

On Skewness and Kurtosis of Econometric Estimators*

Yong Bao[†]
Department of Economics
Purdue University

Aman Ullah[‡]
Department of Economics
University of California, Riverside

July 2008

ABSTRACT

We derive the approximate results for two standardized measures of deviation from normality, namely, the skewness and excess kurtosis coefficients, for a class of econometric estimators. The results are built on a stochastic expansion of the moment condition used to identify the econometric estimator. The approximate results can be used not only to study the finite sample behavior of a particular estimator, but also to compare the finite sample properties of two asymptotically equivalent estimators. We apply the approximate results to the spatial autoregressive model and find that our results approximate the nonnormal behaviors of the maximum likelihood estimator reasonably well. However, when the weights matrix becomes denser, the finite sample distribution of the maximum likelihood estimator departs more severely from normality and our results provide less accurate approximation.

Key Words: skewness; kurtosis; stochastic expansion

JEL Classification: C10, C21

*We would like to thank the co-editor Oliver Linton, two anonymous referees, Prasad Bidarkota, Mohitosh Kejriwal, Jungmin Lee, Peter Thompson, Gautam Tripathi, as well as seminar participants at U Conn, UT Dallas, Florida International and Purdue for very helpful comments.

[†]Corresponding Author: Department of Economics, Krannert School of Management, Purdue University, 403 W. State Street, West Lafayette, IN 47907, (765) 496-2313, Fax: (765) 496-1778, e-mail: ybao@purdue.edu.

[‡]Department of Economics, University of California, Riverside, CA 92521, (951) 827-1591, Fax: (951) 787-5685, e-mail: aman.ullah@ucr.edu.

1 Introduction

Classical statistics and econometrics theory typically relies on asymptotic results for the purpose of estimation and inference. Thanks to various versions of central limit theorems, a typical econometric estimator can be shown to be asymptotically normal and based on this confidence intervals can be constructed and test statistics can be designed. However, in many situations, the asymptotic properties of estimators and test statistics can provide poor approximations of their behavior in finite samples or even moderately large samples. This has been long recognized in the literature with the earliest work dating back to Fisher (1921), while the monograph of Ullah (2004) provides a comprehensive and up-to-date discussion of finite sample econometrics. Usually, the exact results are complicated to analyze and are available only under very restrictive assumptions on the data-generating process. In light of this, approximate techniques have enjoyed popularity, including the large- n , small- σ , Laplace, and saddle-point approximations. We notice that while most of the existing literature focuses on some specific estimators for some specific models, Bao and Ullah (2007b) generalized Rilstone et al. (1996) to develop the approximate first two moments of a large class of estimators in time-series models. An unfinished task in Bao and Ullah (2007b) remains, however, regarding the higher moments of the estimators.

In principle, we can follow the approach of Bao and Ullah (2007b) to expand to a higher order the inverse of the gradient of the moment function in the spirit of Nagar (1959) to derive the approximate third and fourth moments of the estimators, say, approximate the third moment up to order $O(n^{-3})$ and the fourth moment up to order $O(n^{-4})$, where n is the sample size, in the “second-order” sense, see Ullah (2004). However, the “raw” third and fourth moments of an estimator are the absolute measures of skewness and tail behavior of its distribution, relative measures such as the skewness and excess kurtosis *coefficients* should be more useful in situations when we need to judge how the finite sample properties of this estimator can behave differently from the asymptotic properties. Moreover, even though in principle we can derive the third moment up to order $O(n^{-3})$ and the fourth moment up to order $O(n^{-4})$, the terms that are of order $O(n^{-3})$ and $O(n^{-4})$ may be of very small magnitude for a given n and thereby one may wonder how useful they are to be included into the analysis.

The major purpose of this paper is to derive approximate results for two standardized measures of deviation from normality for the estimator, namely, the skewness and excess kurtosis coefficients. Given the knowledge of the nonnormality coefficients, one can not only judge the finite sample behavior of a particular

estimator, but also compare the finite sample properties of two asymptotically equivalent estimators. As an application, we study the finite sample properties of the maximum likelihood estimator (MLE) in the spatial autoregressive model. We find that the departure from normality of the MLE can be very severe in small samples. In these cases, our approximate skewness and kurtosis results can sometimes provide poor approximation to the true tail behaviors of the MLE when the true parameter is approaching its boundary value and the weights matrix is dense. As the sample size increases, the performance of our approximation results improves, especially when the weights matrix is sparse.

The plan of this paper is as follows. In Section 2, we derive our main results. Section 3 gives the application. Section 4 contains some concluding remarks. The appendix collects some technical details in deriving our results and their application in the spatial model.

2 Main Results

We follow Bao and Ullah (2007b) to consider a class of estimators identified by the moment condition

$$\hat{\beta}_n = \arg \{ \psi_n(\beta) = 0 \}, \quad (1)$$

where $\psi_n(\beta) = \psi_n(Z; \beta)$ is a known $k \times 1$ vector-valued function of the observable data $Z = \{Z_i\}_{i=1}^n$, and a parameter vector β , with true value β_0 , of k elements (of the same dimension as $\psi_n(\beta)$) such that $E[\psi_n(\beta)] = 0$ only happens at $\beta = \beta_0$. The type of estimators identified by (1) are general enough to include the maximum likelihood, least squares, method of moments, generalized method of moments, and other extremum estimators, as shown in Rilstone et al. (1996). Usually, the moment condition (1) can be interpreted as the orthogonality condition between regressors and error terms, or as the first-order condition of some optimization criterion. In what follows, we use $\text{tr}A$ to denote the trace, $\|A\|$ to denote the usual norm $(\text{tr}AA')^{1/2}$, and $\nabla^s A(\beta)$ is the matrix of s th order partial derivative of $A(\beta)$ and is obtained recursively (specifically, if $A(\beta)$ is a $k \times 1$ vector function, the j th element of the l th row of $\nabla^s A(\beta)$ (a $k \times k^s$ matrix) is the $1 \times k$ vector $a_{lj}^s(\beta) = \partial a_{lj}^{s-1}(\beta) / \partial \beta^l$). Throughout, the following assumptions are made.

Assumption 1: $\hat{\beta}_n$ exists, and $\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{d} N(0, D)$, where $D = \text{avar}(\sqrt{n}(\hat{\beta}_n - \beta_0)) = O(1)$ is the asymptotic variance of $\sqrt{n}(\hat{\beta}_n - \beta_0)$.

Assumption 2: The first four moments of $\hat{\beta}_n$ exist and are bounded, and for each element of $\hat{\beta}_n$, its r th cumulant is of order $O(n^{(2-r)/2})$, $r = 3, 4$.

Assumption 3: The s th order derivatives of $\psi_n(\beta)$ exist for β in a neighborhood of β_0 and for s up to 3, $E(\|\nabla^s \psi_n(\beta_0)\|^2) < \infty$.

Assumption 4: For β in some neighborhood of β_0 , $[\nabla \psi_n(\beta)]^{-1} = O_p(1)$.

Assumption 5: $\|\nabla^s \psi_n(\beta) - \nabla^s \psi_n(\beta_0)\| \leq \|\beta - \beta_0\| M_n$ for β in some neighborhood of β_0 , where $E(|M_n|) < C < \infty$ for some positive constant C , for s up to 3.

Assumptions 1-5 are fairly standard for a large class of estimators, though excluding nonstationary time-series models involving a unit root. One may lay out a set of primitive conditions to guarantee existence and consistency of $\hat{\beta}_n$. Assumption 2 requires that the first four moments of the estimator exist, which may be regarded as somewhat strong. In general, existence of moments may be difficult to verify. In that case, the results to be derived in this paper may still be informative, though it should then be noted the results are based on “formal” expansions, perhaps not valid expansions. For example, it is well known that moments of the instrumental estimator for a just identified structural equation do not exist. However, it does have a well-defined exact distribution and limiting distribution and the “approximate moments” can still be obtained. Assumption 2 essentially follows Sargan (1974), who showed that for the Nagar-type (Nagar, 1959) large- n approximate moments, which we shall use later to derive our main results, to be valid as asymptotic approximations, the corresponding moments of the exact distribution (of the standardized estimator) must exist and are of order $O(1)$ as $n \rightarrow \infty$, also see Srinivasan (1970) and Basemann (1974). For each element of $\hat{\beta}_n$, the third and fourth cumulants are nothing but the skewness and (excess) kurtosis coefficients of the estimator. Assumptions 3-5 are similar to Rilstone et al. (1996) and Bao and Ullah (2007b) to guarantee that the moment condition is smooth enough so that a stochastic expansion of $\psi_n(\hat{\beta})$ as well as a Nagar-type expansion of the inverse of the gradient of the moment function can be implemented around β_0 .

Following the notational conventions in Rilstone et al. (1996), $\psi_n = \psi_n(\beta_0)$ (we suppress the argument of a function when it is evaluated at β_0), $H_i = \nabla^i \psi_n$, $Q = [E(H_1)]^{-1}$, $V_i = H_i - E(H_i)$, \otimes represent the Kronecker product, $a_{-s/2}$ represent terms of order $O_P(n^{-s/2})$, and put

$$\begin{aligned} a_{-1/2} &= -Q\psi_n, \\ a_{-1} &= -QV_1 a_{-1/2} - \frac{1}{2}QE(H_2)(a_{-1/2} \otimes a_{-1/2}), \\ a_{-3/2} &= -QV_1 a_{-1} - \frac{1}{2}QV_2(a_{-1/2} \otimes a_{-1/2}) - \frac{1}{2}QE(H_2)(a_{-1/2} \otimes a_{-1} + a_{-1} \otimes a_{-1/2}) \\ &\quad - \frac{1}{6}QE(H_3)(a_{-1/2} \otimes a_{-1/2} \otimes a_{-1/2}). \end{aligned}$$

As shown in Rilstone et al. (1996) and Bao and Ullah (2007b), one can write a stochastic expansion

$$\hat{\beta}_n - \beta_0 = a_{-1/2} + a_{-1} + a_{-3/2} + o_P(n^{-3/2}), \quad (2)$$

where the order $O(n^{-1/2})$ term $a_{-1/2}$ represents the asymptotic behavior of $\hat{\beta}_n$ and $D = nE(a_{-1/2}a'_{-1/2})$. Note that $E(a_{-1/2}) = 0$ since $E(\psi_n) = 0$. Based on this and the expansion (2), one can immediately derive the second-order bias and mean squared error (MSE) of $\hat{\beta}_n$, as done in Rilstone et al. (1996) for models with identically and independently distributed (IID) data and Bao and Ullah (2007b) for models with non-IID data. For example, for a single parameter estimator $\hat{\beta}_{n,i}$, $1 \leq i \leq k$, we have the following first two moments:¹

$$\begin{aligned} E(\hat{\beta}_{n,i} - \beta_{0,i}) &= E(a_{-1,i}) + o(n^{-1}), \\ E[(\hat{\beta}_{n,i} - \beta_{0,i})^2] &= E(a_{-1/2,i}^2 + 2a_{-1/2,i}a_{-1,i} + a_{-1,i}^2 + 2a_{-1/2,i}a_{-3/2,i}) + o(n^{-2}). \end{aligned}$$

In principle, one could follow the same strategy to expand $\hat{\beta}_n - \beta_0$ up to order $O(n^{-5/2})$ and derive the second-order third and fourth moments, up to orders $O(n^{-3})$ and $O(n^{-4})$, respectively. However, as stated in the introduction, instead of the absolute measures of skewness and tail behavior of $\hat{\beta}_n$, we are more interested in the relative measures such as the skewness and excess kurtosis coefficients. To facilitate our derivation, we define $T_{n,i} = \sqrt{n}(\hat{\beta}_{n,i} - \beta_{0,i})$. Obviously, the skewness and excess kurtosis coefficients of $\hat{\beta}_{n,i}$ are the same as those of $T_{n,i}$. Corresponding to (2), we write

$$\begin{aligned} T_{n,i} &= \sqrt{n}(a_{-1/2,i} + a_{-1,i} + a_{-3/2,i}) + o_P(n^{-1}) \\ &= \xi_{0,i} + \xi_{-1/2,i} + \xi_{-1,i} + o_P(n^{-1}), \end{aligned} \quad (3)$$

where $\xi_{-s/2,i} = \sqrt{n}a_{-(s+1)/2,i} = O_P(n^{-s/2})$ for $s = 0, 1, 2$. By Assumption 1, $\xi_{0,i} \xrightarrow{d} N(0, D_{ii})$, where D_{ii} denotes the i th element of D . The following theorem gives the approximation skewness and kurtosis results.²

Theorem: The skewness and excess kurtosis coefficients of $\hat{\beta}_{n,i}$ can be approximated by $\gamma_1(\hat{\beta}_{n,i})$ and $\gamma_2(\hat{\beta}_{n,i})$, up to order $O(n^{-1/2})$ and $O(n^{-1})$, respectively, and they are given by

$$\begin{aligned} \gamma_1(\hat{\beta}_{n,i}) &= [E(\xi_{0,i}^2 + 2\xi_{0,i}\xi_{-1/2,i})]^{-3/2} [E(\xi_{0,i}^3 + 3\xi_{0,i}^2\xi_{-1/2,i}) - 3E(\xi_{0,i}^2)E(\xi_{-1/2,i})], \\ \gamma_2(\hat{\beta}_{n,i}) &= \{E(\xi_{0,i}^2 + \xi_{-1/2,i}^2 + 2\xi_{0,i}\xi_{-1/2,i} + 2\xi_{0,i}\xi_{-1,i}) - [E(\xi_{-1/2,i})]^2\}^{-2} \\ &\quad \times \{E(\xi_{0,i}^4 + 4\xi_{0,i}^3\xi_{-1/2,i} + 4\xi_{0,i}^2\xi_{-1,i} + 6\xi_{0,i}^2\xi_{-1/2,i}^2) - 4E(\xi_{0,i}^3 + 3\xi_{0,i}^2\xi_{-1/2,i})E(\xi_{-1/2,i}) \\ &\quad - 4E(\xi_{0,i}^3)E(\xi_{-1,i}) + 6E(\xi_{0,i}^2)[E(\xi_{-1/2,i})]^2\} - 3. \end{aligned} \quad (4)$$

Proof: See the appendix. ■

Note that the skewness result generalizes several expressions that are given in McCullagh (1987) and in Linton (1997) in the context of maximum likelihood estimation. Given the skewness and kurtosis results above, one may follow the lines of Rothenberg (1984) to use the two standardized measures to construct an Edgeworth-type approximation to the distribution of a nonlinear estimator. However, it is still an open question as to whether the Edgeworth distribution is a valid approximation to the true distribution of a general class of (nonlinear) estimators $\hat{\beta}_n$ under the general non-IID setup.

Note that the approximate results in (4) are in terms of expectations of terms involving $\xi_{0,i}$, $\xi_{-1/2,i}$, and $\xi_{-1,i}$. In some cases, these expectations can be worked out explicitly (either analytically or numerically, as demonstrated in the next section). In cases when the expectations are difficult to derive, one may use sample averages to approximate the expectations. In either situation, $\xi_{0,i}$, $\xi_{-1/2,i}$, and $\xi_{-1,i}$ are functions of the unknown $\beta_{0,i}$. In practice, we may have to replace the unknown $\beta_{0,i}$ with its consistent estimator $\hat{\beta}_{n,i}$.

3 Spatial Autoregressive Model

Bao and Ullah (2007a) investigated the finite sample behavior of the MLE of the autoregressive coefficient in a spatial autoregressive model by looking at the second-order bias and MSE. Now we make a more thorough investigation by checking the skewness and kurtosis results.

Consider the following spatial lag model

$$y = \rho_0 W y + \varepsilon, \quad (5)$$

where y is an $n \times 1$ vector of observations on the dependent spatial variable, $W y$ is the corresponding spatially lagged dependent variable for weights matrix W , which is assumed to be known *a priori*, ε is an $n \times 1$ vector of IID Gaussian error terms with zero mean and finite variance σ_0^2 , and ρ_0 is the spatial autoregressive parameter. Under the regularity assumptions in Lee (2004), the average sample likelihood function

$$\mathcal{L}(\rho_0, \sigma_0^2) = \frac{1}{n} \ln |I - \rho_0 W| - \frac{1}{2} \ln (2\pi\sigma_0^2) - \frac{\varepsilon' \varepsilon}{2n\sigma_0^2} \quad (6)$$

is well defined and continuous, and Lee (2004) proved that the MLE has the usual asymptotic properties, including \sqrt{n} -consistency, normality, and asymptotic efficiency. If ρ_0 is known, the MLE of σ_0^2 is given by $\hat{\sigma}_n^2 = (y - \rho_0 W y)'(y - \rho_0 W y)/n = y' C y/n$, where $C = I - \rho_0(W + W') + \rho_0^2 W' W$. Usually, the estimation

procedure is implemented by substituting $\hat{\sigma}_n^2 = y'Cy/n$ into the likelihood function (6) and maximizing a concentrated likelihood function

$$\mathcal{L}(\rho_0) = \frac{1}{n} \sum_{i=1}^n \left[\ln(1 - \rho_0 \omega_i) - \frac{1}{2} \ln \left(\frac{2\pi}{n} y'Cy \right) - \frac{n}{2} \right], \quad (7)$$

where ω_i 's are the eigenvalues of W .

Denote $A = I - \rho_0 W$, $M_i = A^{-1'} [\partial^i (I - \rho_0 (W' + W) + \rho_0^2 W'W) / \partial \rho^i] A^{-1}$, $B_i = \partial^i \ln |A| / \partial \rho_0^i$ (in particular, $B_1 = -\text{tr}(A^{-1}W)$, $B_2 = -\text{tr}[(A^{-1}W)^2]$, $B_3 = -2\text{tr}[(A^{-1}W)^3]$, and $B_4 = -6\text{tr}[(A^{-1}W)^4]$), $b_i = B_i/n$. Then we can write the score function ψ_n , in terms of our notation in Section 2, for the MLE $\hat{\rho}_n$, as well as its higher-order derivatives as follows:

$$\psi_n = b_1 - (\varepsilon'\varepsilon)^{-1} \varepsilon' M_1 \varepsilon / 2,$$

$$H_1 = b_2 - (\varepsilon'\varepsilon)^{-1} \varepsilon' M_2 \varepsilon / 2 + (\varepsilon'\varepsilon)^{-2} (\varepsilon' M_1 \varepsilon)^2 / 2,$$

$$H_2 = b_3 + 3(\varepsilon'\varepsilon)^{-2} \varepsilon' M_1 \varepsilon \varepsilon' M_2 \varepsilon / 2 - (\varepsilon'\varepsilon)^{-3} (\varepsilon' M_1 \varepsilon)^3,$$

$$H_3 = b_4 + 3(\varepsilon'\varepsilon)^{-2} (\varepsilon' M_2 \varepsilon)^2 / 2 - 6(\varepsilon'\varepsilon)^{-3} (\varepsilon' M_1 \varepsilon)^2 \varepsilon' M_2 \varepsilon + 3(\varepsilon'\varepsilon)^{-4} (\varepsilon' M_1 \varepsilon)^4.$$

As it turns out, all the derivatives are in terms of products of ratios of quadratic forms in ε , and the skewness and kurtosis results (4) essentially boil down to expectations of them. Appendix 2 outlines the steps for numerical evaluation and we follow the steps to analyze the skewness and kurtosis behavior of $\hat{\rho}_n$. As in Kelejian and Prucha (1999), we consider three specifications of the weights matrix with different degree of sparseness, namely, the ‘‘one ahead and one behind,’’ ‘‘three ahead and three behind,’’ and ‘‘five ahead and five behind’’ matrices, denoted by $W_{J=2}$, $W_{J=6}$, and $W_{J=10}$, respectively.³ We row-standardize the three matrices and set all the non-zero elements to be equal to each other. We normalize $\sigma_0^2 = 1$. Tables 1-3 give the the theoretical second-order bias (B , up to order $O(n^{-1})$), mean squared error (M , up to order $O(n^{-2})$), the approximate skewness γ_1 (up to order $O(n^{-1/2})$) and excess kurtosis γ_2 (up to order $O(n^{-1})$) of the estimator $\hat{\rho}_n$, as well as the ‘‘true’’ values of them (denoted by Bias, MSE, SK, KR) across 1,000 Monte Carlo replications, for $n = 30, 100, 200$, respectively.

Asymptotically, $\hat{\rho}_n$ should be a normal variable. Bao and Ullah (2007a) documented some evidence indicating that in small samples, when ρ_0 is negatively large and the weights matrix is dense (corresponding to a larger J), the behavior of $\hat{\rho}_n$ can be quite different from what the asymptotic theory predicts by checking the first two moments of $\hat{\rho}_n$. This is supported again by checking the first two moments of $\hat{\rho}_n$. For smaller J , the theoretical bias and MSE results approximate the true bias and MSE quite well.

Of course, the behaviors of the first two moments alone do not necessarily indicate how severe the depar-

ture of the finite sample distribution of $\hat{\rho}_n$ from normality is. More conclusive observations can be possibly made by checking the standardized higher moments of $\hat{\rho}_n$, namely, γ_1 and γ_2 . In general, to approximate higher moments accurately, bigger sample sizes are needed. Moreover, in small samples, $E(\xi_{0,i}^2 + 2\xi_{0,i}\xi_{-1/2,i})$ may be negative, so γ_1 is not defined (this corresponds to the missing values for γ_1 in Table 1). Given this, it is not surprising that when $n = 30$, γ_1 and γ_2 provide in some cases very poor approximations to the true skewness and kurtosis, especially when J is large and/or when ρ_0 is relatively big. Overall, γ_1 seems to provide a better approximation to the skewness coefficient compared with γ_2 (as approximation to the excess kurtosis coefficient) for small J and ρ_0 . Obviously, for a sample of size as small as 30, the behavior of $\hat{\rho}_n$ is quite different from a normal density by looking at SK and KR.

When we move to a sample of bigger size 100, the performance of γ_1 and γ_2 improves significantly. This improvement can also be seen when n goes from 100 to 200. In either case, SK and KR still indicate that the distribution of $\hat{\rho}_n$ is far from being normal. In general, when J is small, both γ_1 and γ_2 provide good approximations to SK and KR. When each spatial unit has more neighbors, however, whereas γ_1 still provides reasonable approximation to the skewness of the distribution of $\hat{\rho}_n$ for moderate ρ_0 , γ_2 proves to approximate the excess kurtosis very poorly, especially for negatively large ρ_0 . In fact, in these cases, the departure of $\hat{\rho}_n$ from normality is most severe, as indicated by SK and KR.

4 Concluding Remarks

We have derived new results on the approximate skewness and excess kurtosis coefficients for a large class of econometric estimators. The knowledge of the two relative measures of departure from normality of econometric estimators may not only enable researchers to judge the finite sample behavior of a particular estimator, but also to compare the finite sample properties of two asymptotically equivalent estimators. Researchers may be tempted to use the two relative measures to construct an Edgeworth-type approximation to the finite sample distribution of the estimator in question. The validity of such an approximation as a distribution function is still an open question to be addressed. In our application, we demonstrate that for the spatial autoregressive model, the departure of the MLE from normality can be quite severe and usually our approximate results capture the true tail behaviors of the MLE quite reasonably well. However, when the departure is most severe, our results do not seem to provide fair approximation in finite samples. As shown in Bao and Ullah (2007a), when a spatial unit is surrounded by many neighbors, the sample estimates of $\hat{\rho}_n$ are quite noisy and we may need really large sample size to achieve convergence. One of the key assumptions

for deriving our large- n approximate results is \sqrt{n} -convergence of the estimator. Therefore more cautions should be called upon to interpret the empirical results and make first-order inferences when we use a dense weights matrix.

Appendix 1: Proof

The standardized i th variable $T_{n,i}^* = [T_{n,i} - E(T_{n,i})]/\sqrt{\text{Var}(T_{n,i})}$ has zero mean and unit variance, and we can easily show that the skewness of $\hat{\beta}_{n,i}$ is equal to $E(T_{n,i}^{*3})$, and its excess kurtosis is equal to $E(T_{n,i}^{*4}) - 3$. By Assumption 2, $E(T_{n,i}^{*3}) = O(n^{-1/2})$ and $E(T_{n,i}^{*4}) - 3 = O(n^{-1})$. Since we do not know $E(T_{n,i})$ and $\text{Var}(T_{n,i})$, we can approximate them to a certain order. Corresponding to the expansion (3),

$$E(T_{n,i}) = u_{i,-1/2} + o(n^{-1/2}), \quad \text{Var}(T_{n,i}) = v_{i,-1/2} + o(n^{-1/2}),$$

where $u_{i,-1/2} = E(\xi_{-1/2,i})$ and $v_{i,-1/2} = E(\xi_{0,i}^2 + 2\xi_{0,i}\xi_{-1/2,i})$ are the approximate mean and variance of $T_{n,i}$, up to $O(n^{-1/2})$. Define the approximate standardized statistic $T_{n,i}^{**} = (T_{n,i} - u_{i,-1/2})/\sqrt{v_{i,-1/2}}$. Obviously, $T_{n,i}^{**} = O_P(1)$. Moreover,

$$\begin{aligned} T_{n,i}^{**} &= \frac{T_{n,i} - E(T_{n,i}) + E(T_{n,i}) - u_{i,-1/2}}{\sqrt{v_{i,-1/2} - \text{Var}(T_{n,i}) + \text{Var}(T_{n,i})}} \\ &= \frac{T_{n,i} - E(T_{n,i})}{\sqrt{\text{Var}(T_{n,i})}} \left[1 + \frac{v_{i,-1/2} - \text{Var}(T_{n,i})}{\text{Var}(T_{n,i})} \right]^{-1/2} + \frac{E(T_{n,i}) - u_{i,-1/2}}{\sqrt{\text{Var}(T_{n,i})}} \left[1 + \frac{v_{i,-1/2} - \text{Var}(T_{n,i})}{\text{Var}(T_{n,i})} \right]^{-1/2} \\ &= T_{n,i}^* \left[1 - \frac{1}{2} \frac{v_{i,-1/2} - \text{Var}(T_{n,i})}{\text{Var}(T_{n,i})} + \dots \right] + \frac{E(T_{n,i}) - u_{i,-1/2}}{\sqrt{\text{Var}(T_{n,i})}} \left[1 - \frac{1}{2} \frac{v_{i,-1/2} - \text{Var}(T_{n,i})}{\text{Var}(T_{n,i})} + \dots \right] \\ &= T_{n,i}^* + o_P(n^{-1/2}) \end{aligned} \tag{8}$$

since $T_{n,i}^* = O_P(1)$, $v_{i,-1/2} - \text{Var}(T_{n,i}) = o(n^{-1/2})$, $u_{i,-1/2} - E(T_{n,i}) = o(n^{-1/2})$, and $\text{Var}(T_{n,i}) = O(1)$. Immediately, $E(T_{n,i}^{**}) = 0 + o(n^{-1/2})$, $\text{Var}(T_{n,i}^{**}) = 1 + o(n^{-1/2})$, and $E(T_{n,i}^{*3}) = E(T_{n,i}^{**3}) + o(n^{-1/2})$, i.e., the third cumulant of $T_{n,i}^*$, or that of $T_{n,i}$, can be approximated by $E(T_{n,i}^{**3})$, up to order $O(n^{-1/2})$. Now we expand $E(T_{n,i}^{**3})$ as follows

$$\begin{aligned} E(T_{n,i}^{**3}) &= v_{i,-1/2}^{-3/2} E[(T_{n,i} - u_{i,-1/2})^3] \\ &= v_{i,-1/2}^{-3/2} [E(T_{n,i}^3) - 3E(T_{n,i}^2)u_{i,-1/2} + 3E(T_{n,i})u_{i,-1/2}^2 - u_{i,-1/2}^3] \\ &= v_{i,-1/2}^{-3/2} [E(\xi_{0,i}^3 + 3\xi_{0,i}^2\xi_{-1/2,i}) - 3E(\xi_{0,i}^2)E(\xi_{-1/2,i})] + o(n^{-1/2}). \end{aligned} \tag{9}$$

Next, to approximate the kurtosis coefficient of $T_{n,i}$, alternatively, the fourth moment of $T_{n,i}^{*4}$, up to order $O(n^{-1})$, we approximate $E(T_{n,i})$ and $Var(T_{n,i})$ in the definition of $T_{n,i}^*$ up to order $O(n^{-1})$,

$$E(T_{n,i}) = u_{i,-1} + o(n^{-1}), \quad Var(T_{n,i}) = v_{i,-1} + o(n^{-1}),$$

where $u_{i,-1} = E(\xi_{-1/2,i} + \xi_{-1,i})$ and $v_{i,-1} = E(\xi_{0,i}^2 + \xi_{-1/2,i}^2 + 2\xi_{0,i}\xi_{-1/2,i} + 2\xi_{0,i}\xi_{-1,i}) - [E(\xi_{-1/2,i})]^2$ are the approximate mean and variance of $T_{n,i}$, up to $O(n^{-1})$. Define the approximate standardized statistic $T_{n,i}^{***} = (T_{n,i} - u_{i,-1})/\sqrt{v_{i,-1}}$. Obviously, $T_{n,i}^{***} = O_P(1)$. Using a similar expansion as (8), we can show $T_{n,i}^{***} = T_{n,i}^* + o(n^{-1})$. Therefore, $T_{n,i}^{***} = 0 + o(n^{-1})$, $Var(T_{n,i}^{***}) = 1 + o(n^{-1})$, and $E(T_{n,i}^{*4}) = E(T_{n,i}^{***4}) + o(n^{-1})$, i.e., the fourth cumulant of $T_{n,i}^*$, or that of $T_{n,i}$, can be approximated by $E(T_{n,i}^{***4}) - 3$, up to order $O(n^{-1})$. Now we expand $E(T_{n,i}^{***4})$ as follows

$$\begin{aligned} E(T_{n,i}^{***4}) &= v_{i,-1}^{-2} E[(T_{n,i} - u_{i,-1})^4] \\ &= v_{i,-1}^{-2} [E(T_{n,i}^4) - 4E(T_{n,i}^3)u_{i,-1} + 6E(T_{n,i}^2)u_{i,-1}^2 - 4E(T_{n,i})u_{i,-1}^3 + u_{i,-1}^4] \\ &= v_{i,-1}^{-2} \{E(\xi_{0,i}^4 + 4\xi_{0,i}^3\xi_{-1/2,i} + 4\xi_{0,i}^2\xi_{-1,i} + 6\xi_{0,i}^2\xi_{-1/2,i}^2 \\ &\quad - 4E(\xi_{0,i}^3 + 3\xi_{0,i}^2\xi_{-1/2,i})E(\xi_{-1/2,i}) - 4E(\xi_{0,i}^3)E(\xi_{-1,i}) + 6E(\xi_{0,i}^2)[E(\xi_{-1/2,i})]^2\} + o(n^{-1}). \end{aligned} \quad (10)$$

The skewness and excess kurtosis follow immediately from (9) and (10). ■

Appendix 2: Spatial Model

Using (2) and (3), we collect the following terms, which are needed in calculating the approximate results in (4) (since we have a scalar parameter, we suppress the subscript i):

$$\begin{aligned} E(\xi_0^2) &= nQ^2E(\psi_n^2), \quad E(\xi_0^3) = -n^{3/2}Q^3E(\psi_n^3), \quad E(\xi_0^4) = n^2Q^4E(\psi_n^4), \\ E(\xi_{-1/2}) &= \sqrt{n}[Q^2E(\psi_n H_1) - \frac{1}{2}Q^3E(H_2)E(\psi_n^2)], \\ E(\xi_0\xi_{-1/2}) &= n[Q^2E(\psi_n^2) - Q^3E(\psi_n^2 H_1) + \frac{1}{2}Q^4E(H_2)E(\psi_n^3)], \\ E(\xi_0^2\xi_{-1/2}) &= -n^{3/2}[Q^3E(\psi_n^3) - Q^4E(\psi_n^3 H_1) + \frac{1}{2}Q^5E(H_2)E(\psi_n^4)], \\ E(\xi_0^3\xi_{-1/2}) &= n^2[Q^4E(\psi_n^4) - Q^5E(\psi_n^4 H_1) + \frac{1}{2}Q^6E(H_2)E(\psi_n^5)], \\ E(\xi_{-1/2}^2) &= n\{Q^2E(\psi_n^2) - 2Q^3E(\psi_n^2 H_1) + Q^4[E(H_2)E(\psi_n^3) + E(\psi_n^2 H_1^2)] - Q^5E(H_2)E(\psi_n^3 H_1) \\ &\quad + \frac{1}{4}Q^6[E(H_2)]^2E(\psi_n^4)\}, \\ E(\xi_0\xi_{-1/2}^2) &= -n^{3/2}\{Q^3E(\psi_n^3) - 2Q^4E(\psi_n^3 H_1) + Q^5[E(H_2)E(\psi_n^4) + E(\psi_n^3 H_1^2)] - Q^6E(H_2)E(\psi_n^4 H_1) \\ &\quad + \frac{1}{4}Q^7[E(H_2)]^2E(\psi_n^5)\}, \\ E(\xi_0^2\xi_{-1/2}^2) &= n^2\{Q^4E(\psi_n^4) - 2Q^5E(\psi_n^4 H_1) + Q^6[E(H_2)E(\psi_n^5) + E(\psi_n^4 H_1^2)] - Q^7E(H_2)E(\psi_n^5 H_1) \end{aligned}$$

$$\begin{aligned}
& +\frac{1}{4}Q^8[E(H_2)]^2E(\psi_n^6)\}, \\
E(\xi_{-1}) &= -\sqrt{n}\{-2Q^2E(\psi_nH_1) + Q^3[E(H_2)E(\psi_n^2) + E(\psi_nH_1^2) + \frac{1}{2}E(\psi_n^2H_2)] \\
& -Q^4[\frac{1}{6}E(H_3)E(\psi_n^3) + \frac{3}{2}E(H_2)E(\psi_n^2H_1)] + \frac{1}{2}Q^5[E(H_2)]^2E(\psi_n^3)\}, \\
E(\xi_0\xi_{-1}) &= n\{Q^2E(\psi_n^2) - 2Q^3E(\psi_n^2H_1) + Q^4[E(H_2)E(\psi_n^3) + E(\psi_n^2H_1^2) + \frac{1}{2}E(\psi_n^3H_2)] \\
& -Q^5[\frac{1}{6}E(H_3)E(\psi_n^4) + \frac{3}{2}E(H_2)E(\psi_n^3H_1)] + \frac{1}{2}Q^6[E(H_2)]^2E(\psi_n^4)\}, \\
E(\xi_0^2\xi_{-1}) &= -n^{3/2}\{Q^3E(\psi_n^3) - 2Q^4E(\psi_n^3H_1) + Q^5[E(H_2)E(\psi_n^4) + E(\psi_n^3H_1^2) + \frac{1}{2}E(\psi_n^4H_2)] \\
& -Q^6[\frac{1}{6}E(H_3)E(\psi_n^5) + \frac{3}{2}E(H_2)E(\psi_n^4H_1)] + \frac{1}{2}Q^7[E(H_2)]^2E(\psi_n^5)\}, \\
E(\xi_0^3\xi_{-1}) &= n^2\{Q^4E(\psi_n^4) - 2Q^5E(\psi_n^4H_1) + Q^6[E(H_2)E(\psi_n^5) + E(\psi_n^4H_1^2) + \frac{1}{2}E(\psi_n^5H_2)] \\
& -Q^7[\frac{1}{6}E(H_3)E(\psi_n^6) + \frac{3}{2}E(H_2)E(\psi_n^5H_1)] + \frac{1}{2}Q^8[E(H_2)]^2E(\psi_n^6)\}.
\end{aligned}$$

To work out the expectations as given above, let $\lambda_{i,j} = E[(\varepsilon'M_1\varepsilon)^i(\varepsilon'M_2\varepsilon)^j / (\varepsilon'\varepsilon)^{i+j}]$, then by substituting ψ_n and H_i , we can write all the expectations in terms of $\lambda_{i,j}$:

$$\begin{aligned}
Q &= (b_2 - \frac{1}{2}\lambda_{0,1} + \frac{1}{2}\lambda_{2,0})^{-1}, \\
E(H_2) &= b_3 + \frac{3}{2}\lambda_{1,1} - \lambda_{3,0}, \\
E(H_3) &= b_4 + \frac{3}{2}\lambda_{0,2} - 6\lambda_{2,1} + 3\lambda_{4,0}, \\
E(\psi_n^2) &= b_1^2 - b_1\lambda_{1,0} + \frac{1}{4}\lambda_{2,0}, \\
E(\psi_n^3) &= b_1^3 - \frac{3}{2}b_1^2\lambda_{1,0} + \frac{3}{4}b_1\lambda_{2,0} - \frac{1}{8}\lambda_{3,0}, \\
E(\psi_n^4) &= b_1^4 - 2b_1^3\lambda_{1,0} + \frac{3}{2}b_1^2\lambda_{2,0} - \frac{1}{2}b_1\lambda_{3,0} + \frac{1}{16}\lambda_{4,0}, \\
E(\psi_n^5) &= b_1^5 - \frac{5}{2}b_1^4\lambda_{1,0} + \frac{5}{2}b_1^3\lambda_{2,0} - \frac{5}{4}b_1^2\lambda_{3,0} + \frac{5}{16}b_1\lambda_{4,0} - \frac{1}{32}\lambda_{5,0}, \\
E(\psi_n^6) &= b_1^6 - 3b_1^5\lambda_{1,0} + \frac{15}{4}b_1^4\lambda_{2,0} - \frac{5}{2}b_1^3\lambda_{3,0} + \frac{15}{16}b_1^2\lambda_{4,0} - \frac{3}{16}b_1\lambda_{5,0} + \frac{1}{64}\lambda_{6,0}, \\
E(\psi_nH_1) &= b_1b_2 - \frac{1}{2}b_1\lambda_{0,1} - \frac{1}{2}b_2\lambda_{1,0} + \frac{1}{4}\lambda_{1,1} + \frac{1}{2}b_1\lambda_{2,0} - \frac{1}{4}\lambda_{3,0}, \\
E(\psi_n^2H_1) &= b_1^2b_2 - \frac{1}{2}b_1^2\lambda_{0,1} - b_1b_2\lambda_{1,0} + \frac{1}{2}b_1\lambda_{1,1} + (\frac{1}{2}b_1^2 + \frac{1}{4}b_2)\lambda_{2,0} - \frac{1}{8}\lambda_{2,1} - \frac{1}{2}b_1\lambda_{3,0} + \frac{1}{8}\lambda_{4,0}, \\
E(\psi_n^3H_1) &= b_1^3b_2 - \frac{1}{2}b_1^3\lambda_{0,1} - \frac{3}{2}b_1^2b_2\lambda_{1,0} + \frac{3}{4}b_1^2\lambda_{1,1} + (\frac{1}{2}b_1^3 + \frac{3}{4}b_1b_2)\lambda_{2,0} - \frac{3}{8}b_1\lambda_{2,1} - (\frac{3}{4}b_1^2 + \frac{1}{8}b_2)\lambda_{3,0} + \frac{1}{16}\lambda_{3,1} \\
& + \frac{3}{8}b_1\lambda_{4,0} - \frac{1}{16}\lambda_{5,0}, \\
E(\psi_n^4H_1) &= b_1^4b_2 - \frac{1}{2}b_1^4\lambda_{0,1} - 2b_1^3b_2\lambda_{1,0} + b_1^3\lambda_{1,1} + (\frac{1}{2}b_1^4 + \frac{3}{2}b_1^2b_2)\lambda_{2,0} - \frac{3}{4}b_1^2\lambda_{2,1} - (b_1^3 + \frac{1}{2}b_1b_2)\lambda_{3,0} + \frac{1}{4}b_1\lambda_{3,1} \\
& + (\frac{3}{4}b_1^2 + \frac{1}{16}b_2)\lambda_{4,0} - \frac{1}{32}\lambda_{4,1} - \frac{1}{4}b_1\lambda_{5,0} + \frac{1}{32}\lambda_{6,0}, \\
E(\psi_n^5H_1) &= b_1^5b_2 - \frac{1}{2}b_1^5\lambda_{0,1} - \frac{5}{2}b_1^4b_2\lambda_{1,0} + \frac{5}{4}b_1^4\lambda_{1,1} + (\frac{1}{2}b_1^5 + \frac{5}{2}b_1^3b_2)\lambda_{2,0} - \frac{5}{4}b_1^3\lambda_{2,1} - (\frac{5}{4}b_1^4 + \frac{5}{4}b_1^2b_2)\lambda_{3,0} + \frac{5}{8}b_1^2\lambda_{3,1} \\
& + (\frac{5}{4}b_1^3 + \frac{5}{16}b_1b_2)\lambda_{4,0} - \frac{5}{32}b_1\lambda_{4,1} - (\frac{5}{8}b_1^2 + \frac{1}{32}b_2)\lambda_{5,0} + \frac{1}{64}\lambda_{5,1} + \frac{5}{32}b_1\lambda_{6,0} - \frac{1}{64}\lambda_{7,0}, \\
E(\psi_nH_1^2) &= b_1b_2^2 - b_1b_2\lambda_{0,1} + \frac{1}{4}b_1\lambda_{0,2} - \frac{1}{2}b_2^2\lambda_{1,0} + \frac{1}{2}b_2\lambda_{1,1} - \frac{1}{8}\lambda_{1,2} + b_1b_2\lambda_{2,0} - \frac{1}{2}b_1\lambda_{2,1} - \frac{1}{2}b_2\lambda_{3,0} + \frac{1}{4}\lambda_{3,1} \\
& + \frac{1}{4}b_1\lambda_{4,0} - \frac{1}{8}\lambda_{5,0}, \\
E(\psi_n^2H_1^2) &= b_1^2b_2^2 - b_1^2b_2\lambda_{0,1} + \frac{1}{4}b_1^2\lambda_{0,2} - b_1b_2^2\lambda_{1,0} + b_1b_2\lambda_{1,1} - \frac{1}{4}b_1\lambda_{1,2} + (b_1^2b_2 + \frac{1}{4}b_2^2)\lambda_{2,0} - (\frac{1}{2}b_1^2 + \frac{1}{4}b_2)\lambda_{2,1} + \frac{\lambda_{2,2}}{16}
\end{aligned}$$

$$\begin{aligned}
& -b_1 b_2 \lambda_{3,0} + \frac{1}{2} b_1 \lambda_{3,1} + \left(\frac{1}{4} b_1^2 + \frac{1}{4} b_2\right) \lambda_{4,0} - \frac{1}{8} \lambda_{4,1} - \frac{1}{4} b_1 \lambda_{5,0} + \frac{\lambda_{6,0}}{16}, \\
E(\psi_n^3 H_1^2) &= b_1^3 b_2^2 - b_1^3 b_2 \lambda_{0,1} + \frac{1}{4} b_1^3 \lambda_{0,2} - \frac{3}{2} b_1^2 b_2^2 \lambda_{1,0} + \frac{3}{2} b_1^2 b_2 \lambda_{1,1} - \frac{3}{8} b_1^2 \lambda_{1,2} + (b_1^3 b_2 + \frac{3}{4} b_1 b_2^2) \lambda_{2,0} - (\frac{1}{2} b_1^3 + \frac{3}{4} b_1 b_2) \lambda_{2,1} \\
&+ \frac{3}{16} b_1 \lambda_{2,2} - (\frac{3}{2} b_1^2 b_2 + \frac{1}{8} b_2^2) \lambda_{3,0} + (\frac{3}{4} b_1^2 + \frac{1}{8} b_2) \lambda_{3,1} - \frac{1}{32} \lambda_{3,2} + (\frac{1}{4} b_1^3 + \frac{3}{4} b_1 b_2) \lambda_{4,0} - \frac{3}{8} b_1 \lambda_{4,1} - (\frac{3}{8} b_1^2 + \frac{1}{8} b_2) \lambda_{5,0} \\
&+ \frac{1}{16} \lambda_{5,1} + \frac{3}{16} b_1 \lambda_{6,0} - \frac{1}{32} \lambda_{7,0}, \\
E(\psi_n^4 H_1^2) &= b_1^4 b_2^2 - b_1^4 b_2 \lambda_{0,1} + \frac{1}{4} b_1^4 \lambda_{0,2} - 2b_1^3 b_2^2 \lambda_{1,0} + 2b_1^3 b_2 \lambda_{1,1} - \frac{1}{2} b_1^3 \lambda_{1,2} + (b_1^4 b_2 + \frac{3}{2} b_1^2 b_2^2) \lambda_{2,0} - (\frac{1}{2} b_1^4 + \frac{3}{2} b_1^2 b_2) \lambda_{2,1} \\
&+ \frac{3}{8} b_1^2 \lambda_{2,2} - (2b_1^3 b_2 + \frac{1}{2} b_1 b_2^2) \lambda_{3,0} + (b_1^3 + \frac{1}{2} b_1 b_2) \lambda_{3,1} - \frac{1}{8} b_1 \lambda_{3,2} + (\frac{1}{4} b_1^4 + \frac{3}{2} b_1^2 b_2 + \frac{1}{16} b_2^2) \lambda_{4,0} - (\frac{3}{4} b_1^2 + \frac{1}{16} b_2) \lambda_{4,1} \\
&+ \frac{1}{64} \lambda_{4,2} - (\frac{1}{2} b_1^3 + \frac{1}{2} b_1 b_2) \lambda_{5,0} + \frac{1}{4} b_1 \lambda_{5,1} + (\frac{3}{8} b_1^2 + \frac{1}{16} b_2) \lambda_{6,0} - \frac{1}{32} \lambda_{6,1} - \frac{1}{8} b_1 \lambda_{7,0} + \frac{1}{64} \lambda_{8,0}, \\
E(\psi_n^2 H_2) &= b_1^2 b_3 - b_1 b_3 \lambda_{1,0} + \frac{3}{2} b_1^2 \lambda_{1,1} + \frac{1}{4} b_3 \lambda_{2,0} - \frac{3}{2} b_1 \lambda_{2,1} - b_1^2 \lambda_{3,0} + \frac{3}{8} \lambda_{3,1} + b_1 \lambda_{4,0} - \frac{1}{4} \lambda_{5,0}, \\
E(\psi_n^3 H_2) &= b_1^3 b_3 - \frac{3}{2} b_1^2 b_3 \lambda_{1,0} + \frac{3}{2} b_1^3 \lambda_{1,1} + \frac{3}{4} b_1 b_3 \lambda_{2,0} - \frac{9}{4} b_1^2 \lambda_{2,1} - (b_1^3 + \frac{1}{8} b_3) \lambda_{3,0} + \frac{9}{8} b_1 \lambda_{3,1} + \frac{3}{2} b_1^2 \lambda_{4,0} - \frac{3}{16} \lambda_{4,1} \\
&- \frac{3}{4} b_1 \lambda_{5,0} + \frac{1}{8} \lambda_{6,0}, \\
E(\psi_n^4 H_2) &= b_1^4 b_3 - 2b_1^3 b_3 \lambda_{1,0} + \frac{3}{2} b_1^4 \lambda_{1,1} + \frac{3}{2} b_1^3 b_3 \lambda_{2,0} - 3b_1^3 \lambda_{2,1} - (b_1^4 + \frac{1}{2} b_1 b_3) \lambda_{3,0} + \frac{9}{4} b_1^2 \lambda_{3,1} + (2b_1^3 + \frac{1}{16} b_3) \lambda_{4,0} \\
&- \frac{3}{4} b_1 \lambda_{4,1} - \frac{3}{2} b_1^2 \lambda_{5,0} + \frac{3}{32} \lambda_{5,1} + \frac{1}{2} b_1 \lambda_{6,0} - \frac{1}{16} \lambda_{7,0}, \\
E(\psi_n^5 H_2) &= b_1^5 b_3 - \frac{5}{2} b_1^4 b_3 \lambda_{1,0} + \frac{3}{2} b_1^5 \lambda_{1,1} + \frac{5}{2} b_1^3 b_3 \lambda_{2,0} - \frac{15}{4} b_1^4 \lambda_{2,1} - (b_1^5 + \frac{5}{4} b_1^2 b_3) \lambda_{3,0} + \frac{15}{4} b_1^3 \lambda_{3,1} + (\frac{5}{2} b_1^4 + \frac{5}{16} b_1 b_3) \lambda_{4,0} \\
&- \frac{15}{8} b_1^2 \lambda_{4,1} - (\frac{5}{2} b_1^3 + \frac{1}{32} b_3) \lambda_{5,0} + \frac{15}{32} b_1 \lambda_{5,1} + \frac{5}{4} b_1^2 \lambda_{6,0} - \frac{3}{64} \lambda_{6,1} - \frac{5}{16} b_1 \lambda_{7,0} + \frac{1}{32} \lambda_{8,0}.
\end{aligned}$$

So we need to evaluate $\lambda_{i,j}$, moments of cross products of ratios of quadratic forms in the normal vector $\varepsilon \sim N(0, \sigma_0^2 I)$. Replacing ε with ε/σ_0 in the definition of $\lambda_{i,j}$ does not change the expectations, so we rewrite $\lambda_{i,j} = E[(\varepsilon' M_1 \varepsilon)^i (\varepsilon' M_2 \varepsilon)^j / (\varepsilon' \varepsilon)^{i+j}]$, where $\varepsilon \sim N(0, I)$. Since M_1 and M_2 are symmetric and (trivially) both are commutative with I (the matrix in the quadratic form for the denominator), we can use immediately the separation result from Bao and Ullah (2007c):

$$E \left[\frac{(\varepsilon' M_1 \varepsilon)^i (\varepsilon' M_2 \varepsilon)^j}{(\varepsilon' \varepsilon)^{i+j}} \right] = \frac{E[(\varepsilon' M_1 \varepsilon)^i (\varepsilon' M_2 \varepsilon)^j]}{E[(\varepsilon' \varepsilon)^{i+j}]} \quad (11)$$

We can easily verify $E[(\varepsilon' \varepsilon)^{i+j}] = n_{(i+j)} = \prod_{k=0}^{i+j-1} (n+2k)$. As for the numerator $E[(\varepsilon' M_1 \varepsilon)^i (\varepsilon' M_2 \varepsilon)^j]$, moments of products of quadratic forms, we can utilize the recursive algorithm in Ghazal (1996) and its generalization in Bao and Ullah (2007c): for $\varepsilon \sim N(0, I)$ and symmetric matrices A_i ,

$$E\left(\prod_{i=1}^q \varepsilon' A_i \varepsilon\right) = E(\varepsilon' A_1 \varepsilon) \cdot E\left(\prod_{i=2}^q \varepsilon' A_i \varepsilon\right) + 2 \sum_{j=2}^q E(\varepsilon' A_j A_1 \varepsilon \cdot \varepsilon' A_2 \varepsilon \cdots \varepsilon' A_{j-1} \varepsilon \cdot \prod_{k=j+1}^q \varepsilon' A_k \varepsilon). \quad (12)$$

Given the separation result (11) and the recursive algorithm (12), we collect in the following the exact expressions of $\lambda_{i,j}$, in terms of products of traces of matrices involving M_1 and M_2 :⁴

$$n_{(1)} \lambda_{0,1} = \text{tr} M_2,$$

$$n_{(1)} \lambda_{1,0} = \text{tr} M_1,$$

$$\begin{aligned}
n_{(2)}\lambda_{2,0} &= 2\text{tr}M_1^2 + (\text{tr}M_1)^2, \\
n_{(2)}\lambda_{0,2} &= 2\text{tr}M_2^2 + (\text{tr}M_2)^2, \\
n_{(3)}\lambda_{3,0} &= 8\text{tr}M_1^3 + 6\text{tr}M_1 \cdot \text{tr}M_1^2 + (\text{tr}M_1)^3, \\
n_{(4)}\lambda_{4,0} &= 48\text{tr}M_1^4 + 32\text{tr}M_1 \cdot \text{tr}M_1^3 + 12\text{tr}M_1^2 \cdot (\text{tr}M_1)^2 + 12(\text{tr}M_1^2)^2 + (\text{tr}M_1)^4, \\
n_{(5)}\lambda_{5,0} &= 384\text{tr}M_1^5 + 240\text{tr}M_1 \cdot \text{tr}M_1^4 + 160\text{tr}M_1^2 \cdot \text{tr}M_1^3 + 80(\text{tr}M_1)^2 \cdot \text{tr}M_1^3 + 60\text{tr}M_1 \cdot (\text{tr}M_1^2)^2 \\
&\quad + 20\text{tr}M_1^2 \cdot (\text{tr}M_1)^3 + (\text{tr}M_1)^5, \\
n_{(6)}\lambda_{6,0} &= 3840\text{tr}M_1^6 + 2304\text{tr}M_1 \cdot \text{tr}M_1^5 + 1440\text{tr}M_1^2 \cdot \text{tr}M_1^4 + 960\text{tr}M_1 \cdot \text{tr}M_1^2 \cdot \text{tr}M_1^3 + 720(\text{tr}M_1)^2 \cdot \text{tr}M_1^4 \\
&\quad + 640(\text{tr}M_1^3)^2 + 180(\text{tr}M_1)^2 \cdot (\text{tr}M_1^2)^2 + 160(\text{tr}M_1)^3 \cdot \text{tr}M_1^3 + 120(\text{tr}M_1^2)^3 + 30(\text{tr}M_1)^4 \cdot \text{tr}M_1^2 \\
&\quad + (\text{tr}M_1)^6, \\
n_{(7)}\lambda_{7,0} &= 46080\text{tr}M_1^7 + 26880\text{tr}M_1 \cdot \text{tr}M_1^6 + 16128\text{tr}M_1^2 \cdot \text{tr}M_1^5 + 13440\text{tr}M_1^3 \cdot \text{tr}M_1^4 + 10080\text{tr}M_1 \cdot \text{tr}M_1^2 \cdot \text{tr}M_1^4 \\
&\quad + 8064(\text{tr}M_1)^2 \cdot \text{tr}M_1^5 + 4480\text{tr}M_1 \cdot (\text{tr}M_1^3)^2 + 3360(\text{tr}M_1^2)^2 \cdot \text{tr}M_1^3 + 1680(\text{tr}M_1)^3 \cdot \text{tr}M_1^4 + 960(\text{tr}M_1^2)^2 \cdot \text{tr}M_1^3 \\
&\quad + 840\text{tr}M_1 \cdot (\text{tr}M_1^2)^3 + 420(\text{tr}M_1)^3 \cdot (\text{tr}M_1^2)^2 + 280(\text{tr}M_1)^4 \cdot \text{tr}M_1^3 + 42(\text{tr}M_1)^5 \cdot \text{tr}M_1^2 + (\text{tr}M_1)^7, \\
n_{(8)}\lambda_{8,0} &= 645120\text{tr}M_1^8 + 368640\text{tr}M_1 \cdot \text{tr}M_1^7 + 215040\text{tr}M_1^2 \cdot \text{tr}M_1^6 + 172032\text{tr}M_1^3 \cdot \text{tr}M_1^5 + 129024\text{tr}M_1 \cdot \text{tr}M_1^2 \cdot \text{tr}M_1^5 \\
&\quad + 107520(\text{tr}M_1)^2 \cdot \text{tr}M_1^6 + 107520\text{tr}M_1 \cdot \text{tr}M_1^3 \cdot \text{tr}M_1^4 + 80640(\text{tr}M_1^4)^2 + 40320(\text{tr}M_1)^2 \cdot \text{tr}M_1^2 \cdot \text{tr}M_1^4 \\
&\quad + 24480\text{tr}M_1 \cdot (\text{tr}M_1^2)^2 \cdot \text{tr}M_1^3 + 23520(\text{tr}M_1^2)^2 \cdot \text{tr}M_1^4 + 21504(\text{tr}M_1)^3 \cdot \text{tr}M_1^5 + 19040\text{tr}M_1^2 \cdot (\text{tr}M_1^3)^2 \\
&\quad + 17920(\text{tr}M_1)^2 \cdot (\text{tr}M_1^3)^2 + 8960(\text{tr}M_1)^3 \cdot \text{tr}M_1^2 \cdot \text{tr}M_1^3 + 3360(\text{tr}M_1)^4 \cdot \text{tr}M_1^4 + 3360(\text{tr}M_1)^2 \cdot (\text{tr}M_1^3)^3 \\
&\quad + 1680(\text{tr}M_1^2)^4 + 840(\text{tr}M_1)^4 \cdot (\text{tr}M_1^2)^2 + 448(\text{tr}M_1)^5 \cdot \text{tr}M_1^3 + 56(\text{tr}M_1)^6 \cdot \text{tr}M_1^2 (\text{tr}M_1)^8, \\
n_{(2)}\lambda_{1,1} &= 2\text{tr}M_1 M_2 + \text{tr}M_1 \cdot \text{tr}M_2, \\
n_{(3)}\lambda_{1,2} &= 8\text{tr}M_1 M_2^2 + 4\text{tr}M_1 M_2 \cdot \text{tr}M_2 + 2\text{tr}M_1 \cdot \text{tr}M_2^2 + \text{tr}M_1 \cdot (\text{tr}M_2)^2, \\
n_{(3)}\lambda_{2,1} &= 8\text{tr}M_1^2 M_2 + 4\text{tr}M_1 M_2 \cdot \text{tr}M_1 + 2\text{tr}M_2 \cdot \text{tr}M_1^2 + \text{tr}M_2 \cdot (\text{tr}M_1)^2, \\
n_{(4)}\lambda_{3,1} &= 48\text{tr}M_1^3 M_2 + 24\text{tr}M_1^2 M_2 \cdot \text{tr}M_1 + 12\text{tr}M_1 M_2 \cdot \text{tr}M_1^2 + 8\text{tr}M_1^3 \cdot \text{tr}M_2 + 6\text{tr}M_1 \cdot \text{tr}M_2 \cdot \text{tr}M_1^2 \\
&\quad + 6\text{tr}M_1 M_2 \cdot (\text{tr}M_1)^2 + (\text{tr}M_1)^3 \cdot \text{tr}M_2, \\
n_{(4)}\lambda_{2,2} &= 32\text{tr}M_1^2 M_2^2 + 16\text{tr}M_1 \cdot \text{tr}M_1 M_2^2 + 16\text{tr}M_2 \cdot \text{tr}M_1^2 M_2 + 16\text{tr}M_1 M_2 M_1 M_2 + 8\text{tr}M_1 \cdot \text{tr}M_2 \cdot \text{tr}M_1 M_2 \\
&\quad + 8(\text{tr}M_1 M_2)^2 + 4\text{tr}M_1^2 \cdot \text{tr}M_2^2 + 2\text{tr}M_1^2 \cdot (\text{tr}M_2)^2 + 2\text{tr}M_2^2 \cdot (\text{tr}M_1)^2 + (\text{tr}M_1)^2 \cdot (\text{tr}M_2)^2, \\
n_{(5)}\lambda_{4,1} &= 384\text{tr}M_1^4 M_2 + 192\text{tr}M_1 \cdot \text{tr}M_1^3 M_2 + 96\text{tr}M_1^2 M_2 \cdot \text{tr}M_1^2 + 64\text{tr}M_1 M_2 \cdot \text{tr}M_1^3 + 48\text{tr}M_1^4 \cdot \text{tr}M_2 \\
&\quad + 48\text{tr}M_1^2 M_2 \cdot (\text{tr}M_1)^2 + 48\text{tr}M_1 \cdot \text{tr}M_1^2 \cdot \text{tr}M_1 M_2 + 32\text{tr}M_1 \cdot \text{tr}M_1^3 \cdot \text{tr}M_2 + 12(\text{tr}M_1)^2 \cdot \text{tr}M_1^2 \cdot \text{tr}M_2 \\
&\quad + 12(\text{tr}M_1^2)^2 \cdot \text{tr}M_2 + 8(\text{tr}M_1)^3 \cdot \text{tr}M_1 M_2 + (\text{tr}M_1)^4 \cdot \text{tr}M_2, \\
n_{(5)}\lambda_{3,2} &= 192\text{tr}M_1^3 M_2^2 + 192\text{tr}M_1 M_2 M_1^2 M_2 + 96\text{tr}M_1 \cdot \text{tr}M_1^2 M_2^2 + 96\text{tr}M_1^2 M_2 \cdot \text{tr}M_1 M_2 + 96\text{tr}M_1^3 M_2 \cdot \text{tr}M_2 \\
&\quad + 48\text{tr}M_1 \cdot \text{tr}M_1 M_2 M_1 M_2 + 48\text{tr}M_1^2 \cdot \text{tr}M_1 M_2^2 + 48\text{tr}M_1 \cdot \text{tr}M_1^2 M_2 \cdot \text{tr}M_2 + 24(\text{tr}M_1)^2 \cdot \text{tr}M_1 M_2^2 \\
&\quad + 24\text{tr}M_1^2 \cdot \text{tr}M_1 M_2 \cdot \text{tr}M_2 + 24\text{tr}M_1 \cdot (\text{tr}M_1 M_2)^2 + 16\text{tr}M_1^3 \cdot \text{tr}M_2^2 + 12(\text{tr}M_1)^2 \cdot \text{tr}M_1 M_2 \cdot \text{tr}M_2
\end{aligned}$$

$$\begin{aligned}
& +12\text{tr}M_1 \cdot \text{tr}M_1^2 \cdot \text{tr}M_2^2 + 8\text{tr}M_1^3 \cdot (\text{tr}M_2)^2 + 6\text{tr}M_1 \cdot \text{tr}M_1^2 \cdot (\text{tr}M_2)^2 + 2(\text{tr}M_1)^3 \cdot \text{tr}M_2^2 + (\text{tr}M_1)^3 (\text{tr}M_2)^2, \\
n_{(6)}\lambda_{4,2} = & 1536\text{tr}M_1^4 M_2^2 + 1536\text{tr}M_1 M_2 M_1^3 M_2 + 768\text{tr}M_1 \cdot \text{tr}M_1^3 M_2^2 + 768\text{tr}M_1^2 M_2 M_1^2 M_2 + 768\text{tr}M_1^3 M_2 \cdot \text{tr}M_1 M_2 \\
& + 768\text{tr}M_1^4 M_2 \cdot \text{tr}M_2 + 768\text{tr}M_1 \cdot \text{tr}M_1 M_2 M_1^2 M_2 + 384(\text{tr}M_1^2 M_2)^2 + 384\text{tr}M_1 \cdot \text{tr}M_1^2 M_2 \cdot \text{tr}M_1 M_2 \\
& + 384\text{tr}M_1 \cdot \text{tr}M_1^3 M_2 \cdot \text{tr}M_2 + 384\text{tr}M_1^2 \cdot \text{tr}M_1^2 M_2^2 + 256\text{tr}M_1^3 \cdot \text{tr}M_1 M_2^2 + 192(\text{tr}M_1)^2 \cdot \text{tr}M_1^2 M_2^2 \\
& + 192\text{tr}M_1^2 \cdot \text{tr}M_1 M_2 M_1 M_2 + 192\text{tr}M_1 \cdot \text{tr}M_1^2 \cdot \text{tr}M_1 M_2^2 + 192\text{tr}M_1^2 \cdot \text{tr}M_1^2 M_2 \cdot \text{tr}M_2 + 128\text{tr}M_1^3 \cdot \text{tr}M_1 M_2 \cdot \text{tr}M_2 \\
& + 96\text{tr}M_1^2 \cdot (\text{tr}M_1 M_2)^2 + 96(\text{tr}M_1)^2 \cdot \text{tr}M_1 M_2 M_1 M_2 + 96(\text{tr}M_1)^2 \cdot \text{tr}M_1^2 M_2 \cdot \text{tr}M_2 + 96\text{tr}M_1 \cdot \text{tr}M_1^2 \cdot \text{tr}M_1 M_2 \cdot \text{tr}M_2 \\
& + 96\text{tr}M_1^4 \cdot \text{tr}M_2^2 + 64\text{tr}M_1 \cdot \text{tr}M_1^3 \cdot \text{tr}M_2^2 + 48(\text{tr}M_1)^2 \cdot (\text{tr}M_1 M_2)^2 + 48\text{tr}M_1^4 \cdot (\text{tr}M_2)^2 + 32(\text{tr}M_1)^3 \cdot \text{tr}M_1 M_2^2 \\
& + 32\text{tr}M_1 \cdot \text{tr}M_1^3 \cdot (\text{tr}M_2)^2 + 24(\text{tr}M_1)^2 \cdot \text{tr}M_1^2 \cdot \text{tr}M_2^2 + 24(\text{tr}M_1^2)^2 \cdot \text{tr}M_2^2 + 16(\text{tr}M_1)^3 \cdot \text{tr}M_1 M_2 \cdot \text{tr}M_2 \\
& + 12(\text{tr}M_1)^2 \cdot \text{tr}M_1^2 \cdot (\text{tr}M_2)^2 + 12(\text{tr}M_1^2)^2 \cdot (\text{tr}M_2)^2 + 2(\text{tr}M_1)^4 \cdot \text{tr}M_2^2 + (\text{tr}M_1)^4 (\text{tr}M_2)^2, \\
n_{(6)}\lambda_{5,1} = & 3840\text{tr}M_1^5 M_2 + 1920\text{tr}M_1 \cdot \text{tr}M_1^4 M_2 + 960\text{tr}M_1^2 \cdot \text{tr}M_1^3 M_2 + 640\text{tr}M_1^3 \cdot \text{tr}M_1^2 M_2 + 480(\text{tr}M_1)^2 \cdot \text{tr}M_1^3 M_2 \\
& + 480\text{tr}M_1 \cdot \text{tr}M_1^2 \cdot \text{tr}M_1^2 M_2 + 480\text{tr}M_1^4 \cdot \text{tr}M_1 M_2 + 384\text{tr}M_1^5 \cdot \text{tr}M_2 + 320\text{tr}M_1 \cdot \text{tr}M_1^3 \cdot \text{tr}M_1 M_2 \\
& + 240\text{tr}M_1 \cdot \text{tr}M_1^4 \cdot \text{tr}M_2 + 160\text{tr}M_1^2 \cdot \text{tr}M_1^3 \cdot \text{tr}M_2 + 120(\text{tr}M_1)^2 \cdot \text{tr}M_1^2 \cdot \text{tr}M_1 M_2 + 120(\text{tr}M_1^2)^2 \cdot \text{tr}M_1 M_2 \\
& + 80(\text{tr}M_1)^3 \cdot \text{tr}M_1^2 M_2 + 80(\text{tr}M_1)^2 \cdot \text{tr}M_1^3 \cdot \text{tr}M_2 + 60\text{tr}M_1 \cdot (\text{tr}M_1^2)^2 \cdot \text{tr}M_2 + 20(\text{tr}M_1)^3 \cdot \text{tr}M_1^2 \cdot \text{tr}M_2 \\
& + 10(\text{tr}M_1)^4 \cdot \text{tr}M_1 M_2 + (\text{tr}M_1)^5 \cdot \text{tr}M_2, \\
n_{(7)}\lambda_{6,1} = & 46080\text{tr}M_1^6 M_2 + 23040\text{tr}M_1 \cdot \text{tr}M_1^5 M_2 + 11520\text{tr}M_1^2 \cdot \text{tr}M_1^4 M_2 + 7680\text{tr}M_1^3 \cdot \text{tr}M_1^3 M_2 \\
& + 5760(\text{tr}M_1)^2 \cdot \text{tr}M_1^4 M_2 + 5760\text{tr}M_1 \cdot \text{tr}M_1^2 \cdot \text{tr}M_1^3 M_2 + 5760\text{tr}M_1^4 \cdot \text{tr}M_1^2 M_2 + 4608\text{tr}M_1^5 \cdot \text{tr}M_1 M_2 \\
& + 3840\text{tr}M_1 \cdot \text{tr}M_1^3 \cdot \text{tr}M_1^2 M_2 + 3840\text{tr}M_1^6 \cdot \text{tr}M_2 + 2880\text{tr}M_1 \cdot \text{tr}M_1^4 \cdot \text{tr}M_1 M_2 + 2304\text{tr}M_1 \cdot \text{tr}M_1^5 \cdot \text{tr}M_2 \\
& + 1920\text{tr}M_1^2 \cdot \text{tr}M_1^3 \cdot \text{tr}M_1 M_2 + 1440(\text{tr}M_1)^2 \cdot \text{tr}M_1^2 \cdot \text{tr}M_1^2 M_2 + 1440(\text{tr}M_1^2)^2 \cdot \text{tr}M_1^2 M_2 + 1440\text{tr}M_1^2 \cdot \text{tr}M_1^4 \cdot \text{tr}M_2 \\
& + 960(\text{tr}M_1)^2 \cdot \text{tr}M_1^3 \cdot \text{tr}M_1 M_2 + 960\text{tr}M_1 \cdot \text{tr}M_1^2 \cdot \text{tr}M_1^3 \cdot \text{tr}M_2 + 960(\text{tr}M_1)^3 \cdot \text{tr}M_1^3 M_2 \\
& + 720\text{tr}M_1 \cdot (\text{tr}M_1^2)^2 \cdot \text{tr}M_1 M_2 + 720(\text{tr}M_1)^2 \cdot \text{tr}M_1^4 \cdot \text{tr}M_2 + 640(\text{tr}M_1^3)^2 \cdot \text{tr}M_2 + 240(\text{tr}M_1)^3 \cdot \text{tr}M_1^2 \cdot \text{tr}M_1 M_2 \\
& + 180(\text{tr}M_1)^2 \cdot (\text{tr}M_1^2)^2 \cdot \text{tr}M_2 + 160(\text{tr}M_1)^3 \cdot \text{tr}M_1^3 \cdot \text{tr}M_2 + 120(\text{tr}M_1)^4 \cdot \text{tr}M_1^2 M_2 + 120(\text{tr}M_1^2)^3 \cdot \text{tr}M_2 \\
& + 30(\text{tr}M_1)^4 \cdot \text{tr}M_1^2 \cdot \text{tr}M_2 + 12(\text{tr}M_1)^5 \cdot \text{tr}M_1 M_2 + (\text{tr}M_1)^6 \cdot \text{tr}M_2.
\end{aligned}$$

In summary, given ρ_0 and W , the steps to evaluate the finite sample skewness and excess kurtosis of $\hat{\rho}_n$ are as follows:

1. Calculate $\lambda_{i,j}$ as given above.
2. Plug $\lambda_{i,j}$ into Q and into expectations of terms involving ψ_n , H_1 , H_2 , and H_3 .
3. Plug the results from Step 2 into expectations of terms involving ξ_0 , $\xi_{-1/2}$, and ξ_{-1} .
4. Plug the results from Step 3 to (4) to calculate the approximate results.

Notes

¹We do not discuss the more complicated issue of cross moments of $\hat{\beta}_i$ and $\hat{\beta}_j$, $i \neq j$, in this paper. This issue is however important if we are interested in the finite sample properties of some test statistics that involve the whole parameter vector. We leave this for our future study.

²In an earlier version of this paper, the authors considered the standardized estimator $D^{-1/2}\sqrt{n}(\hat{\beta}_n - \beta_0)$. However, as pointed out by one referee and the co-editor, since in general D is unknown, in practice $\hat{D}_n^{-1/2}\sqrt{n}(\hat{\beta}_n - \beta_0)$, where \hat{D}_n consistently estimates D and $\hat{D}_n - D = O_P(n^{-1/2})$, is also of interest. The presence of \hat{D}_n will introduce additional terms into the expansion and we can show that for the i th feasible standardized estimator $\hat{D}_{n,ii}^{-1/2}\sqrt{n}(\hat{\beta}_{n,i} - \beta_{0,i}) = \xi_{0,i} + \xi_{-1/2,i} + \xi_{-1,i}$, where $\xi_{0,i} = \sqrt{n}D_{ii}^{-1/2}a_{-1/2,i}$, $\xi_{-1/2,i} = \sqrt{n}D_{ii}^{-1/2}\left[a_{-1,i} - \frac{1}{2}\left(\frac{\hat{D}_{n,ii} - D_{ii}}{D_{ii}}\right)a_{-1/2,i}\right]$, and $\xi_{-1,i} = \sqrt{n}D_{ii}^{-1/2}\left[a_{-3/2,i} + \frac{3}{8}\left(\frac{\hat{D}_{n,ii} - D_{ii}}{D_{ii}}\right)^2 a_{-1/2,i} - \frac{1}{2}\left(\frac{\hat{D}_{n,ii} - D_{ii}}{D_{ii}}\right)a_{-1,i}\right]$. With these newly defined ξ 's, the theorem presented in this paper is still valid for the skewness and kurtosis of the feasible standardized estimator $\hat{D}_{n,ii}^{-1/2}\sqrt{n}(\hat{\beta}_{n,i} - \beta_{0,i})$.

³A “one ahead and one behind” matrix has the i -th row with non-zero elements only in positions $i - 1$ and $i + 1$, $i = 2, \dots, n - 1$, and the first row has non-zero elements only in positions 2 and n while for the last row the non-zeros occur only in positions 1 and $n - 1$. By this, we define the weights matrix in a circular way. The average number of neighboring units J for the “one ahead and one behind” matrix is hence 2. Similarly, we can define the “two ahead and two behind,” “three ahead and three behind” matrices and so on.

⁴Note that in Bao and Ullah (2007a), a different approach, namely, the top-order invariant polynomial approach, was used to derive the second-order bias and MSE of $\hat{\rho}_n$.

References

- Bao, Y. & A. Ullah (2007a) Finite sample moments of maximum likelihood estimator in spatial models. *Journal of Econometrics* 137, 396-413
- Bao, Y. & A. Ullah (2007b) The second-order bias and mean squared error of estimators in time series models. *Journal of Econometrics* 140, 650-669.
- Bao, Y. & A. Ullah (2007c) Expectation of quadratic forms in normal and nonnormal variables with econometric applications. Manuscript, Temple University.
- Basmann, R.L. (1974) Exact finite sample distribution for some econometric estimators and test statistics: a survey and appraisal. In Intriligator, M.D. & D.A. Kendrick (eds.) *Frontiers of Quantitative Economics*, Vol. 2. Amsterdam: North-Holland, pp. 209-288.
- Fisher, R.A. (1921) On the probable error of a coefficient of correlation deduced from a small sample. *Metron* 1, 1-32.
- Ghazal, G.A. (1994) Moments of the ratio of two dependent quadratic forms. *Statistics and Probability Letters* 20, 313-319.
- Kelejian, H.H. & I.R. Prucha (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509-533.
- Lee, L.F. (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899-1925.
- Linton, O. (1997) An Asymptotic Expansion in the GARCH(1,1) Model. *Econometric Theory* 13, 558-581.

- McCullagh, P. (1987) *Tensor Methods in Statistics*. London: Chapman and Hall.
- Nagar, A.L. (1959) The bias and moment matrix of the general k -class estimators of the parameters in simultaneous equations. *Econometrica* 27, 575-595.
- Rilstone, P., V.K. Srivastava & A. Ullah (1996) The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics* 75, 369-395.
- Rothenberg, T.J. (1984) Approximating the distribution of econometric estimators and test statistics. In Griliches, Z. & M.D. Intriligator (eds.) *Handbook of Econometrics, Vol 2*. Amsterdam: North-Holland, pp. 881-935.
- Sargan, J.D. (1974) The validity of Nagar's expansion for the moments of econometric estimators. *Econometrica* 42, 169-176.
- Sawa, T. (1972) Finite sample properties of the k -class estimators. *Econometrica* 40, 653-680.
- Srinivasan, T.N. (1970) Approximation to finite sample moments of estimators where exact sampling distributions are unknown. *Econometrica* 38, 533-541.
- Ullah, A. (2004) *Finite Sample Econometrics*. New York: Oxford University Press.

Table 1: Sample and Theoretical Bias, MSE, Skewness, and Kurtosis, $n = 30$

J	ρ_0	Bias	B	MSE	M	SK	γ_1	KR	γ_2
2	-0.9	0.020	0.018	0.004	0.003	1.404	3.977	2.602	0.428
	-0.4	0.018	0.019	0.025	0.026	0.473	0.251	0.184	0.335
	-0.2	0.013	0.011	0.031	0.033	0.244	0.112	-0.167	0.892
	0	-0.002	0.000	0.031	0.035	-0.091	0.000	-0.265	1.073
	0.2	-0.004	-0.011	0.028	0.033	-0.341	-0.112	-0.052	0.892
	0.4	-0.019	-0.019	0.027	0.026	-0.406	-0.251	-0.187	0.335
	0.9	-0.019	-0.018	0.003	0.003	-2.082	-3.977	13.400	0.428
6	-0.9	0.102	0.002	0.070	0.118	1.122	0.182	0.306	0.593
	-0.4	-0.016	-0.046	0.111	0.114	0.032	-0.140	-0.676	-2.126
	-0.2	-0.063	-0.062	0.111	0.105	-0.224	-0.349	-0.535	-3.668
	0	-0.076	-0.074	0.109	0.093	-0.493	-0.672	-0.102	-4.840
	0.2	-0.084	-0.080	0.102	0.077	-0.759	-1.232	0.508	-5.029
	0.4	-0.085	-0.080	0.076	0.060	-1.022	-2.422	1.343	-3.520
	0.9	-0.050	-0.039	0.014	0.014	-2.448		10.354	9.685
10	-0.9	0.144	-0.056	0.110	0.188	1.081	-0.140	0.139	-8.352
	-0.4	-0.068	-0.122	0.160	0.182	0.192	-1.191	-1.024	-12.644
	-0.2	-0.097	-0.138	0.178	0.178	-0.185	-2.200	-0.916	-10.195
	0	-0.102	-0.147	0.179	0.170	-0.532	-4.327	-0.490	-5.196
	0.2	-0.146	-0.148	0.176	0.157	-0.782	-10.486	0.069	0.932
	0.4	-0.142	-0.139	0.156	0.137	-1.217	-60.228	1.173	6.648
	0.9	-0.084	-0.058	0.041	0.046	-3.669		20.303	4.724

Note: For each J and ρ_0 , Bias is the average bias of the sample estimates over 1,000 replications; B is the theoretical second-order bias; MSE is the mean squared error of the sample estimates over the 1,000 replications; M is the theoretical second-order mean squared error; SK and KR are the sample skewness and excess kurtosis coefficients, respectively, of the estimates over the 1,000 replications; γ_1 and γ_2 are the theoretical approximate skewness and excess kurtosis coefficients.

Table 2: Sample and Theoretical Bias, MSE, Skewness, and Kurtosis, $n = 100$

J	ρ_0	Bias	B	MSE	M	SK	γ_1	KR	γ_2
2	-0.9	0.006	0.005	0.001	0.001	0.839	1.133	1.324	1.015
	-0.4	0.002	0.006	0.007	0.007	0.226	0.224	-0.073	0.072
	-0.2	0.006	0.003	0.009	0.009	0.197	0.106	0.196	0.092
	0	0.004	0.000	0.010	0.010	-0.038	0.000	-0.110	0.100
	0.2	-0.001	-0.003	0.009	0.009	-0.197	-0.106	0.008	0.092
	0.4	-0.007	-0.006	0.007	0.007	-0.330	-0.224	0.159	0.072
	0.9	-0.005	-0.005	0.001	0.001	-0.796	-1.133	0.890	1.015
6	-0.9	0.045	0.000	0.027	0.037	0.987	0.082	0.198	-0.085
	-0.4	-0.012	-0.014	0.039	0.038	-0.245	-0.168	0.022	-0.275
	-0.2	-0.008	-0.019	0.036	0.035	-0.207	-0.287	-0.105	-0.292
	0	-0.023	-0.022	0.030	0.031	-0.462	-0.428	0.367	-0.222
	0.2	-0.022	-0.024	0.026	0.025	-0.448	-0.606	0.351	-0.017
	0.4	-0.024	-0.024	0.018	0.018	-0.707	-0.859	0.789	0.416
	0.9	-0.010	-0.012	0.002	0.002	-1.479	-6.529	4.115	7.066
10	-0.9	0.063	-0.017	0.039	0.079	1.113	-0.141	0.581	-0.580
	-0.4	-0.027	-0.037	0.067	0.071	-0.183	-0.541	-0.429	-0.263
	-0.2	-0.030	-0.042	0.060	0.064	-0.432	-0.766	0.019	0.165
	0	-0.037	-0.045	0.055	0.055	-0.549	-1.062	0.290	0.824
	0.2	-0.039	-0.045	0.041	0.044	-0.541	-1.483	0.038	1.761
	0.4	-0.046	-0.042	0.034	0.032	-0.933	-2.165	1.505	3.082
	0.9	-0.022	-0.018	0.004	0.004	-1.883	-148.590	7.988	10.651

Note: For each J and ρ_0 , Bias is the average bias of the sample estimates over 1,000 replications; B is the theoretical second-order bias; MSE is the mean squared error of the sample estimates over the 1,000 replications; M is the theoretical second-order mean squared error; SK and KR are the sample skewness and excess kurtosis coefficients, respectively, of the estimates over the 1,000 replications; γ_1 and γ_2 are the theoretical approximate skewness and excess kurtosis coefficients.

Table 3: Sample and Theoretical Bias, MSE, Skewness, and Kurtosis, $n = 200$

J	ρ_0	Bias	B	MSE	M	SK	γ_1	KR	γ_2
2	-0.9	0.002	0.003	0.000	0.000	0.530	0.684	0.412	0.579
	-0.4	0.000	0.003	0.004	0.004	0.255	0.176	0.028	0.038
	-0.2	0.002	0.002	0.005	0.005	0.202	0.085	0.068	0.018
	0	-0.001	0.000	0.005	0.005	0.013	0.000	-0.004	0.012
	0.2	-0.002	-0.002	0.004	0.005	-0.108	-0.085	0.118	0.018
	0.4	-0.006	-0.003	0.004	0.004	-0.228	-0.176	0.376	0.038
	0.9	-0.003	-0.003	0.000	0.000	-0.547	-0.684	0.442	0.579
6	-0.9	0.015	0.000	0.013	0.019	0.840	0.054	0.016	-0.056
	-0.4	-0.008	-0.007	0.019	0.019	-0.057	-0.134	0.015	-0.082
	-0.2	-0.007	-0.009	0.018	0.018	-0.171	-0.216	-0.015	-0.055
	0	-0.009	-0.011	0.014	0.015	-0.196	-0.306	-0.180	0.007
	0.2	-0.015	-0.012	0.012	0.012	-0.342	-0.409	0.039	0.118
	0.4	-0.015	-0.012	0.009	0.009	-0.581	-0.539	0.312	0.313
	0.9	-0.006	-0.006	0.001	0.001	-0.968	-1.942	1.784	3.477
10	-0.9	0.034	-0.009	0.022	0.041	1.031	-0.104	0.413	-0.182
	-0.4	-0.013	-0.019	0.033	0.036	-0.182	-0.351	0.079	0.010
	-0.2	-0.019	-0.021	0.032	0.032	-0.388	-0.469	0.347	0.190
	0	-0.028	-0.022	0.027	0.026	-0.509	-0.606	0.486	0.452
	0.2	-0.020	-0.022	0.020	0.021	-0.571	-0.776	0.369	0.826
	0.4	-0.025	-0.021	0.015	0.014	-0.593	-1.008	0.444	1.378
	0.9	-0.008	-0.009	0.001	0.001	-1.440	-4.991	4.147	7.157

Note: For each J and ρ_0 , Bias is the average bias of the sample estimates over 1,000 replications; B is the theoretical second-order bias; MSE is the mean squared error of the sample estimates over the 1,000 replications; M is the theoretical second-order mean squared error; SK and KR are the sample skewness and excess kurtosis coefficients, respectively, of the estimates over the 1,000 replications; γ_1 and γ_2 are the theoretical approximate skewness and excess kurtosis coefficients.