

where $a_{-1} = \Omega_n H_{1n}^{\circ} a_{-1/2} + \frac{1}{2}\Omega_n \mathrm{E}(H_{2n})(a_{-1/2}^2)$. 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 $\Omega_n + \Omega_n^2 H_{1n}^{\circ} + \Omega_n^3 H_{1n}^{\circ 2} + O_p(n^{-\frac{3}{2}})$, $\Omega_n + \Omega_n^2 H_{1n}^{\circ} + O_p(n^{-1})$, and $\Omega_n + O_p(n^{-1/2})$, we obtain a third-order stochastic expansion for $\hat{\lambda}_n$,

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

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

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

**Proof of Corollary 2.2:** We have $b_{-1} = \mathrm{E}(a_{-1/2}) + \mathrm{E}(a_{-1})$, and $b_{-3/2} = \mathrm{E}(a_{-3/2})$. The result follows as the 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 $\mathrm{Var}(\hat{\lambda}_n^{\mathrm{bc3}}) = \mathrm{Var}(\hat{\lambda}_n - \widehat{b}_{-1} - \widehat{b}_{-3/2}) = \mathrm{Var}(\hat{\lambda}_n) - 2\mathrm{Cov}(\hat{\lambda}_n, \widehat{b}_{-1}) + O(n^{-\frac{5}{2}}) = v_{-1} + v_{-3/2} + v_{-2} - 2\mathrm{ACov}(\hat{\lambda}_n, \widehat{b}_{-1}) + O(n^{-\frac{5}{2}})$ as the other terms can all be merged into $O(n^{-\frac{5}{2}})$.

**Proof of Corollary 2.5:** Recall $b_{-1,\vartheta} = \frac{\partial}{\partial\vartheta} b_{-1}(\vartheta)$. The mean value theorem gives

$$b_{-1}(\hat{\vartheta}_n) = b_{-1}(\vartheta_0) + b'_{-1,\vartheta_0}(\hat{\vartheta}_n - \vartheta_0) + (b'_{-1,\bar{\vartheta}_n} - b'_{-1,\vartheta_0})(\hat{\vartheta}_n - \vartheta_0), \tag{A-7}$$

where $\bar{\vartheta}_n$ lies between $\hat{\vartheta}_n$ and $\vartheta_0$. Under Assumption E, $\mathrm{E}(\hat{\vartheta}_n - \vartheta_0) = O(n^{-1})$, $b_{-1,\vartheta_0} = O(n^{-1})$, $b_{-1,\bar{\vartheta}_n} - b_{-1,\vartheta_0} = O_p(n^{-\frac{3}{2}})$, and $\hat{\vartheta}_n - \vartheta_0 = O_p(n^{-\frac{1}{2}})$. Taking expectation on both sides of (A-7) leads to $\mathrm{E}[b_{-1}(\hat{\vartheta}_n)] = b_{-1}(\vartheta_0) + O(n^{-2})$. The result $\mathrm{E}[b_{-3/2}(\hat{\vartheta}_n)] = b_{-3/2}(\vartheta_0) + O(n^{-2})$ follows from the differentiability of $b_{-3/2}(\vartheta)$ and $\sqrt{n}$-consistency of $\hat{\vartheta}_n$.

**Proof of Corollary 2.6:** Under Assumption E, we have by Taylor series expansions,

$$\begin{aligned} b_{-1}(\hat{\theta}_n, \mu_n(\hat{\theta}_n)) &= b_{-1}(\theta_0, \mu_n) + \tilde{b}'_{-1,\theta_0}(\hat{\theta}_n - \theta_0) + O_p(n^{-2}) \\ &= b_{-1}(\theta_0, \mu_0) + b'_{-1,\mu_0}(\mu_n - \mu_0) + \mathrm{E}(\tilde{b}'_{-1,\theta_0})(\hat{\theta}_n - \theta_0) + O_p(n^{-2}) \end{aligned} \tag{A-8}$$

noticing that $\tilde{b}_{-1,\theta_0}$ depends on $\mu_n$ and $\mu_n - \mu_0 = O_p(n^{-\frac{1}{2}})$. It follows that $\mathrm{ACov}(\hat{\lambda}_n, \widehat{b}_{-1}) = b'_{-1,\mu_0}\mathrm{ACov}(\hat{\lambda}_n, \mu_n) + \mathrm{E}(\tilde{b}'_{-1,\theta_0})\mathrm{ACov}(\hat{\lambda}_n, \hat{\theta}_n) + O(n^{-\frac{5}{2}})$. As $\mathrm{ACov}(\hat{\lambda}_n, \mu_n) = \Omega_n \mathrm{Cov}(\tilde{\psi}_n, \mu_n) = O(n^{-\frac{3}{2}})$ by Assumption F($ii$), the first result of Corollary 2.6 thus follows.

The proof of the second result starts from the derivation of (16). First, recall $\psi_n \equiv \psi(\theta_0)$ defined in (5). Let $\mathbf{\Sigma}_n = \mathrm{E}(\psi_n \psi_n')$ and $\mathbb{I}_n = -\mathrm{E}(\frac{\partial}{\partial\theta_0'}\psi_n)$. Partition $\psi_n = (\psi_{1n}, \psi_{2n}')'$, $\mathbf{\Sigma}_n = \{\mathbf{\Sigma}_{ij}\}$, and $\mathbb{I}_n = \{\mathbb{I}_{ij}\}$ according to $\lambda_0$ (or 1) and $\alpha_0$ (or 2). Note that $v_{-1} = \Omega_n^2 \mathrm{Var}(\tilde{\psi}_n)$ and $\Omega_n = -[\mathrm{E}(\frac{d}{d\lambda_0}\tilde{\psi}(\lambda_0))]^{-1}$. A 3rd-order Taylor expansion of $\tilde{\psi}_n = \psi_{1n}(\lambda_0, \hat{\alpha}_n(\lambda_0))$ around $\hat{\alpha}_n(\lambda_0) = \alpha_0$, combined with a 3rd-order stochastic expansion of $\hat{\alpha}_n(\lambda_0) - \alpha_0$ based on $\psi_{2n}(\lambda_0, \hat{\alpha}_n(\lambda_0)) = 0$, gives a 3rd-order stochastic expansion for $\tilde{\psi}_n$ with its first-order term being $\psi_{1n} - \mathbb{I}_{12}\mathbb{I}_{22}^{-1}\psi_{2n}$, which in turn gives a 2nd-order expansion for $-[\mathrm{E}(\frac{d}{d\lambda_0}\tilde{\psi}(\lambda_0))]$ with its first-order term being $\tau_n = \mathbb{I}_{11} - \mathbb{I}_{12}\mathbb{I}_{22}^{-1}\mathbb{I}_{21}$. Combining these two expansions leads to $v_{-1} = \nu_{-1} + \nu_{-3/2} + \nu_{-2} + O(n^{-\frac{5}{2}})$, where the 1st-order term of $\mathrm{Var}(\hat{\lambda}_n)$ based on JEE,

$$\nu_{-1} = \tau_n^{-2}(\mathbf{\Sigma}_{11} - 2\mathbb{I}_{12}\mathbb{I}_{22}^{-1}\mathbf{\Sigma}_{21} + \mathbb{I}_{12}\mathbb{I}_{22}^{-1}\mathbf{\Sigma}_{22}\mathbb{I}_{22}^{-1}\mathbb{I}_{21}), \tag{A-9}$$

and the exact expressions for $\nu_{-3/2}$ and $\nu_{-2}$ are available but not needed as seen below.

Under Assumptions E and F($i$), we have by a Taylor series expansion,

$$v_{-1}(\hat{\vartheta}_n) = v_{-1}(\vartheta_0) + v'_{-1,\vartheta_0}(\hat{\vartheta}_n - \vartheta_0) + \tfrac{1}{2}(\hat{\vartheta}_n - \vartheta_0)' v_{-1,\vartheta_0\vartheta_0'}(\hat{\vartheta}_n - \vartheta_0) + O_p(n^{-\frac{5}{2}}), \tag{A-10}$$

where $v_{-1,\vartheta} = \frac{\partial}{\partial\vartheta} v_{-1}(\vartheta)$ and $v_{-1,\vartheta\vartheta'} = \frac{\partial^2}{\partial\vartheta\partial\vartheta'} v_{-1}(\vartheta)$. By $v_{-1} = \nu_{-1} + \nu_{-3/2} + \nu_{-2} + O(n^{-\frac{5}{2}})$,

$$\begin{aligned} \mathrm{E}[v_{-1}(\hat{\vartheta}_n)] &= v_{-1} + v'_{-1,\vartheta_0}\mathrm{E}(\hat{\vartheta}_n - \vartheta_0) + \tfrac{1}{2}\mathrm{tr}[v_{-1,\vartheta_0\vartheta_0'}\mathrm{AVar}(\hat{\vartheta}_n)] + O(n^{-\frac{5}{2}}) \\ &= v_{-1} + \nu'_{-1,\vartheta_0}\mathrm{E}(\hat{\vartheta}_n - \vartheta_0) + \tfrac{1}{2}\mathrm{tr}[\nu_{-1,\vartheta_0\vartheta_0'}\mathrm{AVar}(\hat{\vartheta}_n)] + O(n^{-\frac{5}{2}}) \\ &\equiv v_{-1} + \nu'_{-1,\theta_0}\mathbf{b}_{-1} + \nu'_{-1,\mu_0}\mathbf{c}_{-1} + \mathbf{d}_{-2} + O(n^{-\frac{5}{2}}). \end{aligned} \tag{A-11}$$

The result of (A-11) immediately suggests a 2nd-order bias-corrected estimator of $\upsilon_{-1}$:

$$\widehat{\upsilon}_{-1}^{\text{bc2}} = \hat{\upsilon}_{-1} - \widehat{\nu}'_{-1,\theta_0}\widehat{\mathbf{b}}_{-1} - \widehat{\nu}'_{-1,\mu_0}\widehat{\mathbf{c}}_{-1} - \widehat{\mathbf{d}}_{-2},$$

as given in (16). Obviously, $\mathbf{b}_{-1}$ is the 2nd-order bias of $\hat{\theta}_n$ and can be obtained from the 2nd-order stochastic expansion based on $\psi_n(\theta)$ defined in (5): $\hat{\theta}_n - \theta_0 = \mathbf{a}_{-1/2} + \mathbf{a}_{-1} + O_p(n^{-\frac{3}{2}})$, where $\mathbf{a}_{-1/2} = \mathbf{\Omega}_n\psi_n$ and $\mathbf{a}_{-1} = \mathbf{\Omega}_n\mathbb{H}_{1n}^{\circ}\mathbf{a}_{-1/2} + \frac{1}{2}\mathbf{\Omega}_n\text{E}(\mathbb{H}_{2n})(\mathbf{a}_{-1/2}\otimes\mathbf{a}_{-1/2})$, $\otimes$ denotes the Kronecker product, $\mathbf{\Omega}_n = -\text{E}(\mathbb{H}_{1n})^{-1}$, $\mathbb{H}_{1n} = \frac{\partial}{\partial\theta_0'}\psi_n(\theta_0)$, $\mathbb{H}_{1n}^{\circ} = \mathbb{H}_{1n} - \text{E}(\mathbb{H}_{1n})$, and $\mathbb{H}_{2n} = \frac{\partial}{\partial\theta_0'}\mathbb{H}_{1n}$. This gives $\text{Bias}(\hat{\theta}_n) = \text{E}(\mathbf{a}_{-1/2} + \mathbf{a}_{-1}) + O(n^{-\frac{3}{2}}) \equiv \mathbf{b}_{-1} + O(n^{-\frac{3}{2}})$ where,

$$\mathbf{b}_{-1} = 2\mathbf{\Omega}_n\text{E}(\psi_n) + \mathbf{\Omega}_n\text{E}(\psi_n'\otimes\mathbb{H}_{1n})\text{vec}(\mathbf{\Omega}_n) + \frac{1}{2}\mathbf{\Omega}_n\text{E}(\mathbb{H}_{2n})(\mathbf{\Omega}_n\otimes\mathbf{\Omega}_n)\text{vec}(\mathbf{\Sigma}_n). \quad \text{(A-12)}$$

Note that, unlike the early works which assume $\text{E}(\psi_n) = 0$, $\text{E}(\psi_n)$ can be $O(n^{-1})$.

Alternatively, $\mathbf{b}_{-1}$ can be defined with its first component being $b_{-1}$ defined in (9), and the second component being the second-order bias of $\hat{\alpha}(\hat{\lambda}_n^{\text{bc2}})$ that can be easily obtained for a given specific model bearing in mind that $\alpha$ is a vector of linear parameters and that $\hat{\alpha}(\lambda_0)$ has an explicit expression, as demonstrated in the subsequent sections.

The quantity $\mathbf{c}_{-1}$ is the 2nd-order bias of $\hat{\mu}_n$ and can be obtained from the expansion of $\hat{\mu}_{n,r} = \frac{1}{n}\sum_{i=1}^n \hat{u}_{ni}^r$, the $r$th moment of the EDF $\hat{\mathcal{F}}_n$ of $\hat{u}_n \equiv u_n(\hat{\theta}_n)$, $r = 3, 4, \ldots$,

$$\begin{aligned}\hat{\mu}_{n,r} &= \mu_{n,r} + \dot{\mu}_{n,r}'(\hat{\theta}_n - \theta_0) + \tfrac{1}{2}(\hat{\theta}_n - \theta_0)'\ddot{\mu}_{n,r}(\hat{\theta}_n - \theta_0) + O_p(n^{-\frac{3}{2}}),\\ &= \mu_{n,r} + \dot{\mu}_{n,r}'(\mathbf{a}_{-1/2} + \mathbf{a}_{-1}) + \tfrac{1}{2}\mathbf{a}_{-1/2}'\ddot{\mu}_{n,r}\mathbf{a}_{-1/2} + O_p(n^{-\frac{3}{2}}), \qquad \text{(A-13)}\end{aligned}$$

where we recall $\mu_{n,r}$ is the $r$th moment of the EDF $\mathcal{F}_n$ of $u_n \equiv u_n(\theta_0)$, $\dot{\mu}_{n,r} = \frac{\partial}{\partial\theta_0}\mu_{n,r}$ and $\ddot{\mu}_{n,r} = \frac{\partial^2}{\partial\theta_0\partial\theta_0'}\mu_{n,r}$. Under Assumption F$(iii)$, we obtain $\text{Bias}(\hat{\mu}_{n,r}) = \mathbf{c}_{-1} + O(n^{-\frac{3}{2}})$, where

$$\mathbf{c}_{-1} = \big\{\text{tr}[\mathbf{\Omega}_n\text{E}(\psi_n\dot{\mu}_{n,r}')] + \text{E}(\dot{\mu}_{n,r}')[\mathbf{b}_{-1} - \mathbf{\Omega}_n\text{E}(\psi_n)] + \text{tr}[\text{E}(\ddot{\mu}_{n,r})\mathbf{\Omega}_n\mathbf{\Sigma}_n\mathbf{\Omega}_n]\big\}_{d\times 1}, \quad \text{(A-14)}$$

and $d$ denotes the dimension of $\mu_n$.

Finally, using $\hat{\theta}_n - \theta_0 = \mathbf{a}_{-1/2} + O_p(n^{-1}) = \mathbf{\Omega}_n\psi_n + O_p(n^{-1})$, and $\hat{\mu}_{n,r} = \mu_{n,r} + \dot{\mu}_{n,r}'(\hat{\theta}_n - \theta_0) + O_p(n^{-1}) = \mu_{n,r} + \dot{\mu}_{n,r}'\mathbf{a}_{-1/2} + O_p(n^{-1}) = \mu_{n,r} + \text{E}(\dot{\mu}_{n,r}')\mathbf{\Omega}_n\psi_n + O_p(n^{-1})$, we obtain

$$\begin{aligned}\mathbf{d}_{-2} &= \tfrac{1}{2}\text{tr}[\nu_{-1,\vartheta_0\vartheta_0'}\text{AVar}(\hat{\vartheta}_n)]\\ &= \tfrac{1}{2}\text{tr}(\nu_{-1,\theta_0\theta_0'}\mathbf{\Omega}_n\mathbf{\Sigma}_n\mathbf{\Omega}_n) + \text{tr}[\nu_{-1,\mu_0\theta_0'}\mathbf{\Omega}_n\text{Cov}(\psi_n, \varpi_n')] + \tfrac{1}{2}\text{tr}[\nu_{-1,\mu_0\mu_0'}\text{Var}(\varpi_n)], \quad \text{(A-15)}\end{aligned}$$

where $\varpi_n = \big\{\mu_{n,r} + \text{E}(\dot{\mu}_{n,r})\mathbf{\Omega}_n\psi_n\big\}_{d\times 1}$.

It is easy to verify that $\widehat{\mathbf{b}}_{-1} = \mathbf{b}_{-1} + O_p(n^{-\frac{3}{2}})$, and that the same results hold for $\widehat{\mathbf{c}}_{-1}$, $\widehat{\upsilon}_{-1,\theta_0}$, and $\widehat{\upsilon}_{-1,\mu_0}$. It follows that $\text{E}(\widehat{\upsilon}_{-1}^{\text{bc2}}) = \upsilon_{-1} + O(n^{-\frac{5}{2}})$ as $\widehat{\mathbf{d}}_{-2} = \mathbf{d}_{-2} + o_p(n^{-\frac{5}{2}})$. By Assumptions E$(i)$ and F$(i)$, $\widehat{\upsilon}_{-3/2} = \upsilon_{-3/2} + (\frac{\partial}{\partial\vartheta_0'}\upsilon_{-3/2})(\hat{\vartheta}_n - \vartheta_0) + O_p(n^{-\frac{5}{2}})$, which leads to $\text{E}(\widehat{\upsilon}_{-3/2}) = \upsilon_{-3/2} + O(n^{-\frac{5}{2}})$; and $\text{E}(\widehat{\upsilon}_{-2}) = \upsilon_{-2} + O(n^{-\frac{5}{2}})$. The rest is trivial.

# Appendix B: Proofs of the Results in Section 3

**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 $\{A_nB_n\}$ are uniformly bounded in both row and column sums,*

*(ii) the elements of $A_n$ are uniformly bounded and $tr(A_n) = O(n)$, and*

*(iii) the elements of $A_nC_n$ and $C_nA_n$ 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) $\lim_{n\to\infty} \frac{1}{n}X_n'X_n$ 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)* $\text{tr}(M_nA_n) = \text{tr}(A_n) + O(1)$, *(ii)* $\text{tr}(A_n'M_nA_n) = \text{tr}(A_n'A_n) + O(1)$,

*(iii)* $\text{tr}[(M_nA_n)^2] = \text{tr}(A_n^2) + O(1)$, *(iv)* $\text{tr}[(A_n'M_nA_n)^2] = \text{tr}[(A_n'A_n)^2] + O(1)$.

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

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

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

**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 $\varepsilon_n$ be a random $n$-vector of iid elements with mean zero, variance $\sigma^2$ and 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)* $\text{E}(\varepsilon_n'A_n\varepsilon_n) = O(\frac{n}{h_n})$, *(ii)* $\text{Var}(\varepsilon_n'A_n\varepsilon_n) = O(\frac{n}{h_n})$,

*(iii)* $\text{Var}(\varepsilon_n'A_n\varepsilon_n + b_n'\varepsilon_n) = O(\frac{n}{h_n})$, *(iv)* $\varepsilon_n'A_n\varepsilon_n = O_p(\frac{n}{h_n})$,

*(v)* $\varepsilon_n'A_n\varepsilon_n - \text{E}(\varepsilon_n'A_n\varepsilon_n) = O_p((\frac{n}{h_n})^{\frac{1}{2}})$, *(vi)* $\varepsilon_n'A_nb_n = O_p((\frac{n}{h_n})^{\frac{1}{2}})$.

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

**Lemma B.5** (Kelejian and Prucha, 2001, p.227, extended): *Let $A_n$ and $D_n$ be $n \times n$ matrices, $b_n$ an $n \times 1$ vector, and $\varepsilon_n$ an $n \times 1$ random vector of iid elements with mean zero, variance $\sigma^2$, skewness $\gamma$, and excess kurtosis $\kappa$. Let $Q_n = \varepsilon_n'A_n\varepsilon_n + b_n'\varepsilon_n$ and $S_n = \varepsilon_n'D_n\varepsilon_n$. Then, (i) $E(Q_n) = \sigma^2\text{tr}(A_n)$ and $E(S_n) = \sigma^2\text{tr}(D_n)$,*

*(ii)* $\text{Var}(Q_n) = \sigma^4\text{tr}[A_n(A_n + A_n')] + \sigma^4\kappa a_n'a_n + \sigma^2 b_n'b_n + 2\sigma^3\gamma a_n'b_n$,

*(iii)* $\text{Var}(S_n) = \sigma^4\text{tr}[D_n(D_n + D_n')] + \sigma^4\kappa d_n'd_n$,

*(iv)* $\text{Cov}(Q_n, S_n) = \sigma^4\text{tr}[A_n(D_n + D_n')] + \sigma^4\kappa a_n'd_n + \sigma^3\gamma b_n'd_n$,

*where $a_n$ and $d_n$ are column vectors of diagonal elements of $A_n$ and $D_n$, respectively.*

**Lemma B.6:** *For the SAR model specified by* (1) *satisfying Assumptions 1-6, the EDF* $\hat{\mathcal{F}}_n$ *of the QML residuals* $\hat{u}_n$ *is such that for each continuity point* $u$ *of* $\mathcal{F}_0$*,*

(*i*) $\mathrm{E}(\hat{\mathcal{F}}_n(u)) = \mathcal{F}_0(u) + O(\frac{h_n}{n})$, *and* $\hat{\mathcal{F}}_n(u) = \mathcal{F}_0(u) + O_p((\frac{h_n}{n})^{\frac{1}{2}})$*;*

*if further* $E|u_{n,i}|^{r+\epsilon}$ *exists for some* $\epsilon > 0$ *and* $r = 3, 4, \ldots$*, then the sample and population moments* $\hat{\mu}_n = \mu(\hat{\mathcal{F}}_n)$ *and* $\mu_0 = \mu(\mathcal{F}_0)$ *are such that*

(*ii*) $\hat{\mu}_n = \mu(\hat{\mathcal{F}}_n) = \mu_0 + O_p((\frac{h_n}{n})^{\frac{1}{2}})$, *and* $\mathrm{E}(\hat{\mu}_n) = \mu_0 + O(\frac{h_n}{n})$.

**Proof.** To prove $(i)$, assume W.L.O.G. $\sigma_0 = 1$, so that $\hat{\varepsilon}_n (= \hat{u}_n) = M_n A_n(\hat{\lambda}_n) Y_n = \tilde{\varepsilon}_n + \hat{r}_n$ where $\tilde{\varepsilon}_n = M_n \varepsilon_n$ is the OLS residuals from regressing $A_n Y_n$ on $X_n$ and $\hat{r}_n = -(\hat{\lambda}_n - \lambda_0) M_n W_n Y_n$. Let $\tilde{\mathcal{F}}_n$ be the EDF of $\tilde{\varepsilon}_n$. We first show that $(i)$ holds for $\tilde{\mathcal{F}}_n$.

Write $\tilde{\varepsilon}_n = \varepsilon_n + \tilde{r}_n$ where $\tilde{r}_n = (I_n - M_n)\varepsilon_n$. Note that $\mathrm{Var}(\tilde{r}_n) = I_n - M_n = O(n^{-1})$, and $\tilde{\mathcal{F}}_n(u) = \frac{1}{n}\sum_{i=1}^n 1(\tilde{\varepsilon}_{n,i} \le u)$ where $1(\cdot)$ is the indicator function. We have, $\mathrm{E}(\tilde{\mathcal{F}}_n(u)) = \frac{1}{n}\sum_{i=1}^n P(\tilde{\varepsilon}_{n,i} \le u) = \frac{1}{n}\sum_{i=1}^n P(\varepsilon_{n,i} + \tilde{r}_{n,i} \le u)$. For some $\epsilon > 0$,

$$
\begin{aligned}
P(\varepsilon_{n,i} + \tilde{r}_{n,i} \le u) &= P(\varepsilon_{n,i} + \tilde{r}_{n,i} \le u, |\tilde{r}_{n,i}| < \epsilon) + P(\varepsilon_{n,i} + \tilde{r}_{n,i} \le u, |\tilde{r}_{n,i}| \ge \epsilon) \\
&\le P(\varepsilon_{n,i} \le u + \epsilon) + P(|\tilde{r}_{n,i}| \ge \epsilon) \\
&= \mathcal{F}_0(u + \epsilon) + O(n^{-1}),
\end{aligned}
$$

where $P(|\tilde{r}_{n,i}| \ge \epsilon) \le \frac{1}{\epsilon^2}\mathrm{Var}(\tilde{r}_{n,i}) = O(n^{-1})$ by Chebyshev's inequality. Similarly, we have, $P(\varepsilon_{n,i} + \tilde{r}_{n,i} \le u) \ge \mathcal{F}_0(u - \epsilon) + O(n^{-1})$. These imply that $P(\tilde{\varepsilon}_{n,i} \le u) = \mathcal{F}_0(u) + O(n^{-1})$. Hence, $\mathrm{E}(\tilde{\mathcal{F}}_n(u)) = \frac{1}{n}\sum_{i=1}^n P(\tilde{\varepsilon}_{n,i} \le u) = \mathcal{F}_0(u) + O(n^{-1})$. For the second part, note that $\mathrm{Var}(\tilde{\mathcal{F}}_n(u)) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}\big(1(\tilde{\varepsilon}_{n,i} \le u)\big) + \frac{2}{n^2}\sum_{i=1}^n \sum_{j=i+1}^n \mathrm{Cov}\big(1(\tilde{\varepsilon}_{n,i} \le u), 1(\tilde{\varepsilon}_{n,j} \le u)\big)$, where $\mathrm{Var}\big(1(\tilde{\varepsilon}_{n,i} \le u)\big) = P(\tilde{\varepsilon}_{n,i} \le u)[1 - P(\tilde{\varepsilon}_{n,i} \le u)] = \mathcal{F}_0(u)[1 - \mathcal{F}_0(u)] + O(n^{-1})$, and $\mathrm{Cov}\big(1(\tilde{\varepsilon}_{n,i} \le u), 1(\tilde{\varepsilon}_{n,j} \le u)\big) = P(\tilde{\varepsilon}_{n,i} \le u,\ \tilde{\varepsilon}_{n,j} \le u) - P(\tilde{\varepsilon}_{n,i} \le u)P(\tilde{\varepsilon}_{n,j} \le u)$. We have,

$$
\begin{aligned}
P(\tilde{\varepsilon}_{n,i} \le u,\ \tilde{\varepsilon}_{n,j} \le u) &= P(\varepsilon_{n,i} + \tilde{r}_{n,i} \le u, \varepsilon_{n,j} + \tilde{r}_{n,j} \le u, |\tilde{r}_{n,i}| < \epsilon, |\tilde{r}_{n,j}| < \epsilon) \\
&\quad + P(\varepsilon_{n,i} + \tilde{r}_{n,i} \le u, \varepsilon_{n,i} + \tilde{r}_{n,j} \le u, (|\tilde{r}_{n,i}| \ge \epsilon \text{ or } |\tilde{r}_{n,j}| \ge \epsilon)) \\
&\le P(\varepsilon_{n,i} \le u + \epsilon, \varepsilon_{n,j} \le u + \epsilon) + P(|\tilde{r}_{n,i}| \ge \epsilon) + P(|\tilde{r}_{n,j}| \ge \epsilon) \\
&= \mathcal{F}_0(u + \epsilon)^2 + O(n^{-1}).
\end{aligned}
$$

Similarly, we have $P(\tilde{\varepsilon}_{n,i} \le u,\ \tilde{\varepsilon}_{n,j} \le u) \ge \mathcal{F}_0(u - \epsilon)^2 + O(n^{-1})$. It follows that $P(\tilde{\varepsilon}_{n,i} \le u,\ \tilde{u}_{\varepsilon,j} \le u) = \mathcal{F}_0(u)^2 + O(n^{-1})$, and that $\mathrm{Cov}\big(1(\tilde{\varepsilon}_{n,i} \le u), 1(\tilde{\varepsilon}_{n,j} \le u)\big) = O(n^{-1})$ as $P(\tilde{\varepsilon}_{n,i} \le u)P(\tilde{\varepsilon}_{n,j} \le u) = (\mathcal{F}_0(u) + O(n^{-1}))(\mathcal{F}_0(u) + O(n^{-1})) = \mathcal{F}_0(u)^2 + O(n^{-1})$. Thus, $\mathrm{Var}(\tilde{\mathcal{F}}_n(u)) = O(n^{-1})$. It follows that $\tilde{\mathcal{F}}_n(u) = \mathcal{F}_0(u) + O_p(n^{-\frac{1}{2}})$.

Now, moving from $\tilde{\mathcal{F}}_n$ to $\hat{\mathcal{F}}_n$, we have $\mathrm{Var}(\hat{r}_n) = \mathrm{Var}[(\hat{\lambda}_n - \lambda_0)M_n\eta_n + (\hat{\lambda}_n - \lambda_0)M_n G_n u_n]$ $= \mathrm{Var}(\hat{\lambda}_n)M_n\eta_n\eta_n' M_n + \mathrm{Var}((\hat{\lambda}_n - \lambda_0)M_n G_n u_n) + 2\mathrm{Cov}[(\hat{\lambda}_n - \lambda_0)M_n\eta_n, (\hat{\lambda}_n - \lambda_0)M_n G_n u_n]$ $= O(\frac{h_n}{n})$, by noticing that the elements of $M_n G_n u_n$ and of $M_n \eta_n u_n' G_n' m_n$ are $O_p(1)$, and that the elements of $M_n\eta_n\eta_n' M_n$ are $O(1)$. Repeating the same process as for $\tilde{\mathcal{F}}_n$ on $\hat{\mathcal{F}}_n$, and applying the results for $\tilde{\mathcal{F}}_n(u)$, we obtain, $\mathrm{E}(\hat{\mathcal{F}}_n(u)) = \mathrm{E}(\tilde{\mathcal{F}}_n(u)) + O(\frac{h_n}{n}) = \mathcal{F}_0(u) + O(\frac{h_n}{n})$ and $\hat{\mathcal{F}}_n(u) = \tilde{\mathcal{F}}_n(u) + O_p((\frac{h_n}{n})^{\frac{1}{2}}) = \mathcal{F}_0(u) + O_p((\frac{h_n}{n})^{\frac{1}{2}})$.

For $(ii)$, we prove the result for the case where $\mu_0$ is the third moment. The cases of higher moments can be proved in a similar manner but with more tedious algebraic work.

It suffices to work with the original residuals $\hat{\varepsilon}_n = M_n A_n(\hat{\lambda}_n)Y_n$ instead of the standardized residuals $\hat{u}_n = \hat{\sigma}_n^{-1}\hat{\varepsilon}_n$ as defined in the main text as bootstrap is invariant of $\hat{\sigma}_n$.

We have $\hat{\varepsilon}_n = \tilde{\varepsilon}_n - (\hat{\lambda}_n - \lambda_0)\bar{Y}_n$, where $\tilde{\varepsilon}_n = M_n\varepsilon_n$ and $\bar{Y}_n = W_nY_n$. Let $\odot$ denote the Hadamard product, then $\hat{\mu}_n = \frac{1}{n}(\hat{\varepsilon}_n \odot \hat{\varepsilon}_n)'\hat{\varepsilon}_n = \frac{1}{n}(\tilde{\varepsilon}_n \odot \tilde{\varepsilon}_n)'\tilde{\varepsilon}_n - \frac{3}{n}(\hat{\lambda}_n - \lambda_0)(\tilde{\varepsilon}_n \odot \tilde{\varepsilon}_n)'\bar{Y}_n + O_p(\frac{h_n}{n}) = \frac{1}{n}(\tilde{\varepsilon}_n \odot \tilde{\varepsilon}_n)'\tilde{\varepsilon}_n + O_p((\frac{h_n}{n})^{\frac{1}{2}})$. Thus $\hat{\mu}_n$ is $\sqrt{n/h_n}$-consistent as $\frac{1}{n}(\tilde{\varepsilon}_n \odot \tilde{\varepsilon}_n)'\tilde{\varepsilon}_n = \mu_0 + O_p(n^{-\frac{1}{2}})$ due to the properties of OLS residuals. If a quantity bounded in probability has a finite expectation, then $\mathrm{E}(\hat{\mu}_n) = \frac{1}{n}\mathrm{E}[(\tilde{\varepsilon}_n \odot \tilde{\varepsilon}_n)'\tilde{\varepsilon}_n] - \frac{3}{n}\mathrm{E}[(\hat{\lambda}_n - \lambda_0)(\tilde{\varepsilon}_n \odot \tilde{\varepsilon}_n)'\bar{Y}_n] + O(\frac{h_n}{n})$, where the first term equals $\mu_0 + O(n^{-1})$ by the property of OLS residuals, and the second term is $O(\frac{h_n}{n})$ because $\mathrm{E}(\hat{\lambda}_n - \lambda_0) = O(\frac{h_n}{n})$, $\mathrm{E}[(\tilde{\varepsilon}_n \odot \tilde{\varepsilon}_n)'\bar{Y}_n] = O(1)$, and $(\tilde{\varepsilon}_n \odot \tilde{\varepsilon}_n)'\bar{Y}_n - \mathrm{E}[(\tilde{\varepsilon}_n \odot \tilde{\varepsilon}_n)'\bar{Y}_n] = O_p((\frac{h_n}{n})^{\frac{1}{2}})$. Thus, $\mathrm{E}(\hat{\mu}_n) = \mu_0 + O(\frac{h_n}{n})$. *Q.E.D.*

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

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

where $\bar{\sigma}_{n0}^2$ lies between $\hat{\sigma}_{n0}^2$ and $\sigma_0^2$. 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 $\hat{\sigma}_{n0}^2 = \frac{1}{n}Y_n'A_n'M_nA_nY_n = \frac{1}{n}u_n'M_nu_n$, by Assumptions 2 and 5, and Lemmas B.2, B.1(*ii*) and B.5(*ii*), $\mathrm{Var}(\hat{\sigma}_{n0}^2) = O(n^{-1})$. By the generalized Chebyshev's inequality: $P(\sqrt{n}|\hat{\sigma}_{n0}^2 - \sigma_0^2| \geq M) \leq \frac{1}{M^2}n\mathrm{Var}(\hat{\sigma}_{n0}^2) = \frac{1}{M^2}O(1)$. 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*), $h_nR_{1n} = \frac{h_n}{n}\hat{\sigma}_{n0}^{-2}Y_n'A_n'M_nW_nY_n = \frac{h_n}{n}\hat{\sigma}_{n0}^{-2}(u_n'M_nG_nu_n + u_n'M_n\eta_n)$. Lemma B.4(*iv*) implies $\frac{h_n}{n}u_n'M_nG_nu_n = O_p(1)$ and Lemma B.4(*vi*) 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

$$h_nR_{1n} = \frac{h_n}{n\sigma_0^2}u_n'M_nG_nu_n + O_p((\tfrac{h_n}{n})^{\frac{1}{2}}) = O_p(1). \tag{B-2}$$

Now, by (B-1), $\mathrm{E}(h_nR_{1n}) = \sigma_0^{-2}\mathrm{E}[(\frac{h_n}{n}Y_n'A_n'M_nW_nY_n)] - \sigma_0^{-4}\mathrm{E}[(\frac{h_n}{n}Y_n'A_n'M_nW_nY_n)(\hat{\sigma}_{n0}^2 - \sigma_0^2)] - \mathrm{E}[(\frac{h_n}{n}Y_n'A_n'M_nW_nY_n)(\bar{\sigma}_{n0}^{-4} - \sigma_0^{-4})(\hat{\sigma}_{n0}^2 - \sigma_0^2))]$, where the 1st term equals $\frac{h_n}{n}\mathrm{tr}(M_nG_n) = O(1)$ and the 3rd term is $O((\frac{h_n}{n})^{\frac{1}{2}})$ by Assumption 7. For the 2nd term, the Cauchy-Schwarz inequality and Lemma B.4(*iii*) lead to,

$$\begin{aligned}
&|\mathrm{E}[(Y_n'A_n'M_nW_nY_n)(\hat{\sigma}_{n0}^2 - \sigma_0^2)]| \\
&\leq |\mathrm{E}[(Y_n'A_n'M_nW_nY_n - \sigma_0^2\mathrm{tr}(M_nG_n))(\hat{\sigma}_{n0}^2 - \sigma_0^2)]| + \sigma_0^2|\mathrm{E}[\mathrm{tr}(M_nG_n)(\hat{\sigma}_{n0}^2 - \sigma_0^2)]| \\
&\leq \tfrac{1}{n}\{\mathrm{Var}(u_n'M_nG_nu_n + u_n'M_n\eta_n)\mathrm{Var}(u_n'M_nu_n - n\sigma_0^2)\}^{\frac{1}{2}} + O(h_n^{-1}) \\
&= \tfrac{1}{n}\{O(\tfrac{n}{h_n})O(n)\}^{\frac{1}{2}} + O(h_n^{-1}) = O(h_n^{-\frac{1}{2}}),
\end{aligned}$$

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

$$\mathrm{E}(h_nR_{1n}) = \tfrac{h_n}{n}\mathrm{tr}(M_nG_n) + O(\tfrac{h_n^{1/2}}{n}) + O((\tfrac{h_n}{n})^{\frac{1}{2}}) = O(1). \tag{B-3}$$

Taking difference between (B-2) and (B-3) and using Lemma B.4(*v*), we obtain $h_nR_{1n} - \mathrm{E}(h_nR_{1n}) = \frac{h_n}{\sigma_0^2 n}u_n'M_nG_nu_n - \frac{h_n}{n}\mathrm{tr}(M_nG_n) + O_p((\frac{h_n}{n})^{\frac{1}{2}}) - O(\frac{h_n^{1/2}}{n}) = O_p((\frac{h_n}{n})^{\frac{1}{2}})$.

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

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

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

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

**Proof of Theorem 3.1.** Clearly, the $\tilde{\psi}(\lambda)$ function given in (13) is differentiable for $\lambda$ in a neighborhood of $\lambda_0$ 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:

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

where $\bar{\lambda}$ lies between $\hat{\lambda}_n$ and $\lambda_0$. 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 $\mathrm{E}(\tilde{\psi}_n) = O(\frac{h_n}{n})$;
$(b)$ $\mathrm{E}(H_{rn}) = O(1)$ 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 $\mathrm{E}(H_{1n})^{-1} = O(1)$; 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 $u_n' M_n G_n u_n - \sigma_0^2 \mathrm{tr}(M_n G_n) = O_p((\frac{n}{h_n})^{\frac{1}{2}})$ by Lemma B.4(v) and $\mathrm{tr}(M_n G_n) = \mathrm{tr}(G_n) + O(1) = nT_{0n} + O(1)$ by Lemma B.3$(i)$, $\tilde{\psi}_n = O_p((\frac{h_n}{n})^{\frac{1}{2}})$. By (B-3), $\mathrm{E}(\tilde{\psi}_n) = -h_n T_{0n} + \frac{h_n}{n}\mathrm{tr}(M_n G_n) + O(\frac{h_n^{1/2}}{n})$. By Lemma B.3$(i)$, $\mathrm{tr}(M_n G_n) = \mathrm{tr}(G_n) + O(1)$. It follows that $\mathrm{E}(\tilde{\psi}_n) = O(\frac{h_n}{n})$.

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

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

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

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

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

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

leading to $\hat{\sigma}_n^{-2}(\bar{\lambda}) = (\hat{\sigma}_{n0}^2 + O_p((h_nn)^{-\frac{1}{2}}))^{-1} = \hat{\sigma}_{n0}^{-2}(1+O_p((h_nn)^{-\frac{1}{2}}))^{-1} = \hat{\sigma}_{n0}^{-2} + O_p((h_nn)^{-\frac{1}{2}})$. Now, as $h_nR_{1n}(\bar{\lambda}) = \hat{\sigma}_n^{-2}(\bar{\lambda})\frac{h_n}{n}Y_n'A_n'(\bar{\lambda})M_nW_nY_n$, we have,

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

Similarly, one shows that $h_nR_{2n}(\bar{\lambda}) = h_nR_{2n} + O_p((\frac{h_n}{n})^{\frac{1}{2}})$. By the mean value theorem, $h_nT_{3n}(\bar{\lambda}) = \frac{h_n}{n}\mathrm{tr}(G_n^4(\bar{\lambda})) = \frac{h_n}{n}\mathrm{tr}(G_n^4) + 4\frac{h_n}{n}\mathrm{tr}(G_n^3(\check{\lambda}))(\bar{\lambda}-\lambda_0)$, where $\check{\lambda}$ lies between $\bar{\lambda}$ and $\lambda_0$. By Assumption 4 and Lemma B.1, $\frac{h_n}{n}\mathrm{tr}(G_n^3(\check{\lambda})) = O(1)$. It follows that $h_nT_{3n}(\bar{\lambda}) - h_nT_{3n} = O_p((\frac{h_n}{n})^{\frac{1}{2}})$. 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.** Straightforward.

**Proof of Corollary 3.5.** As $b_{-1}$ is differentiable to the desired order, the expansion (A-7) holds. By Assumptions 1-6 and Lemma B.6, $\hat{\vartheta}_n$ is $\sqrt{n/h_n}$-consistent and has bias of order $O(\frac{h_n}{n})$. It follows immediately that $\mathrm{E}(\widehat{b}_{-1}) = b_{-1} + O((\frac{h_n}{n})^2)$. Similarly, $b_{-3/2}$ is differentiable, giving $\mathrm{E}(\widehat{b}_{-3/2}) = b_{-3/2} + O((\frac{h_n}{n})^2)$. The results thus follow.

**Proof of Corollary 3.6.** As $\widehat{b}_{-1} = b_{-1}(\hat{\theta}_n, \mu_n(\hat{\theta}_n))$ is differentiable w.r.t. $\hat{\theta}_n$ up to the desired order, the expansion (A-8) holds with $n$ replaced by $n/h_n$, i.e.,

$$\begin{aligned}b_{-1}(\hat{\theta}_n, \mu_n(\hat{\theta}_n)) &= b_{-1}(\theta_0, \mu_n) + \tilde{b}'_{-1,\theta_0}(\hat{\theta}_n - \theta_0) + O_p((\tfrac{h_n}{n})^2) \\ &= b_{-1}(\theta_0, \mu_0) + b'_{-1,\mu_0}(\mu_n - \mu_0) + \mathrm{E}(\tilde{b}'_{-1,\theta_0})(\hat{\theta}_n - \theta_0) + O_p((\tfrac{h_n}{n})^2).\end{aligned}$$

This gives $\mathrm{ACov}(\hat{\lambda}, \widehat{b}_{-1}) = b'_{-1,\mu_0}\Omega_n\mathrm{Cov}(\tilde{\psi}_n, \mu_n) + \mathrm{E}(\tilde{b}'_{-1,\theta_0})\mathrm{ACov}(\hat{\lambda}_n, \hat{\theta}_n) + O((\frac{h_n}{n})^{\frac{5}{2}})$.

It is easy to show that $\tilde{\psi}_n = h_nT_{0n} + h_nR_{1n}^\circ - h_n\mathrm{E}(R_{1n}^\circ)(\hat{\sigma}_{n0}^2 - \sigma_0^2) + O_p(\frac{h_n}{n})$, where $R_{1n}^\circ = \frac{1}{n}(u_n'M_nG_nu_n + u_n'M_n\eta_n)$ and $\hat{\sigma}_{n0}^2 = \frac{1}{n}u_n'M_nu_n$. It follows that

$$\begin{aligned}\mathrm{Cov}(\tilde{\psi}_n, \mu_{n,r}) &= h_n\mathrm{Cov}(R_{1n}^\circ, \mu_{n,r}) - \tfrac{h_n}{n}\mathrm{tr}(M_nG_n)\mathrm{Cov}(\sigma_{n0}^2, \mu_{n,r}) + O((\tfrac{h_n}{n})^{\frac{3}{2}}) \\ &= \tfrac{h_n}{n^2}\mathrm{E}[(u_n'M_nG_nu_n + u_n'M_n\eta_n)\textstyle\sum_{i=1}^n(u_{ni}^r - \mu_{0,r})] \\ &\quad + \tfrac{h_n}{n^2}\mathrm{tr}(M_nG_n)\mathrm{E}[(u_n'M_nu_n)\textstyle\sum_{i=1}^n(u_{ni}^r - \mu_{0,r})] + O((\tfrac{h_n}{n})^{\frac{3}{2}}) \\ &= \tfrac{h_n}{n^2}(\mu_{0,r+2} - \mu_{0,r})\mathrm{tr}(M_nG_n) + \tfrac{h_n}{n^2}\mu_{0,r+1}1_n'M_n\eta_n \\ &\quad - \tfrac{h_n}{n^2}(\mu_{0,r+2} - \mu_{0,r})\mathrm{tr}(M_nG_n) + O((\tfrac{h_n}{n})^{\frac{3}{2}}) \\ &= O((\tfrac{h_n}{n})^{\frac{3}{2}}),\ r = 3, 4,\end{aligned}$$

where $\mu_{0,r}$ is the $r$th moment of $\mathcal{F}_0$. Thus, $b'_{-1,\mu_0}\Omega_n \text{Cov}(\tilde{\psi}_n, \mu_n) = O((\frac{h_n}{n})^{\frac{5}{2}})$, showing that Assumption F$(ii)$ is satisfied, and the first result of Corollary 3.6 follows.

The proof of the second result starts from deriving $\widehat{v}_{-1}^{\text{bc2}}$ under the SAR model. With the joint estimating function: $\psi_n(\theta) = \frac{h_n}{n}\big\{\frac{1}{\sigma^2}[\varepsilon'_n(\lambda,\beta)W_nY_n - \sigma^2\text{tr}(G_n(\lambda))], \quad \frac{1}{\sigma^2}\varepsilon_n(\lambda,\beta)'X_n, \frac{1}{2\sigma^4}[\varepsilon'_n(\lambda,\beta)\varepsilon_n(\lambda,\beta) - n\sigma^2]\big\}'$, where $\varepsilon_n(\lambda,\beta) = A_n(\lambda)Y_n - X_n\beta$, it is straightforward to derive $\mathbf{\Sigma}_n$, $\mathbb{I}_n$, $\mathbb{H}_{1n}$, and $\mathbb{H}_{2n}$, defined around (A-9). The top-left corner element of $\mathbb{I}_n^{-1}\mathbf{\Sigma}_n\mathbb{I}_n^{-1}$ or (A-9) leads to $\nu_{-1}$ given in (33). Obviously, $v_{-1}$ is differentiable up to the desired order. By Lemma B.6, the expansion (A-10) holds with $n$ replaced by $n/h_n$, i.e.,

$$v_{-1}(\hat{\vartheta}_n) = v_{-1}(\vartheta_0) + v'_{-1,\vartheta_0}(\hat{\vartheta}_n - \vartheta_0) + \tfrac{1}{2}(\hat{\vartheta}_n - \vartheta_0)'v_{-1,\vartheta_0\vartheta'_0}(\hat{\vartheta}_n - \vartheta_0) + O_p((\tfrac{h_n}{n})^{\frac{5}{2}}). \quad \text{(B-6)}$$

As $\hat{\theta}_n = \arg\{\psi_n(\theta) = 0\}$, which is $\sqrt{n/h_n}$-consistent in general, we have as in Bao (2013) a second-order stochastic expansion: $\hat{\theta}_n - \theta_0 = \mathbf{a}_{-1/2} + \mathbf{a}_{-1} + O_p((\frac{h_n}{n})^{\frac{3}{2}})$, where $\mathbf{a}_{-1/2} = \mathbf{\Omega}_n\psi_n$, $\mathbf{a}_{-1} = \mathbf{\Omega}_n\mathbb{H}^{\circ}_{1n}\mathbf{a}_{-1/2} + \frac{1}{2}\mathbf{\Omega}_n\text{E}(\mathbb{H}_{2n})(\mathbf{a}_{-1/2}\otimes\mathbf{a}_{-1/2})$, and $\mathbb{H}^{\circ}_{1n} = \mathbb{H}_{1n} - \text{E}(\mathbb{H}_{1n})$.

Further, as $\mu_{n,r}$ is differentiable w.r.t. $\theta_0$, we have similar to (A-13),

$$\begin{aligned}\hat{\mu}_{n,r} &= \mu_{n,r} + \dot{\mu}'_{n,r}(\hat{\theta}_n - \theta_0) + \tfrac{1}{2}(\hat{\theta}_n - \theta_0)'\ddot{\mu}_{n,r}(\hat{\theta}_n - \theta_0) + O_p((\tfrac{h_n}{n})^{\frac{3}{2}}),\\ &= \mu_{n,r} + \dot{\mu}'_{n,r}(\mathbf{a}_{-1/2} + \mathbf{a}_{-1}) + \tfrac{1}{2}\mathbf{a}'_{-1/2}\ddot{\mu}_{n,r}\mathbf{a}_{-1/2} + O_p((\tfrac{h_n}{n})^{\frac{3}{2}}), \quad \text{(B-7)}\end{aligned}$$

where, we recall, $\mu_{n,r}$ is the $r$th moment of the EDF $\mathcal{F}_n$ of $u_n \equiv u_n(\theta_0)$, and $\dot{\mu}_{n,r}$ and $\ddot{\mu}_{n,r}$ are its first- and second-order partial derivatives, $r = 3, 4$, which can easily be seen to satisfy Assumption F$(iii)$ with $n$ replaced by $n/h_n$. Similar arguments as in the proof of Corollary 2.6 lead to $v_{-1} = \nu_{-1} + \nu_{-3/2} + \nu_{-2} + O((\frac{h_n}{n})^{\frac{5}{2}})$. Putting everything together, we obtain

$$\text{E}[v_{-1}(\hat{\vartheta}_n)] = v_{-1} + \nu'_{-1,\theta_0}\mathbf{b}_{-1} + \nu'_{-1,\mu_0}\mathbf{c}_{-1} + \mathbf{d}_{-2} + O((\tfrac{h_n}{n})^{\frac{5}{2}}), \quad \text{(B-8)}$$

where, as $\text{E}(\psi_n) = 0$ and $\nu_{-1,\mu_0\mu'_0} = 0$, the quantities $\mathbf{b}_{-1}$, $\mathbf{c}_{-1}$ and $\mathbf{d}_{-2}$ are simplified to

$$\begin{aligned}&\mathbf{b}_{-1} = \mathbf{\Omega}_n\text{E}(\psi'_n\otimes\mathbb{H}_{1n})\text{vec}(\mathbf{\Omega}_n) + \tfrac{1}{2}\mathbf{\Omega}_n\text{E}(\mathbb{H}_{2n})(\mathbf{\Omega}_n\otimes\mathbf{\Omega}_n)\text{vec}(\mathbf{\Sigma}_n),\\ &\mathbf{c}_{-1} = \big\{\,\text{tr}[\mathbf{\Omega}_n\text{E}(\psi_n\dot{\mu}'_{n,r})] + \text{E}(\dot{\mu}'_{n,r})\mathbf{b}_{-1} + \text{tr}[\text{E}(\ddot{\mu}_{n,r})\mathbf{\Omega}_n\mathbf{\Sigma}_n\mathbf{\Omega}_n]\big\}_{2\times1},\\ &\mathbf{d}_{-2} = \tfrac{1}{2}\text{tr}(\nu_{-1,\theta_0\theta'_0}\mathbf{\Omega}_n\mathbf{\Sigma}_n\mathbf{\Omega}_n) + \text{tr}[\nu_{-1,\mu_0\theta'_0}\mathbf{\Omega}_n\text{E}(\psi_n\varpi'_n)],\end{aligned} \quad \text{(B-9)}$$

and $\varpi_n = \big\{\mu_{n,r} + \text{E}(\dot{\mu}'_{n,r})\mathbf{\Omega}_n\psi_n, r = 3, 4\big\}'$. Therefore, a bias-corrected estimator of $v_{-1}$ is:

$$\widehat{v}_{-1}^{\text{bc2}} = \widehat{v}_{-1} - \widehat{\nu}_{-1,\theta_0}\widehat{\mathbf{b}}_{-1} - \widehat{\nu}_{-1,\mu_0}\widehat{\mathbf{c}}_{-1} - \widehat{\mathbf{d}}_{-2}, \quad \text{(B-10)}$$

which can readily be shown to have a bias of order $O((\frac{h_n}{n})^{\frac{5}{2}})$ for estimating $v_{-1}$. The bias from estimating the other terms in $\text{Var}(\hat{\lambda}_n^{\text{bc3}})$ are all $O((\frac{h_n}{n})^{\frac{5}{2}})$, completing the proof.

**Practical Implementation of $\widehat{v}_{-1}^{\text{bc2}}$:** First note that to avoid the handling of the additional large matrices $\text{E}(\psi'_n\otimes\mathbb{H}_{1n})$ and $\text{E}(\mathbb{H}_{2n})$, $\mathbf{b}_{-1}$ can simply be estimated through $b_{-1}$, (31) and (32) as pointed out in the remarks given below (34).

For estimating $\text{E}(\dot{\mu}_{n,r})$, $\text{E}(\ddot{\mu}_{n,r})$, $\text{E}(\psi_n\dot{\mu}'_{n,r})$, and $\text{E}(\psi_n\varpi'_n) = \{\text{E}(\psi_n\mu_{n,r}) + \mathbf{\Sigma}_n\mathbf{\Omega}_n\text{E}(\dot{\mu}_{n,r})\}$, note that $u_n \equiv u_n(\theta_0) = \sigma_0^{-1}[A(\lambda_0)Y_n - X_n\beta_0]$ and $\mu_{n,r} = \frac{1}{n}\sum_{i=1}^n u_{ni}^r(\theta_0)$. We have,

$$\dot{\mu}_{n,r} = \tfrac{r}{n}\textstyle\sum_{i=1}^n u_{ni}^{r-1}\dot{u}_{ni} \;\text{ and }\; \ddot{\mu}_{n,r} = \tfrac{r}{n}\sum_{i=1}^n[(r-1)u_{ni}^{r-2}\dot{u}_{ni}\dot{u}_{ni}' + u_{ni}^{r-1}\ddot{u}_{ni}], \tag{B-11}$$

where $\dot{u}_{ni} = \frac{\partial}{\partial\theta_0}u_{ni}(\theta_0) = (-\frac{1}{\sigma_0}w_{ni}'Y_n, -\frac{1}{\sigma_0}x_{ni}', -\frac{1}{2\sigma_0^2}u_{ni}(\theta_0))'$ with $w_{ni}'$ being the $i$th row of $W_n$ and $x_{ni}'$ the $i$th row of $X_n$, and $\ddot{u}_{ni} = \frac{\partial^2}{\partial\theta_0\partial\theta_0'}u_{ni}(\theta_0)$ is a symmetric matrix with last column $a = (\frac{1}{2\sigma_0^3}w_{ni}'Y_n, \frac{1}{2\sigma_0^3}x_{ni}', \frac{3}{4\sigma_0^4}u_{ni}(\theta_0))'$, last row $a'$, and other elements zero. To facilitate the calculations, (B-11) can be expressed in matrix form, for $r = 3, 4$,

$$\dot{\mu}_{n,r} = \tfrac{r}{n}\dot{u}_n'u_n^{r-1} \;\text{ and }\; \ddot{\mu}_{n,r} = \tfrac{r(r-1)}{n}\dot{u}_n'(\dot{u}_n \odot (1_{p+2}' \otimes u_n^{r-2})) + \tfrac{r}{n}D_r, \tag{B-12}$$

where $\dot{u}_n = (\dot{u}_{n1}, \ldots, \dot{u}_{nn})' = -(\frac{1}{\sigma_0}W_nY_n, \frac{1}{\sigma_0}X_n, \frac{1}{2\sigma_0^2}u_n(\theta_0))$, $a^r = \{a_i^r\}$ for a vector $a$, $1_m$ is an $m\times 1$ vector of ones, and $D_r$ has last row $a = \big(\frac{1}{2\sigma_0^3}Y_n'W_n'u_n^{r-1}, \frac{1}{2\sigma_0^3}(X_n'u_n^{r-1})', \frac{3}{4\sigma_0^4}1_n'u_n^r\big)$, last column $a'$, and other elements zero. Clearly, $\mathrm{E}(\dot{\mu}_{n,r})$ and $\mathrm{E}(\ddot{\mu}_{n,r})$ can be estimated by their sample analogue, i.e., $\dot{\mu}_{n,r}$ and $\ddot{\mu}_{n,r}$ evaluated at $\hat{\theta}_n$. $\mathrm{E}(\psi_n\dot{\mu}_{n,r}')$ and $\mathrm{E}(\psi_n\mu_{n,r})$ can be easily estimated by bootstrap, although their analytical expressions can be derived.

It left with the estimation of $v_{-1,\vartheta}$ and $v_{-1,\vartheta\vartheta'}$. Rewrite (33) as $v_{-1} = \tau_n^{-1} + \tau_n^{-2}\kappa_n$, where $\tau_n = \sigma_0^{-2}\beta_0'\Phi_n\Phi_n\beta_0 + \mathcal{G}_n$, $\kappa_n = 2\sigma_0^{-1}\mu_{0,3}g_n^{\circ\prime}\Phi_n\beta_0 + (\mu_{0,4}-3)g_n^{\circ\prime}g_n^\circ$, $\Phi_n = M_nG_nX_n$, and $\mathcal{G}_n = \mathrm{tr}[G_n^\circ(G_n^\circ + G_n^{\circ\prime})]$. Noting that $g_n^\circ$, $\Phi_n$ and $\mathcal{G}_n$ depend only on $\lambda_0$, $\tau_n$ depends only on $\theta_0$, and $\kappa_n$ depends on both $\theta_0$ and $\mu_0$ and linear in $\mu_0$, we obtain,

$$\begin{aligned}
v_{-1,\theta_0} &= \tau_n^{-2}(\kappa_{n,\theta_0} - \dot{\tau}_n) - 2\tau_n^{-3}\dot{\tau}_n\kappa_n; & v_{-1,\mu_0} &= \tau_n^{-2}\kappa_{n,\mu_0};\\
v_{-1,\theta_0\mu_0'} &= \tau_n^{-2}\kappa_{n,\theta_0\mu_0'} - 2\tau_n^{-3}\dot{\tau}_n\kappa_{n,\mu_0}'; & v_{-1,\mu_0\mu_0'} &= \tau_n^{-2}\kappa_{n,\mu_0\mu_0'};\\
v_{-1,\theta_0\theta_0'} &= 2\tau_n^{-3}(\dot{\tau}_n\dot{\tau}_n' - \kappa_{n,\theta_0}\dot{\tau}_n' - \ddot{\tau}_n\kappa_n - \dot{\tau}_n\kappa_{n,\theta_0}') + \tau_n^{-2}(\kappa_{n,\theta_0\theta_0'} - \ddot{\tau}_n) + 6\tau_n^{-4}\dot{\tau}_n\dot{\tau}_n'\kappa_n. &&
\end{aligned}$$

where $\dot{\tau}_n$ and $\ddot{\tau}_n$ are the gradient and Hessian of $\tau_n$, having the forms

$$\dot{\tau}_n = \begin{pmatrix} 2\sigma_0^{-2}\beta_0'\Phi_n'\dot{\Phi}_n\beta_0 + \dot{\mathcal{G}}_n \\ 2\sigma_0^{-2}\Phi_n'\Phi_n\beta_0 \\ -\sigma_0^{-4}\beta_0'\Phi_n'\Phi_n\beta_0 \end{pmatrix} \;\text{ and }\; \ddot{\tau}_n = \begin{pmatrix} \ddot{\tau}_{n,11}, & 4\sigma_0^{-2}\beta_0'\Phi_n'\dot{\Phi}_n, & -2\sigma_0^{-4}\beta_0'\Phi_n'\dot{\Phi}_n\beta_0 \\ \sim, & 2\sigma_0^{-2}\Phi_n'\Phi_n, & -2\sigma_0^{-4}\Phi_n'\Phi_n\beta_0 \\ \sim, & \sim, & 2\sigma_0^{-6}\beta_0'\Phi_n'\Phi_n\beta_0 \end{pmatrix}$$

where $\ddot{\tau}_{n,11} = 2\sigma_0^{-2}\beta_0'(\dot{\Phi}_n'\dot{\Phi}_n + \Phi_n'\ddot{\Phi}_n)\beta_0 + \ddot{\mathcal{G}}_n$; and the partial derivatives of $\kappa_n$ have the forms:
$\kappa_{n,\theta_0} = \big\{2\sigma_0^{-1}\mu_{0,3}(\dot{g}_n^{\circ\prime}\Phi_n + g_n^{\circ\prime}\dot{\Phi}_n)\beta_0 + 2(\mu_{0,4}-3)\dot{g}_n^{\circ\prime}g_n^\circ,\ 2\sigma_0^{-1}\mu_{0,3}g_n^{\circ\prime}\Phi_n,\ -\sigma_0^{-3}\mu_{0,3}g_n^{\circ\prime}\Phi_n\beta_0\big\}'$,
$\kappa_{n,\mu_0} = \big(2\sigma_0^{-1}g_n^{\circ\prime}\Phi_n\beta_0,\ g_n^{\circ\prime}g_n^\circ\big)'$, $\kappa_{n,\mu_0\mu_0'} = 0_{2\times 2}$,

$$\kappa_{n,\mu_0\theta_0'} = \begin{pmatrix} 2\sigma_0^{-1}(\dot{g}_n^{\circ\prime}\Phi_n + g_n^{\circ\prime}\dot{\Phi}_n)\beta_0, & 2\sigma_0^{-1}g_n^{\circ\prime}\Phi_n, & -\sigma_0^{-3}g_n^{\circ\prime}\Phi_n\beta_0 \\ 2\dot{g}_n^{\circ\prime}g_n^\circ, & 0, & 0 \end{pmatrix}, \quad \text{and}$$

$$\kappa_{n,\theta_0\theta_0'} = \begin{pmatrix} \kappa_{n\lambda_0\lambda_0'}, & 2\sigma_0^{-1}\mu_{0,3}(\dot{g}_n^{\circ\prime}\Phi_n + g_n^{\circ\prime}\dot{\Phi}_n), & -\sigma_0^{-3}\mu_{0,3}(\dot{g}_n^{\circ\prime}\Phi_n + g_n^{\circ\prime}\dot{\Phi}_n)\beta_0 \\ \sim, & 0, & -\sigma_0^{-3}\mu_{0,3}\Phi_n'g_n^\circ \\ \sim, & \sim, & \frac{3}{2}\sigma_0^{-5}\mu_{0,3}\dot{g}_n^{\circ\prime}\Phi_n\beta_0 \end{pmatrix},$$

where $\kappa_{n\lambda_0\lambda_0'} = 2\sigma_0^{-1}\mu_{0,3}(\ddot{g}_n^{\circ\prime}\Phi_n + 2\dot{g}_n^{\circ\prime}\dot{\Phi}_n + g_n^{\circ\prime}\ddot{\Phi}_n)\beta_0 + 2(\mu_{0,4}-3)(\ddot{g}_n^{\circ\prime}g_n^\circ + \dot{g}_n^{\circ\prime}\dot{g}_n^\circ)$, and the single dotted and double dotted $g_n^\circ$, $\Phi_n$ and $\mathcal{G}_n$ denote the first and second derivatives.

Finally, we have $\dot{\mathcal{G}}_n = 2\mathrm{tr}[\dot{G}_n^\circ(G_n^\circ + G_n^{\circ\prime})]$ and $\ddot{\mathcal{G}}_n = 2\mathrm{tr}[\ddot{G}_n^\circ(G_n^\circ + G_n^{\circ\prime}) + \dot{G}_n^\circ(\dot{G}_n^\circ + \dot{G}_n^{\circ\prime})]$, which can be easily calculated based on $\dot{G}_n = G_n^2$, and $\ddot{G}_n = 2G_n^3$.

# 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., 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review* 26, 153-166.

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., 2013. Finite sample bias of the QMLE in spatial autoregressive models. *Econometric Theory* 29, 68-88.

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.

Bao, Y., Ullah, A., 2009. Higher-Order Bias and MSE of Nonlinear Estimators. *Pakistan Journal of Statistics* 25, 587-594.

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., Mullaninathan, 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.

Horn, R. A., Johnson, C. R., 1985. *Matrix Analysis.* Cambridge University Press.

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.

Kundhi, G., Rilstone, P., 2008. The third-order bias of nonlinear estimators. *Communications in Statistics – Theory and Methods* 37, 2617-2633.

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.

Yang, Z. L., Tse, Y. K., 2006. Modeling the firm-size distribution using Box-Cox heteroscedastic regression. *Journal of Applied Econometrics* 21, 541-653.

Yang, Z. L., Tse, Y. K., 2007. A corrected plug-in method for quantile interval construction through a transformed regression. *Journal of Business and Economic Statistics* 25, 356-376.

**Table 1.** Empirical Mean[rmse](sd) of Estimators of $\lambda$, and Averaged Bootstrap SEs

| $\lambda$ | $n$ | $\hat{\lambda}_n$ | $\hat{\lambda}_n^{\text{bc2}}$ | $\hat{\lambda}_n^{\text{bc3}}$ | $\overline{se}_1$ | $\overline{se}_2$ | $\overline{se}_3$ | $\overline{se}_3^c$ |
|---|---|---|---|---|---|---|---|---|
| | | `(a) Queen Contiguity, Normal Errors, MRSAR-A` | | | | | | |
| .50 | 50 | .411 [.195](.174) | .492 [.175](.175) | .497 [.175](.175) | .159 | .171 | .179 | .172 |
| | 100 | .459 [.123](.116) | .498 [.117](.117) | .500 [.117](.117) | .113 | .116 | .120 | .116 |
| | 200 | .480 [.078](.076) | .499 [.075](.075) | .499 [.075](.075) | .073 | .074 | .076 | .075 |
| | 500 | .493 [.049](.048) | .501 [.048](.048) | .501 [.048](.048) | .048 | .048 | .049 | .049 |
| .25 | 50 | .163 [.222](.204) | .242 [.209](.209) | .246 [.210](.210) | .190 | .203 | .202 | .204 |
| | 100 | .212 [.146](.140) | .248 [.142](.142) | .250 [.143](.143) | .136 | .139 | .139 | .140 |
| | 200 | .231 [.094](.092) | .250 [.093](.093) | .250 [.093](.093) | .090 | .092 | .092 | .092 |
| | 500 | .242 [.060](.060) | .250 [.060](.060) | .250 [.060](.060) | .060 | .060 | .060 | .060 |
| .00 | 50 | -.078 [.229](.216) | -.006 [.224](.224) | -.003 [.226](.226) | .210 | .222 | .215 | .225 |
| | 100 | -.034 [.157](.153) | -.002 [.156](.156) | -.001 [.157](.157) | .151 | .154 | .151 | .155 |
| | 200 | -.018 [.106](.104) | -.000 [.105](.105) | .000 [.105](.105) | .103 | .104 | .104 | .104 |
| | 500 | -.008 [.068](.067) | -.000 [.068](.068) | -.000 [.068](.068) | .068 | .068 | .068 | .068 |
| -.25 | 50 | -.317 [.233](.223) | -.255 [.236](.236) | -.254 [.237](.237) | .221 | .232 | .220 | .236 |
| | 100 | -.279 [.164](.161) | -.253 [.166](.166) | -.253 [.166](.166) | .158 | .161 | .156 | .163 |
| | 200 | -.266 [.112](.111) | -.252 [.112](.112) | -.251 [.112](.112) | .110 | .112 | .110 | .112 |
| | 500 | -.256 [.073](.072) | -.250 [.073](.073) | -.250 [.073](.073) | .072 | .073 | .072 | .072 |
| -.50 | 50 | -.552 [.228](.222) | -.504 [.236](.236) | -.504 [.237](.237) | .223 | .232 | .217 | .237 |
| | 100 | -.519 [.162](.161) | -.501 [.166](.166) | -.501 [.166](.166) | .159 | .161 | .155 | .163 |
| | 200 | -.514 [.113](.113) | -.502 [.114](.114) | -.502 [.114](.114) | .113 | .114 | .112 | .114 |
| | 500 | -.505 [.073](.073) | -.500 [.073](.073) | -.500 [.073](.073) | .074 | .074 | .073 | .074 |
| | | `(b) Queen Contiguity, Normal Mixture Errors, MRSAR-A` | | | | | | |
| .50 | 50 | .420 [.182](.164) | .494 [.165](.165) | .498 [.165](.165) | .149 | .160 | .167 | .162 |
| | 100 | .462 [.120](.114) | .499 [.114](.114) | .500 [.114](.114) | .108 | .111 | .115 | .112 |
| | 200 | .482 [.076](.074) | .500 [.074](.074) | .500 [.074](.074) | .071 | .072 | .074 | .074 |
| | 500 | .492 [.049](.048) | .500 [.048](.048) | .500 [.048](.048) | .048 | .048 | .048 | .048 |
| .25 | 50 | .169 [.207](.190) | .241 [.195](.195) | .244 [.195](.195) | .179 | .190 | .191 | .194 |
| | 100 | .213 [.140](.135) | .248 [.136](.136) | .249 [.137](.137) | .130 | .133 | .134 | .136 |
| | 200 | .230 [.092](.090) | .249 [.090](.090) | .249 [.090](.090) | .088 | .089 | .090 | .090 |
| | 500 | .242 [.060](.060) | .250 [.060](.060) | .250 [.060](.060) | .059 | .059 | .060 | .060 |
| .00 | 50 | -.070 [.217](.206) | -.004 [.213](.213) | -.002 [.214](.214) | .197 | .207 | .204 | .215 |
| | 100 | -.032 [.150](.147) | -.002 [.150](.150) | -.001 [.150](.150) | .145 | .148 | .146 | .151 |
| | 200 | -.018 [.104](.103) | -.001 [.103](.103) | -.001 [.103](.103) | .100 | .101 | .101 | .102 |
| | 500 | -.008 [.068](.067) | -.001 [.067](.067) | -.001 [.067](.067) | .067 | .067 | .067 | .067 |
| -.25 | 50 | -.314 [.223](.213) | -.258 [.224](.224) | -.257 [.225](.225) | .208 | .216 | .209 | .226 |
| | 100 | -.275 [.155](.153) | -.251 [.157](.157) | -.250 [.157](.157) | .152 | .154 | .151 | .158 |
| | 200 | -.263 [.111](.110) | -.249 [.112](.112) | -.249 [.112](.112) | .108 | .109 | .108 | .111 |
| | 500 | -.257 [.072](.072) | -.251 [.072](.072) | -.251 [.072](.072) | .071 | .072 | .071 | .072 |
| -.50 | 50 | -.550 [.218](.212) | -.506 [.224](.224) | -.505 [.225](.225) | .210 | .216 | .207 | .228 |
| | 100 | -.520 [.155](.153) | -.503 [.158](.158) | -.503 [.158](.158) | .152 | .154 | .150 | .159 |
| | 200 | -.513 [.112](.111) | -.502 [.113](.113) | -.502 [.113](.113) | .111 | .111 | .110 | .112 |
| | 500 | -.505 [.074](.073) | -.500 [.074](.074) | -.500 [.074](.074) | .073 | .073 | .073 | .074 |

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

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

| $\lambda$ | $n$ | $\hat{\lambda}_n$ | $\hat{\lambda}_n^{\text{bc2}}$ | $\hat{\lambda}_n^{\text{bc3}}$ | $\overline{se}_1$ | $\overline{se}_2$ | $\overline{se}_3$ | $\overline{se}_3^c$ |
|---|---|---|---|---|---|---|---|---|
| | | (c) Queen Contiguity, Lognormal Errors, MRSAR-A | | | | | | |
| .50 | 50 | .426 [.163](.146) | .491 [.146](.146) | .493 [.146](.146) | .138 | .144 | .154 | .151 |
| | 100 | .465 [.110](.105) | .498 [.105](.105) | .498 [.105](.105) | .102 | .103 | .106 | .104 |
| | 200 | .482 [.072](.069) | .499 [.069](.069) | .499 [.069](.069) | .067 | .067 | .069 | .069 |
| | 500 | .491 [.047](.046) | .499 [.046](.046) | .499 [.046](.046) | .046 | .046 | .046 | .046 |
| .25 | 50 | .179 [.185](.171) | .241 [.174](.174) | .244 [.174](.174) | .163 | .170 | .176 | .180 |
| | 100 | .216 [.128](.124) | .247 [.126](.125) | .248 [.126](.125) | .123 | .123 | .124 | .127 |
| | 200 | .232 [.087](.085) | .249 [.085](.085) | .249 [.085](.085) | .084 | .083 | .085 | .085 |
| | 500 | .242 [.058](.057) | .249 [.057](.057) | .249 [.057](.057) | .057 | .057 | .057 | .057 |
| .00 | 50 | -.067 [.198](.186) | -.011 [.192](.191) | -.008 [.192](.192) | .180 | .186 | .190 | .200 |
| | 100 | -.029 [.139](.136) | -.003 [.138](.138) | -.002 [.138](.138) | .135 | .136 | .136 | .141 |
| | 200 | -.017 [.099](.097) | -.002 [.098](.098) | -.001 [.098](.098) | .095 | .095 | .096 | .097 |
| | 500 | -.007 [.065](.064) | -.000 [.065](.065) | .000 [.065](.065) | .065 | .064 | .065 | .065 |
| -.25 | 50 | -.307 [.199](.191) | -.258 [.198](.198) | -.256 [.199](.199) | .189 | .194 | .197 | .212 |
| | 100 | -.272 [.142](.140) | -.252 [.144](.144) | -.251 [.144](.144) | .141 | .142 | .141 | .149 |
| | 200 | -.264 [.105](.104) | -.251 [.105](.105) | -.250 [.105](.105) | .102 | .102 | .102 | .104 |
| | 500 | -.256 [.070](.070) | -.250 [.070](.070) | -.250 [.070](.070) | .069 | .069 | .069 | .070 |
| -.50 | 50 | -.548 [.196](.190) | -.509 [.200](.199) | -.507 [.200](.200) | .191 | .195 | .196 | .215 |
| | 100 | -.514 [.145](.144) | -.500 [.148](.148) | -.499 [.148](.148) | .141 | .141 | .141 | .150 |
| | 200 | -.511 [.106](.106) | -.501 [.107](.107) | -.501 [.107](.107) | .105 | .105 | .105 | .106 |
| | 500 | -.505 [.070](.070) | -.501 [.070](.070) | -.500 [.070](.070) | .071 | .070 | .070 | .070 |
| | | (d) Group Interaction with $k = n^{0.5}$, Normal Errors, MRSAR-B | | | | | | |
| .50 | 50 | .426 [.145](.124) | .495 [.122](.122) | .499 [.122](.122) | .111 | .123 | .128 | .121 |
| | 100 | .449 [.112](.099) | .498 [.097](.097) | .500 [.097](.097) | .092 | .097 | .102 | .096 |
| | 200 | .474 [.073](.068) | .499 [.067](.067) | .500 [.067](.067) | .065 | .067 | .069 | .068 |
| | 500 | .491 [.042](.041) | .500 [.041](.041) | .500 [.041](.041) | .040 | .041 | .041 | .041 |
| .25 | 50 | .142 [.209](.179) | .239 [.178](.178) | .244 [.179](.178) | .159 | .175 | .180 | .173 |
| | 100 | .177 [.160](.143) | .248 [.141](.141) | .250 [.141](.141) | .133 | .140 | .146 | .139 |
| | 200 | .212 [.108](.102) | .249 [.101](.101) | .249 [.101](.101) | .096 | .098 | .101 | .100 |
| | 500 | .236 [.062](.060) | .250 [.061](.061) | .250 [.061](.061) | .060 | .060 | .060 | .061 |
| .00 | 50 | -.134 [.261](.224) | -.013 [.226](.226) | -.007 [.227](.227) | .202 | .220 | .223 | .220 |
| | 100 | -.094 [.206](.183) | -.005 [.182](.182) | -.003 [.182](.182) | .172 | .180 | .186 | .180 |
| | 200 | -.047 [.137](.128) | .001 [.128](.128) | .001 [.128](.128) | .125 | .128 | .131 | .129 |
| | 500 | -.019 [.081](.079) | -.000 [.079](.079) | -.000 [.079](.079) | .078 | .079 | .079 | .079 |
| -.25 | 50 | -.404 [.303](.260) | -.265 [.266](.266) | -.259 [.267](.267) | .240 | .258 | .259 | .261 |
| | 100 | -.358 [.243](.218) | -.253 [.219](.219) | -.251 [.219](.219) | .207 | .215 | .221 | .217 |
| | 200 | -.308 [.170](.160) | -.251 [.160](.160) | -.250 [.160](.160) | .153 | .156 | .159 | .159 |
| | 500 | -.274 [.101](.098) | -.252 [.098](.098) | -.252 [.098](.098) | .097 | .098 | .098 | .098 |
| -.50 | 50 | -.670 [.337](.291) | -.516 [.302](.301) | -.511 [.303](.302) | .273 | .290 | .288 | .297 |
| | 100 | -.620 [.281](.254) | -.501 [.257](.257) | -.499 [.257](.257) | .238 | .246 | .252 | .250 |
| | 200 | -.568 [.198](.186) | -.502 [.187](.187) | -.501 [.187](.187) | .179 | .183 | .186 | .186 |
| | 500 | -.527 [.118](.115) | -.500 [.116](.116) | -.500 [.116](.116) | .115 | .116 | .116 | .116 |

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

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

| $\lambda$ | $n$ | $\hat{\lambda}_n$ | $\hat{\lambda}_n^{\text{bc2}}$ | $\hat{\lambda}_n^{\text{bc3}}$ | $\overline{se}_1$ | $\overline{se}_2$ | $\overline{se}_3$ | $\overline{se}_3^c$ |
|---|---|---|---|---|---|---|---|---|
| | | (e) Group Interaction with $k = n^{0.5}$, Normal Mixture Errors, MRSAR-B | | | | | | |
| .50 | 50 | .427 [.144](.124) | .489 [.121](.120) | .493 [.120](.120) | .104 | .114 | .120 | .115 |
| | 100 | .449 [.111](.098) | .495 [.096](.096) | .497 [.096](.096) | .088 | .093 | .098 | .093 |
| | 200 | .474 [.073](.068) | .498 [.068](.068) | .499 [.068](.068) | .064 | .065 | .067 | .067 |
| | 500 | .490 [.042](.041) | .500 [.041](.041) | .500 [.041](.041) | .040 | .040 | .041 | .041 |
| .25 | 50 | .147 [.203](.175) | .234 [.172](.171) | .239 [.172](.171) | .149 | .163 | .171 | .167 |
| | 100 | .179 [.157](.140) | .245 [.138](.137) | .248 [.138](.138) | .128 | .133 | .140 | .135 |
| | 200 | .214 [.104](.097) | .249 [.097](.097) | .249 [.097](.097) | .093 | .095 | .099 | .098 |
| | 500 | .236 [.062](.060) | .250 [.060](.060) | .250 [.060](.060) | .059 | .059 | .060 | .060 |
| .00 | 50 | -.126 [.250](.215) | -.017 [.214](.213) | -.011 [.214](.214) | .191 | .207 | .215 | .214 |
| | 100 | -.089 [.203](.182) | -.006 [.180](.180) | -.004 [.181](.181) | .164 | .171 | .179 | .174 |
| | 200 | -.047 [.137](.129) | -.002 [.129](.128) | -.001 [.129](.129) | .122 | .124 | .128 | .128 |
| | 500 | -.020 [.082](.080) | -.002 [.080](.080) | -.002 [.080](.080) | .078 | .078 | .079 | .080 |
| -.25 | 50 | -.393 [.294](.256) | -.267 [.259](.259) | -.261 [.260](.259) | .227 | .244 | .251 | .255 |
| | 100 | -.358 [.241](.216) | -.259 [.216](.216) | -.257 [.216](.216) | .199 | .205 | .213 | .211 |
| | 200 | -.306 [.165](.156) | -.251 [.155](.155) | -.250 [.155](.155) | .149 | .151 | .156 | .156 |
| | 500 | -.274 [.101](.098) | -.251 [.099](.099) | -.251 [.099](.099) | .096 | .096 | .097 | .099 |
| -.50 | 50 | -.658 [.330](.289) | -.519 [.298](.297) | -.513 [.298](.298) | .260 | .277 | .282 | .291 |
| | 100 | -.618 [.270](.243) | -.507 [.245](.245) | -.505 [.245](.245) | .230 | .237 | .245 | .245 |
| | 200 | -.566 [.197](.185) | -.503 [.186](.186) | -.502 [.186](.186) | .175 | .178 | .183 | .184 |
| | 500 | -.528 [.120](.117) | -.501 [.117](.117) | -.501 [.117](.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](.106) | .094 | .102 | .117 | .110 |
| | 100 | .452 [.104](.092) | .492 [.090](.090) | .494 [.090](.090) | .083 | .085 | .092 | .087 |
| | 200 | .475 [.071](.067) | .497 [.066](.066) | .497 [.066](.066) | .060 | .060 | .065 | .064 |
| | 500 | .490 [.041](.039) | .499 [.040](.040) | .499 [.040](.040) | .038 | .037 | .039 | .040 |
| .25 | 50 | .165 [.180](.158) | .236 [.156](.155) | .241 [.155](.155) | .135 | .146 | .166 | .160 |
| | 100 | .184 [.146](.131) | .242 [.128](.128) | .244 [.128](.128) | .118 | .120 | .130 | .126 |
| | 200 | .215 [.101](.094) | .247 [.094](.093) | .247 [.093](.093) | .088 | .088 | .094 | .094 |
| | 500 | .235 [.060](.059) | .248 [.059](.059) | .248 [.059](.059) | .056 | .056 | .058 | .059 |
| .00 | 50 | -.109 [.231](.203) | -.020 [.202](.201) | -.014 [.201](.201) | .174 | .187 | .210 | .206 |
| | 100 | -.081 [.179](.160) | -.008 [.159](.159) | -.006 [.159](.158) | .151 | .152 | .164 | .161 |
| | 200 | -.046 [.131](.123) | -.005 [.122](.122) | -.004 [.122](.122) | .114 | .114 | .122 | .122 |
| | 500 | -.018 [.078](.076) | -.001 [.077](.077) | -.001 [.077](.076) | .074 | .073 | .077 | .077 |
| -.25 | 50 | -.377 [.274](.243) | -.272 [.245](.244) | -.265 [.244](.244) | .210 | .225 | .253 | .248 |
| | 100 | -.346 [.216](.193) | -.261 [.194](.194) | -.259 [.194](.193) | .181 | .181 | .194 | .193 |
| | 200 | -.307 [.159](.148) | -.258 [.148](.148) | -.257 [.148](.148) | .140 | .139 | .148 | .148 |
| | 500 | -.273 [.099](.096) | -.252 [.097](.097) | -.252 [.097](.097) | .092 | .091 | .095 | .095 |
| -.50 | 50 | -.639 [.312](.279) | -.524 [.286](.285) | -.516 [.285](.285) | .241 | .258 | .285 | .285 |
| | 100 | -.610 [.245](.219) | -.516 [.222](.221) | -.514 [.222](.221) | .209 | .208 | .220 | .221 |
| | 200 | -.565 [.186](.174) | -.508 [.174](.174) | -.507 [.174](.174) | .165 | .164 | .173 | .174 |
| | 500 | -.527 [.117](.114) | -.503 [.114](.114) | -.502 [.114](.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 $\overline{se}_3^c = \text{mean}(\widehat{V}_3(\hat{\lambda}_n^{\text{bc3}})^{\frac{1}{2}})$.

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

| | | | | | Empirical Tail Probabilities: L=left, R=right | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | Test | dgp | Mean | SD | 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 | 0.0151 | 1.0400 | 0.0122 | 0.0279 | 0.0527 | 0.0565 | 0.0323 | 0.0156 |
| | | 2 | -0.0206 | 1.0105 | 0.0138 | 0.0283 | 0.0539 | 0.0487 | 0.0282 | 0.0131 |
| | | 3 | -0.0191 | 0.9636 | 0.0175 | 0.0324 | 0.0533 | 0.0354 | 0.0162 | 0.0063 |
| 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 | 0.0345 | 1.0163 | 0.0094 | 0.0244 | 0.0484 | 0.0564 | 0.0299 | 0.0135 |
| | | 2 | 0.0455 | 1.0131 | 0.0099 | 0.0242 | 0.0451 | 0.0579 | 0.0298 | 0.0134 |
| | | 3 | -0.0070 | 0.9977 | 0.0094 | 0.0242 | 0.0496 | 0.0506 | 0.0261 | 0.0112 |
| 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 | 0.0455 | 1.0044 | 0.0093 | 0.0238 | 0.0467 | 0.0531 | 0.0282 | 0.0130 |
| | | 2 | 0.0399 | 1.0202 | 0.0100 | 0.0243 | 0.0477 | 0.0594 | 0.0325 | 0.0137 |
| | | 3 | 0.0182 | 1.0088 | 0.0099 | 0.0230 | 0.0464 | 0.0557 | 0.0301 | 0.0116 |

Note: (1) $X_1$ and $X_2$ are generated from MRSAR-B 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 $k = n^{0.5}$

| | | | | | Empirical Tail Probabilities: L=left, R=right | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | stat | dgp | Mean | SD | 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 | 0.0789 | 1.0401 | 0.0102 | 0.0244 | 0.0463 | 0.0651 | 0.0384 | 0.0185 |
| | | 2 | 0.0332 | 1.0210 | 0.0125 | 0.0260 | 0.0467 | 0.0553 | 0.0303 | 0.0144 |
| | | 3 | 0.0275 | 0.9763 | 0.0102 | 0.0238 | 0.0422 | 0.0497 | 0.0260 | 0.0122 |
| 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 | 0.0638 | 1.0149 | 0.0094 | 0.0234 | 0.0479 | 0.0560 | 0.0304 | 0.0137 |
| | | 2 | 0.0590 | 1.0145 | 0.0085 | 0.0220 | 0.0435 | 0.0627 | 0.0332 | 0.0146 |
| | | 3 | 0.0301 | 0.9899 | 0.0099 | 0.0226 | 0.0440 | 0.0525 | 0.0273 | 0.0112 |
| 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 | 0.0437 | 1.0078 | 0.0084 | 0.0215 | 0.0448 | 0.0580 | 0.0303 | 0.0121 |
| | | 2 | 0.0475 | 1.0119 | 0.0101 | 0.0228 | 0.0464 | 0.0574 | 0.0306 | 0.0123 |
| | | 3 | 0.0346 | 0.9970 | 0.0083 | 0.0201 | 0.0465 | 0.0559 | 0.0310 | 0.0135 |

Note: (1) $X_1$ and $X_2$ are generated from `MRSAR-B` scheme, $\sigma = 1$, 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.