Approximated Likelihood and Root Estimators for Spatial Interaction in Spatial Autoregressive Models[☆]

Fei Jin[a,*], Lung-fei Lee[a]

[a]*Department of Economics, Ohio State University, Columbus, OH 43210 USA*

Abstract

In this paper, we first generalize an approximate measure of spatial dependence, the $APLE$ statistic (Li et al., 2007), to a spatial Durbin (SD) model. This generalized $APLE$ takes into account exogenous variables directly and can be used to detect spatial dependence originating from either a spatial autoregressive (SAR), spatial error (SE) or SD process. However, that measure is not consistent. Secondly, by examining carefully the first order condition of the concentrated log likelihood of the SD (or SAR) model, whose first order approximation generates the $APLE$, we construct a moment equation quadratic in the autoregressive parameter that generalizes an original estimation approach in Ord (1975) and yields a closed-form consistent root estimator of the autoregressive parameter. With a specific moment equation constructed from an initial consistent estimator, the root estimator can be as efficient as the MLE under normality. Furthermore, when there is unknown heteroskedasticity in the disturbances, we derive a modified $APLE$ and a root estimator which can be robust to unknown heteroskedasticity. The root estimators are computationally much simpler than the quasi-maximum likelihood estimators.

Keywords: spatial autoregressive model, spatial error model, spatial Durbin model, $APLE$, GMM
JEL classification: C21, R15

1. Introduction

Li et al. (2007) propose a closed-form measure of spatial dependence, an approximate profile-likelihood estimator ($APLE$), based on a pure spatial autoregressive (SAR) model. Their Monte Carlo experiments for spatial weights matrices defined according to a second-order neighborhood structure on toroidal lattices show that the $APLE$ provides a better assessment of the strength of spatial dependence for data generated by the pure SAR model than alternative measures such as Moran's I (Moran, 1950). Thus, the $APLE$ provides a better measure of spatial dependence than Moran's I for exploratory analyses. It has been

[☆]We are grateful to the editor, two anonymous referees and seminar participants at Ohio State University and the 20th Annual Meetings of the Midwest Econometrics Group for their helpful comments.
[*]Corresponding author. Tel.: +1 614 247 8481; fax: +1 614 292 3906.
Email addresses: jin.120@osu.edu (Fei Jin), lee.1777@osu.edu (Lung-fei Lee)

shown in Martellosio (2010) that Moran's I has zero power to detect spatial correlation in a SAR model when the autoregressive coefficient is large and close to one. Li et al. (2012) generalize the $APLE$ statistic to the spatial error (SE) model to account for exogenous variables. As both the SAR and SE models are constrained forms of the more general spatial Durbin (SD) model, an approximate measure for spatial dependence of interest should account for exogenous variables directly and provide a good approximation to the autoregressive parameter in the SD model. This approximate measure can be used to detect spatial dependence originating from either the SAR, SE or SD model. In this paper, we extend the $APLE$ to the SD model, which, similarly to Li et al. (2007), is based on a first order approximation to the first order condition of the concentrated log likelihood of the SD model. The original $APLE$ as well as the extended $APLE$ from the first order approximations are not consistent for the autoregressive parameter due to a systematic bias. Higher order approximations of the first order condition may generate more accurate measures of the autoregressive parameter, but they involve multiple roots and generally do not yield closed-form solutions. Treating the first order condition differently, we obtain a moment equation quadratic in the autoregressive parameter that generates a closed-form root estimator. Our proposed root estimator generalizes an estimator originated in Ord (1975) in a general setting. For the quadratic moment equation, conditions under which one of the roots is consistent will be specified. With an initial consistent estimator, a moment equation can be designed to generate a second step root estimator which is asymptotically as efficient as the maximum likelihood estimator (MLE) under normality. Once an estimate of the autoregressive parameter is available, other parameters in a SD model may be estimated by least squares (LS) after applying a spatial filter to the data on the dependent variable. The modified $APLE$ and the root estimator can be used as measures of spatial dependence or simple estimators for the autoregressive parameter in a SD or SAR model, as the SD model nests the SAR model.

The proposed estimators can further be extended to possess some robust properties. The original $APLE$ in Li et al. (2007) has not accounted for possible heteroskedasticity in the disturbances. Li et al. (2012) argue that a valid transformation can be applied to the SE model, so the extended $APLE$ may be calculated with the transformed data. This is so when the heteroskedastic variance has a known functional form.[1] However, if we do not know the form of heteroskedasticity, the data could not be properly transformed. A misspecified transformation can lead to errors in inference. With unknown heteroskedasticity, we may adjust the first order condition to derive a modified $APLE$ statistic, which we call an approximate concentrated moment estimator ($ACME$), and we can also adjust the moment equation to derive a root estimator that is robust to unknown heteroskedasticity.

Existing estimation methods for SAR (SD) models do not have a closed form and are usually computa-

[1] In this situation, we can also easily modify our root estimates to accommodate the known heteroskedasticity because after proper transformation, it results in a SAR model with homogenous disturbances.

tionally involved.[2] The MLE or quasi-maximum likelihood estimator (QMLE) does not have a closed form (Anselin, 1988; Lee, 2004a). The computation involves the evaluation of the log-determinant of a square matrix with dimension equal to the sample size at different parameter values, so it might be computationally demanding when the sample size is large.[3] Some empirical applications may create large matrices, for example, the US Census Bureau collects data at over 250,000 census block group locations and the Home Mortgage Disclosure Act data have over 100 million observations. Because of the computational burden of the MLE, even with sample sizes that might not be too large, researchers may turn to less efficient estimation methods such as the two stage least squares (2SLS) proposed by Kelejian and Prucha (1998).[4] For example, Helms (2012) uses the 2SLS estimation when the sample size is 16,638. Lee (2007a) considers the generalized method of moments (GMM) estimation, which combines the quadratic moments that capture the correlation across the spatial units with the linear moments used in the 2SLS approach. Compared to the QMLE, the GMM estimator is computationally simpler and it can be as efficient as the MLE under normality.[5] Lee (2007b) proposes a computationally simpler GMM for the estimation of SAR models. The method reduces the GMM estimation of a vector of parameters into nonlinear estimation of only the autoregressive parameter. It can reduce the computational burden substantially and it may be as efficient as the joint GMM estimator under certain conditions. But it still does not generate a closed-form solution and searching over a parameter space is necessary. Even though the GMM avoids computing log-determinants of matrices, searching over a parameter space with large matrices involved could still be computationally intensive. Our root estimator is asymptotically as efficient as the MLE under normality since the designed second step moment equation automatically combines the linear and quadratic moment conditions in an efficient way.[6] For SAR models with unknown heteroskedasticity, Lin and Lee (2010) study the GMM estimation

[2] Because of the correlation of the spatially lagged dependent variable with disturbances, the LS estimator is only consistent for a subclass of models (Lee, 2002).

[3] Various techniques and simplifications have been proposed to tackle this problem, see, for example, Martin (1993), Griffith and Sone (1995), Pace (1997), Pace and Barry (1997a,b), Barry and Pace (1999), Griffith (2000), Smirnov and Anselin (2001), Pace and LeSage (2004), Pace and LeSage (2009) and Smirnov and Anselin (2009). Even with these techniques and simplifications, the computation can be still time-consuming. Alternative simplifications often lead to less accurate estimates. We note that Pace and LeSage (2009) propose a sampling approach to estimate the log determinant. Their Monte Carlo study shows that the approach can be very fast in estimating the log determinant. Given that the log determinant needs to be evaluated many times at different parameter values, the actual time of computing an MLE or QMLE may be much longer.

[4] Their model is more general one with both a spatial lag of the dependent variable and a SAR process in the disturbances. While the autoregressive parameter for the spatial lag of the dependent variable is estimated by 2SLS, the autoregressive parameter in the disturbance process is estimated by GMM with three moment equations.

[5] Liu et al. (2010) and Lee and Liu (2010) consider the efficient GMM estimation of the regular and high order SAR models with properly modified moment equations. Their estimator is as efficient as the MLE under normality and is more efficient than the QMLE otherwise.

[6] Both the modified GMM and our root estimator reduce the estimation to that of only the autoregressive parameter, which might lead to better finite sample performance when a bias correction might be constructed and applied to this single estimate,

where linear and quadratic moment equations involving both the autoregressive parameter and parameters for other exogenous variables are used.[7] Our (robust) root estimator is obtained with a properly modified and combined moment equation quadratic in the autoregressive parameter. Thus, for the closed-form root estimator (see Eq. (23)), no searching over a parameter space is needed. Because of the closed form, the root estimator requires little programming effort. Our Monte Carlo study shows that the root estimator has similar finite sample performance as the QMLE under normality and the robust root estimator performs well under unknown heteroskedasticity. Computing the root estimates only takes slightly longer time than computing the $APLE$, which is much faster than computing the QMLE. As the computational burden of both the modified $APLE$ and the root estimate is minimal, they can be applied to SAR, SE or SD models for huge data sets.

The rest of the paper is organized as follows. Section 2 introduces related models and develops the $APLE$ and $ACME$; Section 3 establishes the consistency and asymptotic distribution of our root estimators in both the homoskedastic and heteroskedastic cases; Section 4 presents some Monte Carlo results; Section 5 concludes. Some lemmas and proofs are collected in the Appendix.

2. The Models, $APLE$ and $ACME$

In this section, we introduce the related models, and then derive the $APLE$ for the SD model when ϵ_{ni}'s are i.i.d., and the $ACME$ when ϵ_{ni}'s may be only independent but with different and unknown variances.

A SAR model is specified as

$$y_n = \rho W_n y_n + X_n\beta + \epsilon_n, \tag{1}$$

where n is the sample size, y_n is an n-dimensional vector of observations, W_n is an $n \times n$ spatial weights matrix with a zero diagonal, X_n is an $n \times k$ matrix of exogenous variables, $\epsilon_n = (\epsilon_{n1}, \ldots, \epsilon_{nn})'$ with ϵ_{ni}'s being independent with mean zero, and ρ is an autoregressive parameter. If the spatial dependence is in the disturbances instead, we have a SE model which is

$$y_n = X_n\beta + u_n, \quad u_n = \rho W_n u_n + \epsilon_n. \tag{2}$$

Let I_n denote the n-dimensional identity matrix. Pre-multiplying both sides of Eq. (2) by $(I_n - \rho W_n)$ yields

$$y_n = \rho W_n y_n + X_n\beta + W_n X_n(-\rho\beta) + \epsilon_n, \tag{3}$$

compared to the case when a complete vector of parameters are estimated jointly and then the bias correction is applied to this vector of estimates.

[7] Kelejian and Prucha (2010) also consider the specification and estimation of the SAR model with SAR disturbances that has heteroskedastic innovations. As in Kelejian and Prucha (1998), the autoregressive parameter for the spatial lagged dependent variable is estimated by 2SLS and the autoregressive parameter in the disturbance process is estimated by GMM with multiple moment equations.

which is a constrained form of the SD model[8]

$$y_n = \rho W_n y_n + X_n\beta + W_n X_n \gamma + \epsilon_n. \tag{4}$$

That is, γ in the SD model (4) is required to be equal to minus ρ times β for the SE process. A regression model with the SAR process is just the SD model (4) with $\gamma = 0$, so it is also a constrained form of the SD model. Without the constraints, the SD model may also have an interest of its own. The $W_n X_n$ as regressors may capture externality arising form neighbors' characteristics (see, e.g., LeSage and Pace 2009, p. 30). If W_n is row-normalized and X_n contains an intercept term, i.e., $X_n = [l_n, X_{1n}]$, where l_n is an n-dimensional column vector of ones and X_{1n} is an $n \times (k-1)$ matrix, then $W_n X_n$ will generate a column vector of ones as $W_n X_n = [l_n, W_n X_{1n}]$. Coefficients on these two column vectors of ones should be collected together. If W_n is not row-normalized, the columns of X_n and $W_n X_n$ are in general linearly independent. In this case, for l_n in X_n, $W_n l_n$ is the vector of row sums which is an extra regressor.[9] To make later narrative easier, we write the SD model as

$$y_n = \rho W_n y_n + Z_n\theta + \epsilon_n, \tag{5}$$

where $Z_n = [X_n, W_n X_{1n}]$ or $Z_n = [X_n, W_n X_n]$, depending on whether both X_n and $W_n X_n$ contain a column vector of ones or not, and θ is the corresponding vector of coefficients. The Z_n is $n \times d$ with $d = 2k - 1$ or $d = 2k$. The $APLE$ and $ACME$ are derived for the SD model (5). Our root estimators are also stated with the setting of Eq. (5). When a SAR model rather than a more general SD model is considered, just take Z_n to be X_n. Let the true parameters of ρ and θ be ρ_0 and θ_0. When ϵ'_{ni}'s are i.i.d. $(0, \sigma^2)$, the true parameter for σ^2 is σ_0^2; when there is unknown heteroskedasticity, $E(\epsilon_n \epsilon'_n) = Diag(\sigma_{n1}^2, \ldots, \sigma_{nn}^2) = \Sigma_n$, where $Diag(a_n)$ denote a diagonal matrix with the diagonal elements being those of the vector a_n. Let $S_n(\rho) = I_n - \rho W_n$ and $G_n(\rho) = W_n S_n^{-1}(\rho)$. Denote $S_n = S_n(\rho_0)$ and $G_n = G_n(\rho_0)$ for short.

When ϵ_{ni}'s are i.i.d. with variance σ^2, the log likelihood function for the model (5) is

$$L_n(\rho, \theta, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) + \ln|S_n(\rho)| - \frac{1}{2\sigma^2}[S_n(\rho)y_n - Z_n\theta]'[S_n(\rho)y_n - Z_n\theta].$$

Maximizing the function with a fixed ρ, we obtain the QMLEs for θ and σ^2 as:

$$\hat{\theta}_n = (Z'_n Z_n)^{-1} Z'_n S_n(\rho) y_n, \tag{6}$$

$$\hat{\sigma}_n^2 = \frac{1}{n} y'_n S'_n(\rho) M_{Z_n} S_n(\rho) y_n, \tag{7}$$

where $M_{Z_n} = I_n - Z_n(Z'_n Z_n)^{-1} Z'_n$. Eqs. (6) and (7) are just like the LS estimators after the spatial filter $S_n(\rho)$ has been applied to y_n. Substituting these expressions into the log likelihood function, we have the

[8] We use this terminology following LeSage and Pace (2009).

[9] If elements in W_n are 0 or 1 as in a network, a row sum refers to an outdegree. So in such a case, the outdegrees of individuals form an explanatory variable.

concentrated log (or profile) likelihood function of ρ:

$$L_n(\rho) = -\frac{n}{2}[\ln(2\pi/n)+1] + \ln|S_n(\rho)| - \frac{n}{2}\ln[y_n' S_n'(\rho) M_{Z_n} S_n(\rho) y_n].$$

The first order condition for the maximization of the concentrated log likelihood function is:

$$\frac{n y_n' S_n'(\rho) M_{Z_n} W_n y_n}{y_n' S_n'(\rho) M_{Z_n} S_n(\rho) y_n} - tr[G_n(\rho)] = 0, \tag{8}$$

where $tr(A_n)$ denotes the trace of a square matrix A_n. Multiplying both sides by $\frac{1}{n} y_n' S_n'(\rho) M_{Z_n} S_n(\rho) y_n$ yields

$$y_n' S_n'(\rho) M_{Z_n} W_n y_n - y_n' S_n'(\rho) M_{Z_n} S_n(\rho) y_n \frac{tr[G_n(\rho)]}{n} = 0. \tag{9}$$

Similar to the derivation of the $APLE$ in Li et al. (2007), an approximate measure of ρ can be obtained from a first order approximation of the left-hand side of Eq. (9). Note that $tr[G_n(\rho)] = tr[W_n(I_n + \rho W_n + \dots)] \approx \rho tr(W_n^2)$ as W_n has a zero diagonal, the approximation yields

$$APLE_{sd} = \frac{y_n' M_{Z_n} W_n y_n}{y_n' W_n' M_{Z_n} W_n y_n + y_n' M_{Z_n} y_n \frac{tr(W_n^2)}{n}}, \tag{10}$$

For the convenience of later reference, we also write down the $APLE$ for the SAR model as

$$APLE_{sar} = \frac{y_n' M_{X_n} W_n y_n}{y_n' W_n' M_{X_n} W_n y_n + y_n' M_{X_n} y_n \frac{tr(W_n^2)}{n}}, \tag{11}$$

where $M_{X_n} = I_n - X_n(X_n' X_n)^{-1} X_n'$. Furthermore, for a pure SAR process, $M_{Z_n} = I_n$. As $y_n' W_n y_n = y_n'[(W_n' + W_n)/2] y_n$ and $tr(W_n^2) = \lambda'\lambda$, where λ is the vector of W_n's eigenvalues, Eq. (10) would reduce to

$$\frac{y_n'[(W_n' + W_n)/2] y_n}{y_n' W_n' W_n y_n + y_n' y_n \frac{\lambda'\lambda}{n}},$$

which is that given in Li et al. (2007).

Using the same approach, Li et al. (2012) derive the $APLE$ for the SE model as

$$APLE_{se} = \frac{y_n' M_{X_n}[(W_n + W_n')/2] M_{X_n} y_n}{A_n}, \tag{12}$$

where $A_n = y_n' M_{X_n} W_n' W_n M_{X_n} y_n - y_n' M_{X_n}(W_n + W_n')(I_n - M_{X_n})(W_n + W_n') M_{X_n} y_n + y_n' M_{X_n} y_n \frac{tr(W_n^2)}{n}$. The $APLE$ for the SAR model in Eq. (11) and that for the SE model in Eq. (12) have different forms. Alternatively, we could use the $APLE$ based the SD model in Eq. (10) as an approximate measure of spatial dependence originating from either the SAR, SE or SD model.

Eq. (9) can be rewritten as

$$y_n' S_n'(\rho)\Big[G_n'(\rho) - \frac{tr[G_n'(\rho)]}{n} I_n\Big] M_{Z_n} S_n(\rho) y_n = 0. \tag{13}$$

When there is unknown heteroskedasticity, the expectation of the left-hand side of the above equation over n at the true parameters $\rho_0, \theta_0, \sigma_{n1}^2, \ldots, \sigma_{nn}^2$ does not converge to zero in general, since

$$
\begin{aligned}
&\frac{1}{n}E\Big\{y_n' S_n'\Big[G_n' - \frac{tr(G_n')}{n}I_n\Big]M_{Z_n}S_n y_n\Big\} \\
&= \frac{1}{n}E\Big\{(Z_n\theta_0 + \epsilon_n)'\Big[G_n' - \frac{tr(G_n')}{n}I_n\Big]M_{Z_n}(Z_n\theta_0 + \epsilon_n)\Big\} \\
&= \frac{1}{n}tr\Big\{\Big[G_n' - \frac{tr(G_n')}{n}I_n\Big]M_{Z_n}\Sigma_n\Big\} \\
&= \frac{1}{n}tr\Big\{\Big[G_n' - \frac{tr(G_n')}{n}I_n\Big]\Sigma_n\Big\} + o(1) \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big[G_n' - \frac{tr(G_n')}{n}I_n\Big]_{ii}\sigma_{ni}^2 + o(1),
\end{aligned}
\tag{14}
$$

by Lemma 1 in the Appendix. Under unknown heteroskedasticity, we may modify Eq. (13) into the following equation

$$
y_n' S_n'(\rho)\big[G_n'(\rho) - Diag[G_n'(\rho)]\big]M_{Z_n}S_n(\rho)y_n = 0, \tag{15}
$$

which is a valid moment equation because the zero diagonal of $G_n'(\rho) - Diag[G_n'(\rho)]$ implies that the expectation of the left-hand side of the equation over n at ρ_0 converges to zero. Taking a first order Taylor expansion of the left-hand side of Eq. (15) with ρ and setting it to zero yield a modified $APLE$ statistic, which we call $ACME$:

$$
ACME_{sd} = \frac{y_n' M_{Z_n} W_n y_n}{y_n' W_n' M_{Z_n} W_n y_n + y_n' Diag(W_n^2) M_{Z_n} y_n}. \tag{16}
$$

For the SAR model, the $ACME$ is

$$
ACME_{sar} = \frac{y_n' M_{X_n} W_n y_n}{y_n' W_n' M_{X_n} W_n y_n + y_n' Diag(W_n^2) M_{X_n} y_n}. \tag{17}
$$

For a pure SAR process, Eq. (16) simplifies to

$$
\frac{y_n' W_n y_n}{y_n' W_n' W_n y_n + y_n' Diag(W_n^2) y_n}. \tag{18}
$$

Eqs. (16)—(18) can be used as approximate measures of ρ when unknown heteroskedasticity exists. Eqs. (10) and (16) (or Eqs. (11) and (17)) only differ in the second terms of their denominators.

3. A Root Estimator

3.1. A Root Estimator: Homoskedastic Case

Eq. (13) also motivates an extended GMM root estimator for ρ of the SD model when ϵ_{ni}'s are i.i.d.. The matrix $G_n'(\rho) - I_n \cdot tr[G_n'(\rho)]/n$ in Eq. (13) has a zero trace. Not accounting for the ρ's in the matrix, Eq. (13) is quadratic in ρ. Replacing the matrix with any $n \times n$ constant matrix P_n satisfying $tr(P_n M_{Z_n}) = 0$

(or $tr(P_n) = 0$)[10], a consistent GMM root estimator can be derived by solving the equation

$$g_n(\rho) = y_n' S_n'(\rho) P_n M_{Z_n} S_n(\rho) y_n = 0, \tag{19}$$

because the expectation of $g_n(\rho_0)$ is zero:

$$\begin{aligned} E[g_n(\rho_0)] &= E[(Z_n\theta_0 + \epsilon_n)' P_n M_{Z_n}(Z_n\theta_0 + \epsilon_n)] \\ &= \sigma_0^2 tr(P_n M_{Z_n}) = 0. \end{aligned}$$

The $P_n = G_n' - I_n \cdot tr(G_n')/n$ or $P_n = G_n' - I_n \cdot tr(G_n' M_{Z_n})/n$ is expected to generate a root estimator that is asymptotically as efficient as the MLE under normality since Eq. (19) with $P_n = G_n' - I_n \cdot tr(G_n')/n$ is essentially the first order condition of the concentrated log likelihood function Eq. (13), even though there is a single moment equation.[11] The form of the moment equation automatically combines the linear and quadratic moments in a way such that the root estimator can be efficient under normality, unlike Lee (2007a) or Lee (2007b), where linear moments are used together with the quadratic moments as a system with optimum weighting by the inverse of their variance-covariance matrix. This is not surprising because the single moment equation is motivated from the first order condition of the concentrated log likelihood function. Once a consistent estimator of ρ is available, Eqs. (6) and (7) can be used to calculate estimates for β and σ^2, respectively.

To establish the consistency of the root estimator, the following regularity conditions are assumed.

Assumption 1. *ϵ_{ni}'s in $\epsilon_n = (\epsilon_{n1}, \ldots, \epsilon_{nn})'$ are i.i.d. $(0, \sigma_0^2)$ and the moment $E(|\epsilon_{ni}^{4+\eta}|)$ exists for some $\eta > 0$.*

Assumption 2. *Matrices $\{W_n\}$ and $\{S_n^{-1}\}$ are bounded in both row and column sum norms (Horn and Johnson, 1985). The diagonal elements of W_n are zero.*

Assumption 3. *Elements of X_n are uniformly bounded constants, Z_n has full column rank and $\lim_{n\to\infty} \frac{Z_n' Z_n}{n}$ exists and is nonsingular.*

Assumption 4. *Constant n-dimensional square matrices $\{P_n = [p_{n,ij}]\}$ which satisfy $tr(P_n M_{Z_n}) = 0$ are bounded in both row and column sum norms.*

[10] We still have a consistent estimator if $tr(P_n) = 0$ instead of $tr(P_n M_{Z_n}) = 0$. This is so because for the expectation of the left-hand side of Eq. (19) at ρ_0, the additional term divided by n is $-\frac{\sigma_0^2}{n} tr[P_n Z_n (Z_n' Z_n)^{-1} Z_n'] = -\frac{\sigma_0^2}{n} tr[Z_n' P_n Z_n (Z_n' Z_n)^{-1}] = O(\frac{d}{n})$, which is not exactly zero but converges to zero as n goes to infinity. However, using a matrix P_n such that $tr(P_n M_{Z_n}) = 0$ might have better small sample properties. Given any matrix A_n, such a P_n matrix can be constructed as $P_n = A_n - \frac{tr(A_n M_{Z_n})}{n-d} I_n$.

[11] For a pure SAR process, Eq. (19) will be reduced to $y_n' S_n'(\rho) P_n S_n(\rho) y_n = 0$. In Ord (1975), based on the motivation of a modified LS estimation, he considered the quadratic moment $y_n' S_n'(\rho) W_n S_n(\rho) y_n = 0$, but dismissed it in favor of the MLE in terms of efficiency. The Eq. (19) with a class of P_n provides a general framework including the Ord's moment equation. One can overcome the relative inefficiency of the Ord's moment estimator by the selection of an efficient P_n as above.

The existence of a moment higher than the fourth order of the disturbances in Assumption 1 is needed for the application of the central limit theorem for linear and quadratic forms (Kelejian and Prucha, 2001). The boundedness in row and column sum norms of a sequence of matrices in Assumption 2 originated in Kelejian and Prucha (1998, 1999, 2001). Assumption 3 is required for convenience, as in Lee (2004a). As P_n is often generated from W_n, it is reasonable to assume that $\{P_n\}$ are bounded in both row and column sum norms.

The quadratic moment equation Eq. (19) has two roots in general. Under certain conditions, one of the roots is consistent. Let $B_n^s = B_n + B_n'$ for any n-dimensional square matrix B_n.

Proposition 1. *Under Assumptions 1—4, if* $(Z_n\theta_0)'P_nM_{Z_n}G_n(Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(P_n^sG_n^s)$ *were non-negative, the consistent root for* ρ_0 *of Eq. (19) would be*

$$\hat{\rho}_{1n} = \frac{b_n - \sqrt{b_n^2 - 4a_nc_n}}{2a_n}, \tag{20}$$

where $a_n = y_n'W_n'P_nM_{Z_n}W_ny_n$, $b_n = y_n'(P_nM_{Z_n})^sW_ny_n$ *and* $c_n = y_n'P_nM_{Z_n}y_n$; *but if* $(Z_n\theta_0)'P_nM_{Z_n}G_n(Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(P_n^sG_n^s)$ *were negative, the consistent root would be*

$$\hat{\rho}_{2n} = \frac{b_n + \sqrt{b_n^2 - 4a_nc_n}}{2a_n}, \tag{21}$$

when $\lim_{n\to\infty} \frac{1}{n}[(Z_n\theta_0)'G_n'P_nM_{Z_n}G_n(Z_n\theta_0) + \sigma_0^2 tr(G_n'P_nG_n)] \neq 0$. *In the case that* $\lim_{n\to\infty} \frac{1}{n}[(Z_n\theta_0)'G_n'P_nM_{Z_n}G_n(Z_n\theta_0) + \sigma_0^2 tr(G_n'P_nG_n)] = 0$, $\hat{\rho}_{3n} = c_n/b_n$ *is the unique consistent root if* $\lim_{n\to\infty} \frac{1}{n}[(Z_n\theta_0)'P_nM_{Z_n}G_n(Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(P_n^sG_n^s)] \neq 0$.

The conditions that $\lim_{n\to\infty} \frac{1}{n}[(Z_n\theta_0)'G_n'P_nM_{Z_n}G_n(Z_n\theta_0) + \sigma_0^2 tr(G_n'P_nG_n)] \neq 0$ and $\lim_{n\to\infty} \frac{1}{n}[(Z_n\theta_0)'P_nM_{Z_n}G_n(Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(P_n^sG_n^s)] \neq 0$ guarantee that a_n/n and b_n/n do not converge to zero in probability, respectively. Let $H_n(\rho) = G_n'(\rho) - \frac{tr(G_n'(\rho)M_{Z_n})}{n-d}M_{Z_n}$, $H_n = H_n(\rho_0)$,[12] and $f(P_n) = (Z_n\theta_0)'P_nM_{Z_n}G_n(Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(P_n^sG_n^s) = (Z_n\theta_0)'P_nM_{Z_n}H_n'(Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(P_n^sH_n^s)$. The sign of $f(P_n)$ depends on the correlation between P_n and H_n. If $P_n = H_n$, then $f(H_n) \geq 0$ and Eq. (20) is the consistent root when $a_n/n \neq o_P(1)$. By continuity, $f\big(H_n(\rho)\big)$ is non-negative when ρ is close to ρ_0. In empirical applications, ρ_0 is often positive, then $P_n = H_n(0.5)$ or $P_n = H_n(0) = W_n' - \frac{tr(W_n'M_{Z_n})}{n-d}M_{Z_n}$ could generate a consistent root estimator of the form Eq. (20). Given P_n, the scalars a_n, b_n and c_n are products of vectors and matrices, so the computational cost of Eq. (20) or (21) is minimal.

The asymptotic distribution of the consistent root $\hat{\rho}_n$ can be derived from a first order expansion of $g_n(\hat{\rho}_n) = 0$ at ρ_0. As $g_n(\rho_0)$ is quadratic in the disturbances, the central limit theorem for linear and quadratic forms is applicable.

[12]Note that using $P_n = G_n' - \frac{tr(G_n'M_{Z_n})}{n-d}M_{Z_n}$ and $P_n = G_n' - \frac{tr(G_n'M_{Z_n})}{n-d}I_n$ generate the same root estimator. We use H_n for narrative convenience, but we may use $G_n' - \frac{tr(G_n'M_{Z_n})}{n-d}I_n$ when calculating a root estimate.

Proposition 2. *The consistent root $\hat{\rho}_n$ in Proposition 1 has the asymptotic distribution that*

$$\sqrt{n}(\hat{\rho}_n - \rho_0) \xrightarrow{D} N\big(0, \Omega\big),$$

where $\Omega = V_\rho \Sigma_\rho^{-2}$ *with* $V_\rho = \lim_{n\to\infty} \frac{1}{n}\big\{\sigma_0^2 (Z_n\theta_0)' P_n M_{Z_n} P_n' (Z_n\theta_0) + 2E(\epsilon_{ni}^3)(Z_n\theta_0)' P_n M_{Z_n} Diag(P_n M_{Z_n}) l_n$ $+ [E(\epsilon_{ni}^4) - 3\sigma_0^4] \sum_{i=1}^n p_{n,ii}^2 + \frac{1}{2}\sigma_0^4 tr(P_n^s P_n^s)\big\}$ *and* $\Sigma_\rho = \lim_{n\to\infty} \frac{1}{n}[(Z_n\theta_0)' P_n M_{Z_n} G_n (Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(P_n^s G_n^s)]$ *being assumed to exist and be non-zero.*

The V_ρ in the above proposition is the limit of the variance of $\frac{1}{\sqrt{n}} g_n(\rho_0)$, so it is generally positive.

When $E(\epsilon_{ni}^3) = E(\epsilon_{ni}^4) - 3\sigma_0^4 = 0$, e.g., ϵ_{ni}'s are i.i.d. normal, the asymptotic variance of $\hat{\rho}_n$ reduces to $\Omega = \lim_{n\to\infty} n \frac{\sigma_0^2 (Z_n\theta_0)' P_n M_{Z_n} P_n' (Z_n\theta_0) + \frac{1}{2}\sigma_0^4 tr(P_n^s P_n^s)}{[(Z_n\theta_0)' P_n M_{Z_n} G_n (Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(P_n^s G_n^s)]^2}$. Then, by applying the Cauchy inequality, H_n is the best P_n matrix such that the asymptotic variance of this consistent root estimator is the smallest. As pointed out earlier, with the best $P_n (= H_n)$ matrix, the consistent root estimator has the form $(b_n - \sqrt{b_n^2 - 4a_n c_n})/(2a_n)$ when $a_n/n \neq o_P(1)$.

Proposition 3. *When* $E(\epsilon_{ni}^3) = E(\epsilon_{ni}^4) - 3\sigma_0^4 = 0$, *suppose that* $\lim_{n\to\infty} \frac{1}{n}\{(Z_n\theta_0)' G_n M_{Z_n} G_n (Z_n\theta_0)$ $+ \frac{1}{2}\sigma_0^2 [tr(G_n^s G_n^s) - \frac{1}{n} tr^2(G_n^s)]\}$ *exists and is non-zero, the best root estimator is*

$$\hat{\rho}_{b,n} = \frac{b_n - \sqrt{b_n^2 - 4a_n c_n}}{2a_n}, \tag{22}$$

where $a_n = y_n' W_n' H_n M_{Z_n} W_n y_n$, $b_n = y_n' (H_n M_{Z_n})^s y_n$ *and* $c_n = y_n' H_n M_{Z_n} y_n$, *in the sense that* $\sqrt{n}(\hat{\rho}_{b,n} - \rho_0) \xrightarrow{D} N(0, \Omega_b)$ *with* $\Omega_b \leq \Omega$, *where* $\Omega_b = \sigma_0^2 \big\{\lim_{n\to\infty} \frac{1}{n}\big[(Z_n\theta_0)' G_n M_{Z_n} G_n (Z_n\theta_0) + \frac{1}{2}\sigma_0^2 \big(tr(G_n^s G_n^s) - \frac{1}{n} tr^2(G_n^s)\big)\big]\big\}^{-1}$.

When $E(\epsilon_{ni}^3) = E(\epsilon_{ni}^4) - 3\sigma_0^4 = 0$, the asymptotic variance Ω_b for the best root estimator in the above proposition is the same as that for the QMLE (Lee, 2004a). When the condition $E(\epsilon_{ni}^3) = E(\epsilon_{ni}^4) - 3\sigma_0^4 = 0$ does not hold, the root estimator $\hat{\rho}_{b,n}$ may lose efficiency. Note that no matter whether the condition holds or not, $\hat{\rho}_{b,n}$ in the above proposition is the consistent root estimator when H_n is used as the P_n matrix.

As H_n involves the unknown parameter ρ_0, it can be estimated by using an initial consistent estimator for ρ_0. An estimated H_n would generate a root estimator with the same limiting distribution as $\hat{\rho}_{b,n}$.

Proposition 4. *Suppose that $\hat{\rho}_n$ is a $\sqrt{n}$-consistent estimator of ρ_0, and* $\lim_{n\to\infty} \frac{1}{n}[(Z_n\theta_0)' G_n' H_n M_{Z_n} G_n (Z_n\theta_0) + \sigma_0^2 tr(G_n' H_n G_n)] \neq 0$. *Then the root estimator*

$$\tilde{\rho}_{b,n} = \frac{\hat{b}_n - \sqrt{\hat{b}_n^2 - 4\hat{a}_n \hat{c}_n}}{2\hat{a}_n}, \tag{23}$$

where $\hat{a}_n = y_n' W_n' H_n(\hat{\rho}_n) M_{Z_n} W_n y_n$, $\hat{b}_n = y_n' \big(H_n(\hat{\rho}_n) M_{Z_n}\big)^s W_n y_n$ *and* $\hat{c}_n = y_n' H_n(\hat{\rho}_n) M_{Z_n} y_n$, *is consistent and has the same limiting distribution as* $\hat{\rho}_{b,n}$.

An initial consistent estimator $\hat{\rho}_n$ may be derived by using $H_n(0)$ as the P_n matrix. Based on $H_n(\hat{\rho}_n)$, Eq. (23) is the best root estimator when $E(\epsilon_{ni}^3) = E(\epsilon_{ni}^4) - 3\sigma_0^4 = 0$.[13] We shall use the notation RE_{sd} for this root estimator based on the SD model. Replacing Z_n with X_n everywhere above, we obtain the root estimator RE_{sar} specifically for the SAR model.

Note that the expression for $H_n(\rho)$ involves a matrix inverse $(I_n - \rho W_n)^{-1}$, which is computationally intensive for large sample sizes.[14] If $||\rho W_n|| < 1$ with a matrix norm $||\cdot||$, then we have the expansion $(I_n - \rho W_n)^{-1} = I_n + \rho W_n + \rho^2 W_n^2 + \dots$ and $||(I_n - \rho W_n)^{-1} - [I_n + \rho W_n + \cdots + \rho^r W_n^r]|| = ||\rho^{r+1} W_n^{r+1}(I_n - \rho W_n)^{-1}|| < ||\rho W_n||^{r+1}/(1 - ||\rho W_n||)$. In the second step of computing a root estimate, we may start from using a few term approximation $(I_n + \hat{\rho}_n W_n' + \cdots + \hat{\rho}_n^r W_n'^r)W_n' - \frac{tr[(I_n + \hat{\rho}_n W_n' + \cdots + \hat{\rho}_n^r W_n'^r)W_n' M_{Z_n}]}{n-d} M_{Z_n}$ of $H_n(\hat{\rho}_n)$ in Eq. (23). If the change of the root estimate in absolute value is smaller than a chosen tolerance level, we can stop and report the estimate; otherwise, we may use $(r+1)$ term approximation of $H_n(\rho)$ and also use the newly computed estimate from the last step in computing a new $\hat{\rho}_n$ using Eq. (23). We could use more and more terms to approximate $H_n(\rho)$ until the tolerance criterion is met. This procedure turns out to be very efficient in our Monte Carlo study.

3.2. A Root Estimator: Heteroskedastic Case

When there is unknown heteroskedasticity in disturbances, from Eq. (14), the expectation of the left-hand side of the moment equation Eq. (19) over n is $\frac{1}{n} tr(P_n M_{Z_n} \Sigma_n)$, which generally does not converge to zero even if $tr(P_n M_{Z_n}) = 0$. In order to derive a consistent root estimator from solving Eq. (19), we require $P_n M_{Z_n}$ to have a zero diagonal, so that the expectation of the left-hand side of Eq. (19) at ρ_0 would be zero.[15]

---

[13] Because an initial consistent estimator $\hat{\rho}_n$, e.g., derived with $H_n(0)[= W_n' - \frac{tr(W_n' M_{Z_n})}{n-d} M_{Z_n}]$, has a closed form expression, the feasible two step root estimator will also have a closed form expression as the analytical expression of the initial $\hat{\rho}_n$ can be substituted into $H_n(\hat{\rho}_n)$ in its derivation. Li et al. (2007) have emphasized on closed form statistics for exploratory analyses. They propose the *APLE* but dismiss Ord's quadratic root estimator because of the need to solve the quadratic moment equation (as well as its possible inefficiency). They have overlooked possible analytical solutions of a quadratic equation.

[14] The best GMM estimator in Lee (2007a) or Lee (2007b) also involves this matrix inverse.

[15] As in the homoskedastic case, if P_n, instead of $P_n M_{Z_n}$, is required to have a zero diagonal, then we can still obtain a consistent GMM estimator, since the expectation of the moment equation over n at ρ_0 converges to zero as n goes to infinity in this case. We require $P_n M_{Z_n}$ to have a zero diagonal so that better small sample properties may be obtained. Let $P_n = [P_{n1}, \dots, P_{nn}]'$ and $M_{Z_n} = [M_{n1}, \dots, M_{nn}]$, where P_{ni} and M_{ni} are n-dimensional vectors, then $P_n M_{Z_n}$ has a zero diagonal means that P_n satisfies $P_{ni}' M_{ni} = 0$, $i = 1, \dots, n$. So we have many choices of the P_n matrix. In particular, given any n-dimensional square matrix A_n, we may let $P_n = A_n - Diag(A_n M_{Z_n})[Diag(M_{Z_n})]^{-1}$ or $P_n = A_n - Diag(A_n M_{Z_n})[Diag(M_{Z_n})]^{-1} M_{Z_n}$, if every diagonal element of M_{Z_n} is non-zero. In the case that some diagonal elements of M_{Z_n} are zero, we may simply let $P_n = A_n - Diag(A_n)$, or adjust A_n to be A_n^* such that the corresponding diagonal elements of $A_n^* M_{Z_n}$ are zero and let $P_n = A_n^* - Diag(A_n^* M_{Z_n})[Diag(M_{Z_n})]^{-}$ or $P_n = A_n^* - Diag(A_n^* M_{Z_n})[Diag(M_{Z_n})]^{-} M_{Z_n}$, where B^- denotes a generalized matrix inverse for a matrix B.

Assumption 5. *The constant n-dimensional square matrices $\{P_n\}$, which satisfy that $P_n M_{Z_n}$ has a zero diagonal, are bounded in both row and column sum norms.*

We make the following assumption about the unknown heteroskedasticity.

Assumption 6. *$\epsilon_{ni}'s$ in $\epsilon_n = (\epsilon_{n1}, \ldots, \epsilon_{nn})'$ are independent $(0, \sigma_{ni}^2)$ and the moments $E|\epsilon_{ni}^{4+\eta}|$ for some $\eta > 0$ exist and are uniformly bounded for all n and i.*

The consistent root is described in the following proposition. The regularity conditions are similar to those in Proposition 1 after taking into account the heteroskedastic variance matrix Σ_n.

Proposition 5. *Under Assumptions 2, 3, 5 and 6, if $(Z_n\theta_0)' P_n M_{Z_n} G_n (Z_n\theta_0) + tr(\Sigma_n P_n^s G_n)$ were non-negative, the consistent root would be*

$$\hat{\rho}_{1n} = \frac{b_n - \sqrt{b_n^2 - 4a_n c_n}}{2a_n}, \tag{24}$$

where $a_n = y_n' W_n' P_n M_{Z_n} W_n y_n$, $b_n = y_n'(P_n M_{Z_n})^s W_n y_n$ and $c_n = y_n' P_n M_{Z_n} y_n$; if $(Z_n\theta_0)' P_n M_{Z_n} G_n (Z_n\theta_0)$ $+ tr(\Sigma_n P_n^s G_n)$ were negative, the consistent root would be

$$\hat{\rho}_{2n} = \frac{b_n + \sqrt{b_n^2 - 4a_n c_n}}{2a_n}, \tag{25}$$

when $\lim_{n\to\infty} \frac{1}{n}[(Z_n\theta_0)' G_n' P_n M_{Z_n} G_n (Z_n\theta_0) + tr(\Sigma_n G_n' P_n G_n)] \neq 0$. In the case that $\lim_{n\to\infty} \frac{1}{n}[(Z_n\theta_0)' G_n' P_n M_{Z_n} G_n (Z_n\theta_0) + tr(\Sigma_n G_n' P_n G_n)] = 0$, $\hat{\rho}_{3n} = c_n/b_n$ is the unique consistent root if $\lim_{n\to\infty} \frac{1}{n}[(Z_n\theta_0)' P_n M_{Z_n} G_n (Z_n\theta_0) + tr(\Sigma_n P_n^s G_n)] \neq 0$.

The conditions that $\lim_{n\to\infty} \frac{1}{n}[(Z_n\theta_0)' G_n' P_n M_{Z_n} G_n (Z_n\theta_0) + tr(\Sigma_n G_n' P_n G_n)] \neq 0$ and $\lim_{n\to\infty} \frac{1}{n}[(Z_n\theta_0)' P_n M_{Z_n} G_n (Z_n\theta_0) + tr(\Sigma_n P_n^s G_n)] \neq 0$ are equivalent to none-zero probability limits of a_n/n and b_n/n, respectively.

Proposition 6. *The consistent root $\hat{\rho}_n$ in Proposition 5 has the asymptotic distribution that*

$$\sqrt{n}(\hat{\rho}_n - \rho_0) \xrightarrow{D} N(0, \Omega),$$

where $\Omega = V_\rho \Sigma_\rho^{-2}$ with $V_\rho = \lim_{n\to\infty} \frac{1}{n}[(Z_n\theta_0)' P_n M_{Z_n} \Sigma_n M_{Z_n} P_n' (Z_n\theta_0) + tr(\Sigma_n P_n \Sigma_n P_n^s)]$ and $\Sigma_\rho = \lim_{n\to\infty} \frac{1}{n}[(Z_n\theta_0)' P_n M_{Z_n} G_n (Z_n\theta_0) + tr(\Sigma_n P_n^s G_n)]$ being assumed to exist and be non-zero.

Note that V_ρ is the limit of the variance of $\frac{1}{\sqrt{n}} g_n(\rho_0)$. As contrary to the homogenous variance case, the third and fourth moments of non-normal disturbances do not play a role in the asymptotic variance of the estimator due to the design of $P_n M_{Z_n}$ having a zero diagonal.

Since the asymptotic variance of the consistent root in the above proposition involves unknown heteroskedasticity terms, the best selection of the matrix P_n may be unavailable. A possible choice of P_n in practice might be the consistently estimated $G_n' - Diag(G_n' M_{Z_n})[Diag(M_{Z_n})]^{-1}$, if none of the diagonal elements of M_{Z_n} is zero. To get an estimator for G_n, we may first derive an initial consistent estimator $\hat{\rho}_n$ for ρ_0 based on the moment equation $y_n' S_n'(\rho)\{W_n' - Diag(W_n' M_{Z_n})[Diag(M_{Z_n})]^{-1}\} M_{Z_n} S_n(\rho) y_n = 0$. Then using $G_n'(\hat{\rho}_n) - Diag[G_n'(\hat{\rho}_n) M_{Z_n}][Diag(M_{Z_n})]^{-1}$ as the P_n matrix in the moment equation, we derive the root estimator $\hat{\rho}_{1n}$. We shall call this root estimator RE_{sd}. The root estimator specifically for the SAR model is denoted by RE_{sar}.

4. Monte Carlo Study

We conduct some Monte Carlo experiments to investigate finite sample performances and computing times of the QMLE and various versions of RE, $APLE$ and $ACME$. The DGP is either the SAR or SE model. For the QMLE, the likelihood function is derived as follows: ignore (unknown) heteroskedasticity even though there might be and form the likelihood based on the SAR model if the DGP is the SAR process or based on the SE model if the DGP is the SE process. For RE_{sd}, in the homoskedastic case, the initial estimator $\hat{\rho}_n$ is the root $\hat{\rho}_{1n}$ of Eq. (19) with $P_n = W_n' - \frac{tr(W_n' M_{Z_n})}{n-d} I_n$, and RE_{sd} denotes the corresponding root estimator Eq. (23) in Proposition 4 with $P_n = G_n'(\hat{\rho}_n) - \frac{tr[G_n'(\hat{\rho}_n) M_{Z_n}]}{n-d} I_n$; in the heteroskedastic case, the initial estimator is the root $\hat{\rho}_{1n}$ of Eq. (24) with $P_n = W_n' - Diag(W_n' M_{Z_n})[Diag(M_{Z_n})]^{-1}$, and RE_{sd} denotes the corresponding root estimator $\hat{\rho}_{1n}$ of Eq. (24) with $P_n = G_n'(\hat{\rho}_n) - Diag[G_n'(\hat{\rho}_n) M_{Z_n}][Diag(M_{Z_n})]^{-1}$. The root estimate RE_{sar} specific for the SAR model is derived similarly.

We consider three different spatial weights matrices W_{1n}, W_{2n} and W_{3n}. The W_{1n} is the "circular world matrix" considered in Arraiz et al. (2010). Specifically, each of the first $n/3$ and last $n/3$ rows except the first and last rows only has two non-zero elements,[16] which are in the positions $(i, i-1)$ and $(i, i+1)$ and are equal to 0.5. For the first row, the non-zero elements are in the positions $(1, 2)$ and $(1, n)$ and they are equal to 0.5; for the last row, the non-zero elements are in the positions $(n, 1)$ and $(n, n-1)$ and they are also equal to 0.5. Each of the middle $n/3$ rows has 10 non-zero elements, which are in the positions $(i, i-5)$, ..., $(i, i-1)$, $(i, i+1)$,..., $(i, i+5)$ and are equal to 0.1. The W_{2n} and W_{3n} are generated according to, respectively, the queen and rook criteria on regular $m \times m$ grids, leading to a sample size of $n = m^2$. We use the row-normalized W_{2n} and W_{3n}. The exogenous variable matrix X_n consists of an intercept term, an exogenous variable drawn from the normal distribution $N(3, 1)$, and the third one drawn from the uniform distribution $U(-1, 2)$. The true parameter vector corresponding to these exogenous variables is $\beta_0 = (0.8, 0.2, 1.5)'$. The design of exogenous variables and corresponding parameters has been used in Lin and Lee (2010).

For the homoskedastic case, the error terms are randomly drawn from the normal distribution $N(0, 0.5^2)$. For the heteroskedastic case, two designs of heteroskedasticity are considered:

[16] When $n/3$ is not an integer, the smallest integer larger than $n/3$ is taken.

- Heteroskedasticity design 1 (HD-1): For W_{1n}, the standard deviation (STD) is equal to a constant times the number of non-zero elements in each row;[17] for W_2 and W_3, the STD is equal to a constant times the absolute value of the second exogenous variable.[18] The constants are chosen such that the average STD is equal to 0.5.

- Heteroskedasticity design 2 (HD-2): For W_{1n}, the STD is equal to a constant times the inverse of the number of non-zero elements in each row; for W_{2n} and W_{3n}, the STD is equal to a constant times the inverse of the absolute value of the second exogenous variable. Again the constants are chosen to make the average STD be equal to 0.5.

We calculate various measures of the autoregressive coefficient by focusing on non-negative ρ_0 values, as this is usually the case in empirical applications. For each case of ρ_0, the number of repetitions is 2000.

Figs. 1—6 compare the mean, STD and root mean square error (RMSE) of the QMLE and different versions of RE, $APLE$ and $ACME$. For these figures, we have $n = 400$.

Figs. 1 and 2 are the case when there is no unknown heteroskedasticity in the disturbances. When the DGP is the SAR process, from Fig. 1, the QMLE and RE_{sar} have similarly small bias (in absolute value) for different spatial weights matrices and ρ_0's, while the biases of $APLE_{sar}$, $ACME_{sar}$, $APLE_{sd}$ and $ACME_{sd}$ are only small when ρ_0 is close to zero and generally increase as ρ_0 increases. In terms of bias, $APLE_{sar}$ has not shown an advantage over $APLE_{sd}$, though $APLE_{sar}$ is based on the DGP. The $APLE_{sar}$ and $APLE_{sd}$ have similar bias for W_{1n} and W_{2n}, but $APLE_{sar}$ has large bias for large ρ_0's in the case of W_{2n}. The QMLE, RE_{sar}, $APLE_{sar}$ and $ACME_{sar}$ have similar STD that is smaller than those of RE_{sd}, $APLE_{sd}$ and $ACME_{sd}$, which is expected since the latter ones are based on the more general SD model. It is noted that for W_{1n}, the bias of $ACME_{sd}$ is significantly larger than that of other statistics. The RMSEs of different statistics show similar patterns as their biases. When the DGP is the SE process, the bias, STD, RMSE of the QMLE, RE_{sd}, $APLE_{se}$, $APLE_{sd}$ and $ACME_{sd}$ are plotted in Fig. 2. The biases of statistics other than $APLE_{se}$ have similar patterns as the corresponding ones in Fig. 1. The $APLE_{se}$ have not shown an advantage over $APLE_{sd}$ in terms of smaller bias, which is obvious for W_{2n} for which $APLE_{se}$ usually has larger bias than $APLE_{sd}$. The STDs of all statistics are very similar. The $APLE_{se}$, $APLE_{sd}$ and $ACME_{sd}$ have larger RMSEs than the QMLE and RE_{sd} for W_{1n} and W_{3n}, while all statistics have similar RMSEs for W_{2n}.

Figs. 3—6 show the results when there is unknown heteroskedasticity in the disturbances. Figs. 3 and 4 correspond to the DGP being the SAR process but with different designs of heteroskedasticity, and Figs. 5 and 6 correspond to the DGP being the SE process with different variances. In general, RE_{sar} and RE_{sd}

[17] This design is one used in Arraiz et al. (2010).

[18] For W_{2n} and W_{3n}, $(m-2)^2$ rows would have the same number of non-zero elements, which is approximately $100[(m-4)/m]\%$ of the total number of rows. If the same heteroskedasticity design as for W_{1n} is used, there would be little heteroskedasticity.

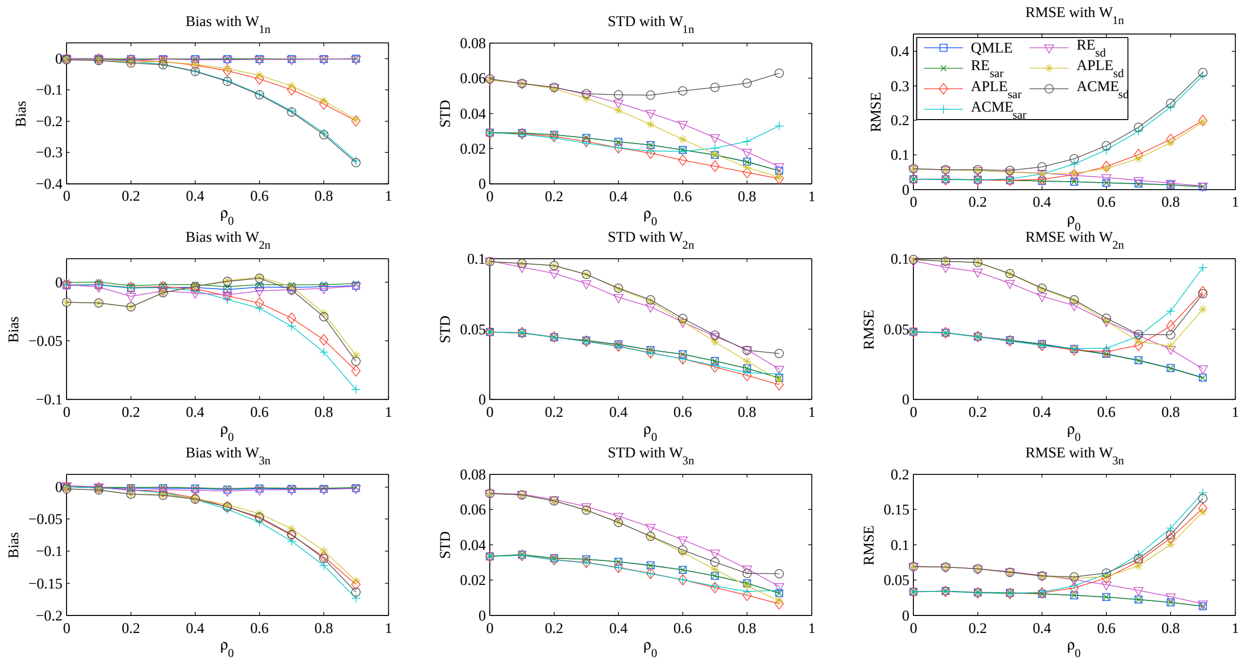

Figure 1: Comparison of the bias, STD and RMSE of the QMLE, RE_{sar}, $APLE_{sar}$, $ACME_{sar}$, RE_{sd}, $APLE_{sd}$ and $ACME_{sd}$ when the DGP is the SAR model under homoskedasticity.

have the smallest bias in Figs. 3 and 4 and RE_{sd} has the smallest bias in Figs. 5 and 6. Since the QMLE has ignored the heteroskedasticity, it may generate large bias in some cases, e.g., its bias is close to 0.2 when $\rho_0 = 0.6$ for W_{1n} in Fig. 5. In most figures, however, the QMLE has relatively small bias. The statistics derived from homoskedastic models—$APLE_{sar}$, $APLE_{se}$ and $APLE_{sd}$—generally have relatively small bias when ρ_0 is small and relatively large bias when ρ_0 is large. We note that for W_{1n} in Fig. 5, both $APLE_{se}$ and $APLE_{sd}$ have very large bias for positive ρ_0's. In Figs. 3 and 4, like $APLE_{sar}$ and $APLE_{sd}$, $ACME_{sar}$ and $ACME_{sd}$ have large bias for large ρ_0's; in Figs. 5 and 6, $ACME_{sd}$ have large bias for large ρ_0's except for the case with W_{1n} in Fig. 5, where the bias of $ACME_{sd}$ is smaller than those of the QMLE, $APLE_{se}$ and $APLE_{se}$. The STDs and RMSEs in Figs. 3 and 4 are similar to the corresponding ones in Fig. 1, and the STDs and RMSEs in Figs. 5 and 6 are similar to the corresponding ones in Fig. 2.

Table 1 compares the computing times and finite sample properties of different statistics when the sample size is large. The DGP is the SAR model. We focus on the QMLE, RE_{sar} and $APLE_{sar}$, as computing the other statistics above are expected to take similar time. To compute RE_{sar}, we use the procedure described in the last paragraph of Subsection 3.1 which starts from using a two term approximation $(I_n + \hat{\rho}_n W_n' + \hat{\rho}_n^2 W_n'^2)W_n' - \frac{tr[(I_n+\hat{\rho}_n W_n' + \hat{\rho}_n^2 W_n'^2)W_n' M_{Z_n}]}{n-d} M_{Z_n}$ of $H_n(\hat{\rho}_n)$ in Eq. (23). The tolerance criteria for RE_{sar} and the QMLE are both set to be 0.0001. The reported results are from Matlab on a desktop computer with Intel Core i7-2600 processor and 8 gigabyte memory. For the same sample size and spatial

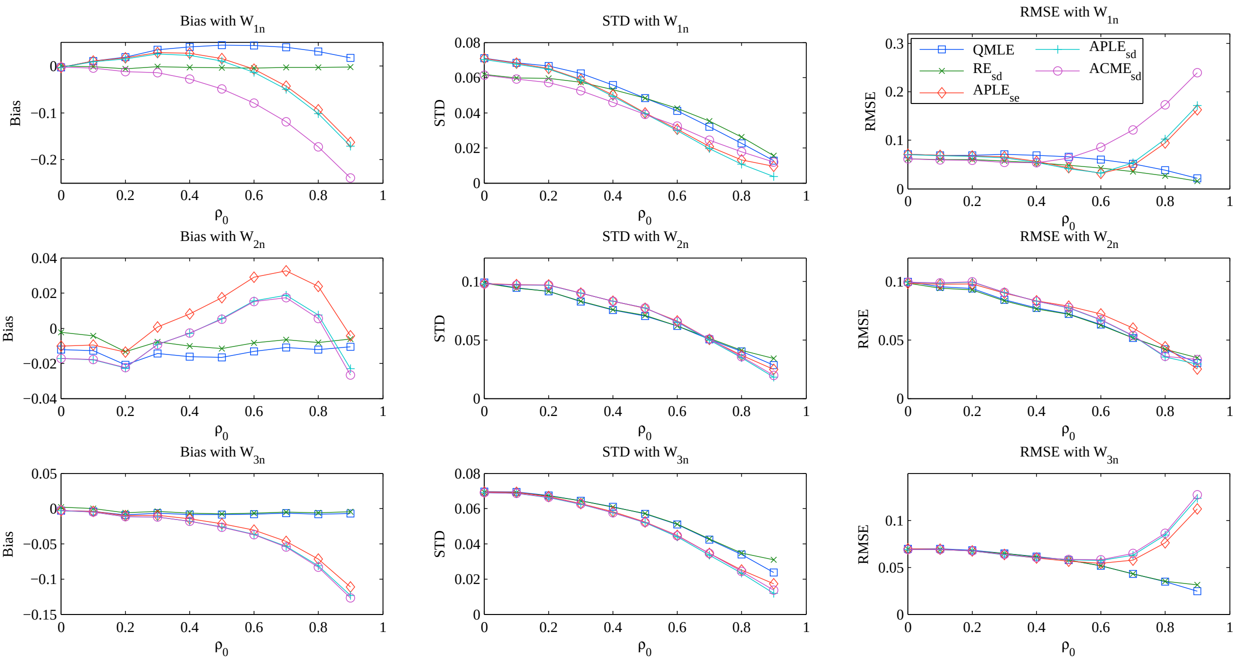

Figure 2: Comparison of the bias, STD and RMSE of the QMLE, RE_{sd}, $APLE_{se}$, $APLE_{sd}$ and $ACME_{sd}$ when the DGP is the SE model under homoskedasticity.

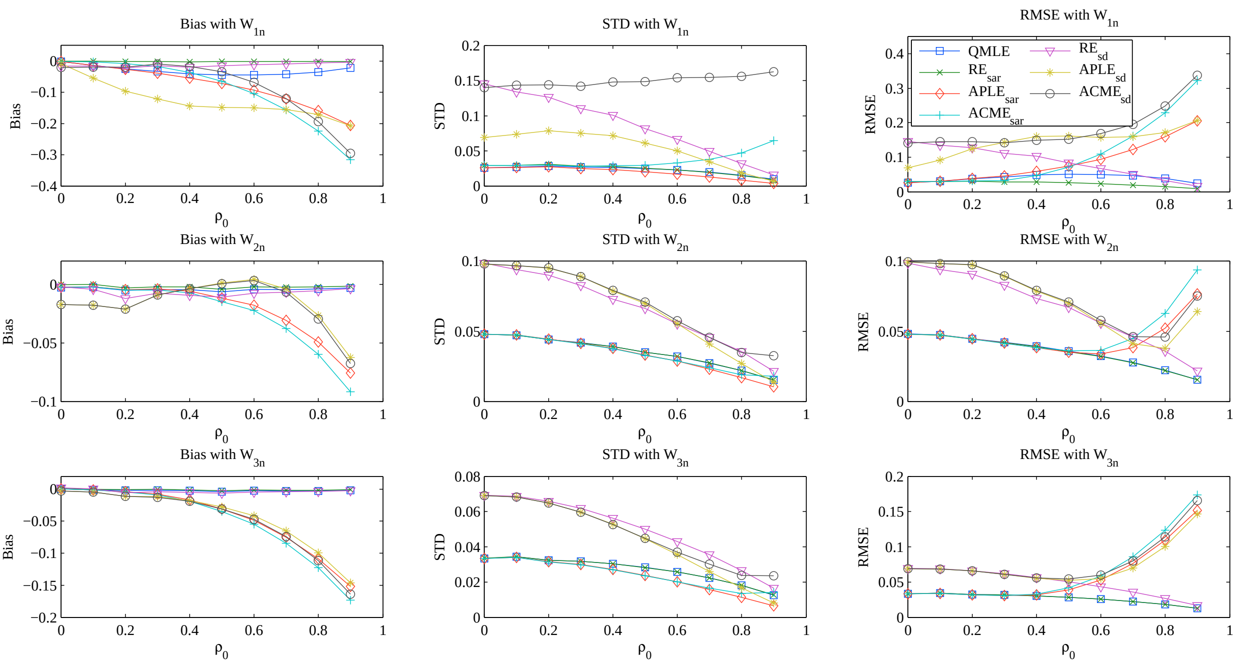

Figure 3: Comparison of the bias, STD and RMSE of the QMLE, RE_{sar}, $APLE_{sar}$, $ACME_{sar}$, RE_{sd}, $APLE_{sd}$ and $ACME_{sd}$ when the DGP is the SAR model under heteroskedasticity (HD-1).

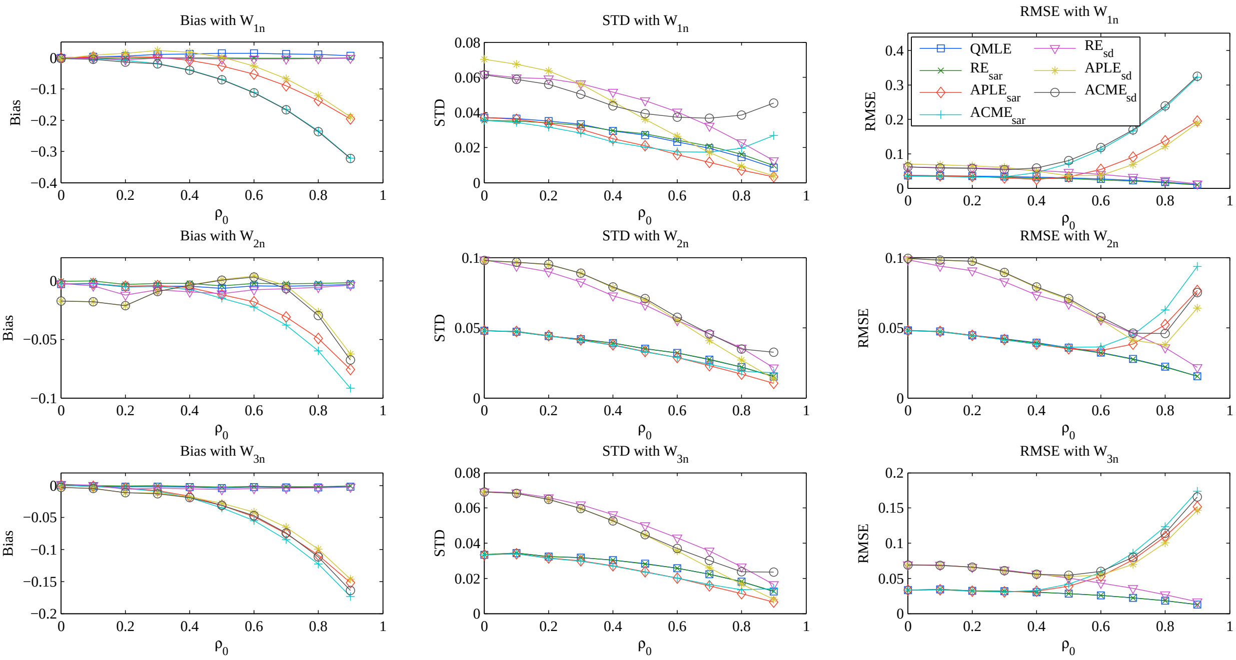

Figure 4: Comparison of the bias, STD and RMSE of the QMLE, RE_{sar}, $APLE_{sar}$, $ACME_{sar}$, RE_{sd}, $APLE_{sd}$ and $ACME_{sd}$ when the DGP is the SAR model under heteroskedasticity (HD-2).

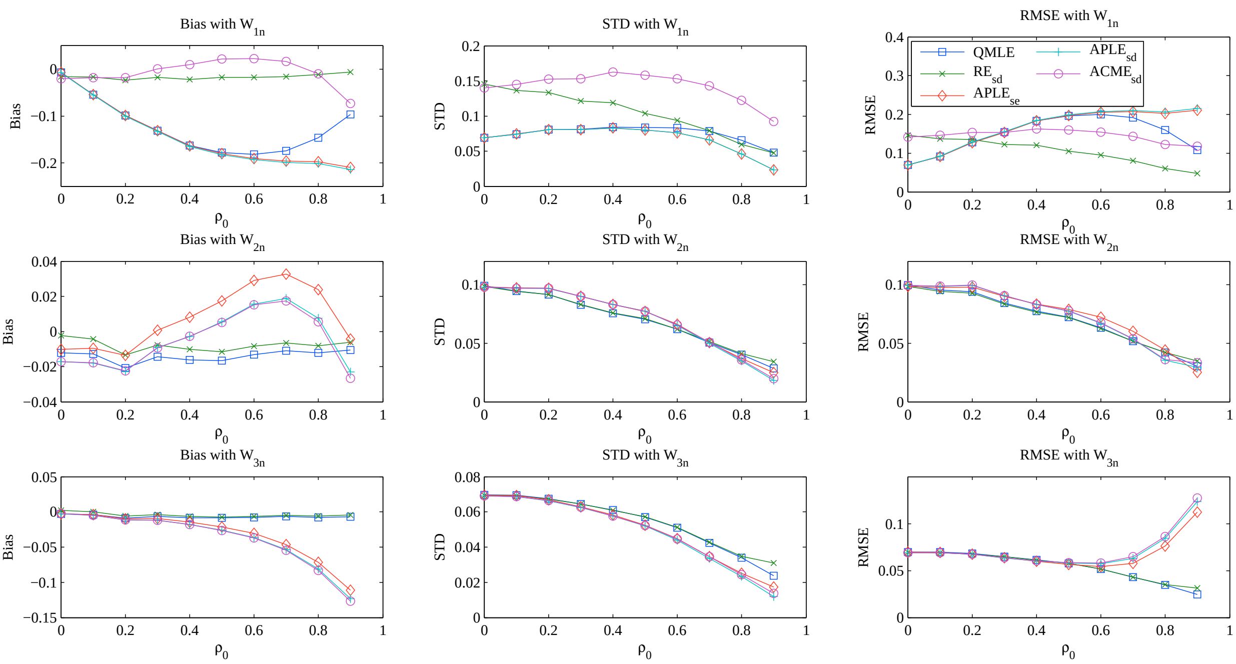

Figure 5: Comparison of the bias, STD and RMSE of the the QMLE, RE_{sd}, $APLE_{se}$, $APLE_{sd}$ and $ACME_{sd}$ when the DGP is the SE model under heteroskedasticity (HD-1).

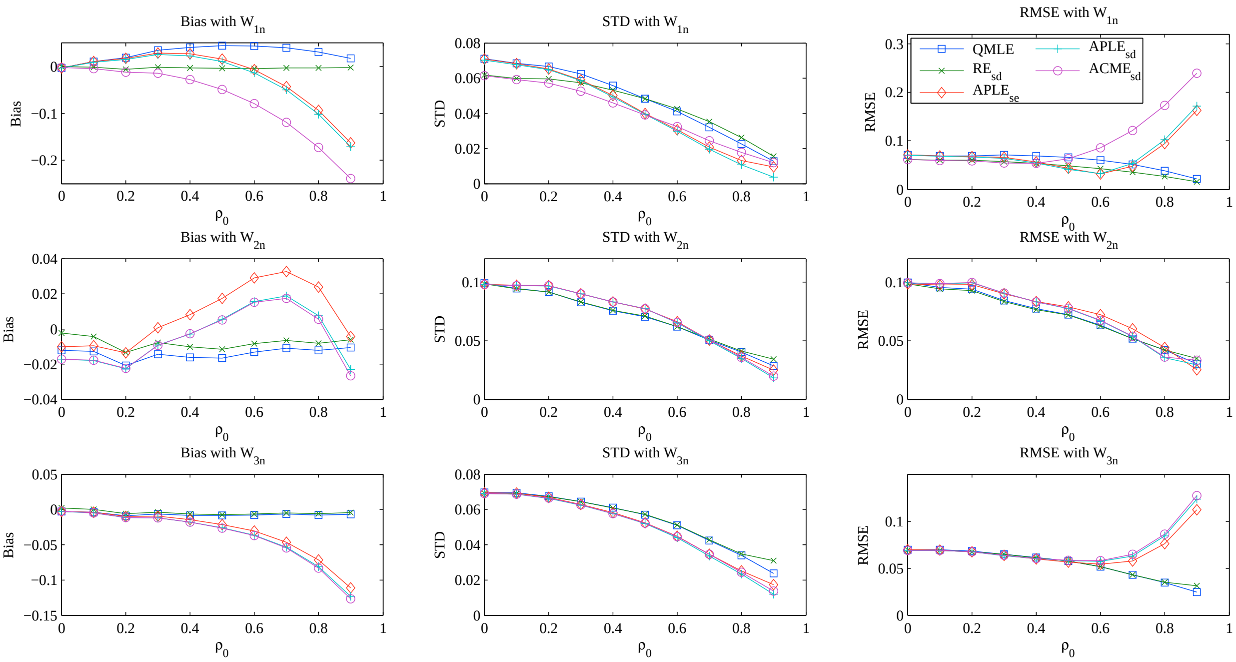

Figure 6: Comparison of the bias, STD and RMSE of the QMLE, RE_{sd}, $APLE_{se}$, $APLE_{sd}$ and $ACME_{sd}$ when the DGP is the SE model under heteroskedasticity (HD-2).

weights matrix in the DGP, while computing an $APLE_{sar}$ takes about the same time for different ρ_0's, computing the QMLE and RE_{sar} take more time when ρ_0 becomes larger. For moderate values of ρ_0, computing RE_{sar} only takes slightly longer time than computing the $APLE_{sar}$ and is at least 8 times faster than computing the QMLE. The bias, STD and RMSE have the same pattern as we have seen in Fig. 1.

5. Conclusion

In this paper, an approximate measure of spatial dependence, the $APLE$, is generalized to the SD model so that exogenous variables are directly taken into account and it may be used to detect spatial dependence originating from either the SAR, SE or SD process. The $APLE$ is derived from a first order approximation of the first order condition for the SD model. Following the first order condition, we further construct a moment condition quadratic in the autoregressive parameter of the SD model which generates a closed-form root estimator. We specify conditions under which a root of the moment equation is consistent. With an initial consistent estimator, a second step root estimator from a properly designed moment equation can be asymptotically as efficient as that of the MLE under normality. Our root estimator involves minimal computational burden. This estimator also applies to the SAR model as it is a constrained form of the SD model.

When there is unknown heteroskedasticity, we adjust the first order condition to derive a modified $APLE$

Table 1: Comparison of the computing time, bias, STD and RMSE of the QMLE, RE_{sar} and $APLE_{sar}$ when the sample size is large.

		n=4900				n=10000			
		$\rho_0 = 0$	0.3	0.6	0.9	0	0.3	0.6	0.9
The spatial weights matrix is W_{1n} in the DGP.									
Time†	QMLE	1.388	1.564	1.747	2.214	5.572	6.390	7.111	9.008
	RE_{sar}	0.188	0.192	0.199	0.212	0.767	0.778	0.791	0.800
	$APLE_{sar}$	0.184	0.183	0.183	0.183	0.756	0.757	0.758	0.755
Bias	QMLE	-2.41E-04	2.03E-05	-2.07E-04	-4.14E-05	1.03E-04	-3.30E-05	-1.55E-04	-9.75E-05
	RE_{sar}	-1.71E-04	8.97E-05	-1.58E-04	-8.67E-06	1.37E-04	4.28E-08	-1.35E-04	-7.80E-05
	$APLE_{sar}$	-2.37E-04	-8.18E-03	-6.45E-02	-2.01E-01	1.05E-04	-8.24E-03	-6.46E-02	-2.02E-01
STD	QMLE	7.95E-03	7.07E-03	5.23E-03	1.85E-03	5.62E-03	5.03E-03	3.54E-03	1.29E-03
	RE_{sar}	7.95E-03	7.07E-03	5.23E-03	1.90E-03	5.62E-03	5.03E-03	3.53E-03	1.36E-03
	$APLE_{sar}$	7.95E-03	6.49E-03	3.61E-03	7.31E-04	5.62E-03	4.61E-03	2.44E-03	5.14E-04
RMSE	QMLE	7.95E-03	7.07E-03	5.23E-03	1.85E-03	5.62E-03	5.03E-03	3.54E-03	1.30E-03
	RE_{sar}	7.95E-03	7.07E-03	5.24E-03	1.90E-03	5.62E-03	5.03E-03	3.53E-03	1.36E-03
	$APLE_{sar}$	7.95E-03	1.04E-02	6.46E-02	2.01E-01	5.62E-03	9.45E-03	6.47E-02	2.02E-01
The spatial weights matrix is W_{2n} in the DGP.									
Time†	QMLE	1.786	1.971	2.219	2.658	6.944	7.586	8.467	10.115
	RE_{sar}	0.197	0.213	0.258	0.391	0.787	0.826	0.889	1.014
	$APLE_{sar}$	0.184	0.184	0.184	0.185	0.760	0.765	0.759	0.760
Bias	QMLE	-4.06E-04	-2.86E-04	-3.59E-04	-1.84E-04	5.34E-05	-4.19E-05	-4.66E-05	-6.04E-05
	RE_{sar}	-2.28E-04	-1.03E-04	-1.85E-04	-4.02E-05	1.39E-04	4.77E-05	3.95E-05	1.57E-05
	$APLE_{sar}$	-4.01E-04	1.42E-05	-1.49E-02	-7.30E-02	5.37E-05	2.55E-04	-1.47E-02	-7.31E-02
STD	QMLE	1.38E-02	1.14E-02	8.52E-03	3.83E-03	9.41E-03	8.07E-03	5.94E-03	2.54E-03
	RE_{sar}	1.38E-02	1.14E-02	8.52E-03	3.88E-03	9.41E-03	8.07E-03	5.94E-03	2.57E-03
	$APLE_{sar}$	1.38E-02	1.13E-02	7.64E-03	2.64E-03	9.41E-03	7.99E-03	5.32E-03	1.76E-03
RMSE	QMLE	1.38E-02	1.14E-02	8.53E-03	3.83E-03	9.41E-03	8.07E-03	5.94E-03	2.54E-03
	RE_{sar}	1.38E-02	1.14E-02	8.52E-03	3.88E-03	9.41E-03	8.07E-03	5.94E-03	2.57E-03
	$APLE_{sar}$	1.38E-02	1.13E-02	1.67E-02	7.31E-02	9.41E-03	7.99E-03	1.56E-02	7.31E-02
The spatial weights matrix is W_{3n} in the DGP.									
Time†	QMLE	1.604	1.795	2.018	2.419	6.249	7.042	7.857	9.410
	RE_{sar}	0.189	0.198	0.221	0.331	0.770	0.786	0.823	0.935
	$APLE_{sar}$	0.183	0.183	0.183	0.183	0.758	0.755	0.754	0.754
Bias	QMLE	-1.95E-04	-5.03E-05	-2.47E-04	-1.86E-04	-2.17E-04	-1.95E-06	-1.94E-04	-1.38E-04
	RE_{sar}	-1.06E-04	5.49E-05	-1.32E-04	-5.41E-05	-1.73E-04	5.05E-05	-1.36E-04	-5.45E-05
	$APLE_{sar}$	-1.95E-04	-5.92E-03	-4.55E-02	-1.47E-01	-2.18E-04	-5.82E-03	-4.53E-02	-1.47E-01
STD	QMLE	9.55E-03	8.97E-03	6.82E-03	3.20E-03	6.75E-03	6.21E-03	4.90E-03	2.17E-03
	RE_{sar}	9.56E-03	8.97E-03	6.84E-03	3.30E-03	6.75E-03	6.21E-03	4.90E-03	2.26E-03
	$APLE_{sar}$	9.56E-03	8.45E-03	5.32E-03	1.69E-03	6.75E-03	5.85E-03	3.82E-03	1.15E-03
RMSE	QMLE	9.55E-03	8.97E-03	6.83E-03	3.20E-03	6.75E-03	6.21E-03	4.90E-03	2.17E-03
	RE_{sar}	9.56E-03	8.97E-03	6.84E-03	3.30E-03	6.76E-03	6.21E-03	4.90E-03	2.26E-03
	$APLE_{sar}$	9.56E-03	1.03E-02	4.58E-02	1.47E-01	6.76E-03	8.25E-03	4.54E-02	1.47E-01

† The average time in seconds to compute an estimate.

statistic, the $ACME$, by a first order approximation. The moment equation can also be modified so that a consistent root estimator under unknown heteroskedasticity is available.

Our Monte Carlo results show that the root estimator has similar bias and STD as the QMLE in the homoskedastic case, and it also has small bias which is generally smaller than those of the QMLE and various versions of $APLE$ which ignore the heteroskedasticity in the unknown heteroskedastic case. Different versions of $APLE$ in the homoskedastic case and $ACME$ in the heteroskedastic case can generate small bias when the autoregressive parameter is not large in magnitude. The $APLE_{sar}$ and $APLE_{se}$ have not shown advantages over $APLE_{sd}$ in terms of smaller bias when the DGPs are, respectively, the SAR and SE models. For moderate true values of the autoregressive parameter, computing a root estimate only takes slightly longer time than computing the $APLE$, and it is much faster than computing the QMLE.

Appendix: Lemmas and Proofs

Lemma 1. *Suppose that $n \times n$ matrices $\{A_n = [a_{n,ij}]\}$ are bounded in both row and column sum norms. Elements of $n \times k$ matrices $\{X_n = [x_{n,ij}]\}$ are uniformly bounded and $\lim_{n\to\infty} \frac{X_n' X_n}{n}$ exists and is nonsingular. Let $M_{X_n} = I_n - X_n(X_n'X_n)^{-1}X_n'$. Then*

(1) matrices $\{M_{X_n}\}$ are bounded in both row and column sum norms,

(2) $tr(M_{X_n}A_n) = tr(A_n) + O(1)$,

(3) $\sum_{i=1}^n (M_{X_n}A_n)_{ii}^2 = \sum_{i=1}^n a_{n,ii}^2 + O(1)$.

Proof. See Lee (2004b). □

Lemmas 2—5 are from, for example, Lin and Lee (2010).[19]

Lemma 2. *Suppose that $A_n = [a_{n,ij}]$ and $B_n = [b_{n,ij}]$ are two square matrices of dimension n and ϵ_{ni}'s in $\epsilon_n = (\epsilon_{n1}, \ldots, \epsilon_{nn})'$ are independently distributed with mean zero (but may not be i.i.d.). Then,*

(1) $E(\epsilon_n \cdot \epsilon_n' A_n \epsilon_n) = (a_{n,11}E(\epsilon_{n1}^3), \ldots, a_{n,nn}E(\epsilon_{nn}^3))'$,

(2) $E[A_n\epsilon_n(B_n\epsilon_n)'] = A_n\Sigma_n B_n'$, and

(3) $E(\epsilon_n' A_n \epsilon_n \cdot \epsilon_n' B_n \epsilon_n) = \sum_{i=1}^n a_{n,ii}b_{n,ii}[E(\epsilon_{ni}^4) - 3\sigma_{ni}^4] + \left(\sum_{i=1}^n a_{n,ii}\sigma_{ni}^2\right)\left(\sum_{i=1}^n b_{n,ii}\sigma_{ni}^2\right) + \sum_{i=1}^n \sum_{j=1}^n a_{n,ij}(b_{n,ij} + b_{n,ji})\sigma_{ni}^2\sigma_{nj}^2 = \sum_{i=1}^n a_{n,ii}b_{n,ii}[E(\epsilon_{ni}^4) - 3\sigma_{ni}^4] + tr(\Sigma_n A_n)tr(\Sigma_n B_n) + tr[\Sigma_n A_n \Sigma_n(B_n + B_n')]$,

where $\Sigma_n = Diag(\sigma_{n1}^2, \ldots, \sigma_{nn}^2)$ with $\sigma_{ni}^2 = E(\epsilon_{ni}^2)$, $i = 1, \ldots, n$.

Lemma 3. *Suppose that n-dimensional square matrices $\{A_n\}$ are bounded in both row and column sum norms and ϵ_{ni}'s in $\epsilon_n = (\epsilon_{n1}, \ldots, \epsilon_{nn})'$ are independent $(0, \sigma_{ni}^2)$. Sequence of the variances $\{\sigma_{ni}^2\}$ and fourth moments $\{E(\epsilon_{ni}^4)\}$ are bounded. Then, $E(\epsilon_n' A_n \epsilon_n) = O(n)$, $var(\epsilon_n' A_n \epsilon_n) = O(n)$, $\epsilon_n' A_n \epsilon_n = O_P(n)$ and $\frac{1}{n}\epsilon_n' A_n \epsilon_n - \frac{1}{n}E(\epsilon_n' A_n \epsilon_n) = o_P(1)$.*

[19] The (3) of Lemma 2 has corrected an error in a matrix expression in Lin and Lee (2010). Lin and Lee (2010) have the right result in the summation form, but an error occurs when the summation is transformed into the trace of matrix products.

Lemma 4. *Suppose that A_n is an $n \times n$ matrix with its column sum norm being bounded, elements of the $n \times k$ matrix C_n are uniformly bounded, and elements ϵ_{ni}'s of $\epsilon_n = (\epsilon_{n1}, \ldots, \epsilon_{nn})'$ are independent $(0, \sigma_{ni}^2)$ with finite third absolute moments, which are uniformly bounded for all n and i. Then $\frac{1}{\sqrt{n}} C_n' A_n \epsilon_n = O_P(1)$ and $\frac{1}{n} C_n' A_n \epsilon_n = o_P(1)$. Furthermore, if the limit of $\frac{1}{n} C_n' A_n \Sigma_n A_n' C_n$ exists and is positive definite, where $\Sigma_n = Diag(\sigma_{n1}^2, \ldots, \sigma_{nn}^2)$, then $\frac{1}{\sqrt{n}} C_n' A_n \epsilon_n \xrightarrow{D} N(0, \lim_{n\to\infty} \frac{1}{n} C_n' A_n \Sigma_n A_n' C_n)$.*

Lemma 5. *Suppose that $\{A_n\}$ is a sequence of symmetric $n \times n$ matrices with row and column sum norms bounded and $b_n = (b_{n1}, \ldots, b_{nn})'$ is an n-dimensional column vector such that $\sup_n \frac{1}{n} \sum_{i=1}^n |b_{ni}|^{2+\eta_1} < \infty$ for some $\eta_1 > 0$. Furthermore, suppose that $\epsilon_{n1}, \cdots, \epsilon_{nn}$ are mutually independent with zero means and the moments $E(|\epsilon_{ni}|^{4+\eta_2})$ for some $\eta_2 > 0$ exist and are uniformly bounded for all n and i.*

Let $\sigma_{Q_n}^2$ be the variance of Q_n where $Q_n = \epsilon_n' A_n \epsilon_n + b_n' \epsilon_n - tr(A_n \Sigma_n)$. Assume that $\frac{1}{n}\sigma_{Q_n}^2$ is bounded away from zero. Then, $\frac{Q_n}{\sigma_{Q_n}} \xrightarrow{D} N(0, 1)$.

Lemma 6. *Suppose that the sequence $\{||S_n^{-1}||\}$, where $||\cdot||$ is a matrix norm, is bounded. Then the sequence $\{S_n^{-1}(\rho)\}$ is uniformly bounded in a neighborhood of ρ_0.*

Proof. See Lee (2004b). □

Proof of Proposition 1. $y_n = S_n^{-1}(Z_n\theta_0 + \epsilon_n)$, so $a_n = (Z_n\theta_0 + \epsilon_n)' A_n (Z_n\theta_0 + \epsilon_n)$, $b_n = (Z_n\theta_0 + \epsilon_n)' B_n (Z_n\theta_0 + \epsilon_n)$ and $c_n = (Z_n\theta_0 + \epsilon_n)' C_n (Z_n\theta_0 + \epsilon_n)$, where $A_n = G_n' P_n M_{Z_n} G_n$, $B_n = G_n' P_n M_{Z_n} S_n^{-1} + S_n'^{-1} P_n M_{Z_n} G_n = G_n' P_n M_{Z_n} + P_n M_{Z_n} G_n + 2\rho_0 A_n$, and $C_n = S_n'^{-1} P_n M_{Z_n} S_n^{-1} = P_n M_{Z_n} + \rho_0 (G_n' P_n M_{Z_n} + P_n M_{Z_n} G_n) + \rho_0^2 A_n$, using the fact that $S_n^{-1} = I_n + \rho_0 G_n$. The M_{Z_n} is bounded in both row and column sum norms by Lemma 1. Then, A_n, B_n and C_n are bounded in both row and column sum norms as G_n, P_n, M_{Z_n} and S_n^{-1} are bounded in both row and column sum norms.

By Lemma 4, $\frac{1}{n}(Z_n\theta_0)' A_n \epsilon_n = o_P(1)$, $\frac{1}{n}(Z_n\theta_0)' B_n \epsilon_n = o_P(1)$ and $\frac{1}{n}(Z_n\theta_0)' C_n \epsilon_n = o_P(1)$. As elements of Z_n are uniformly bounded, we have $a_{1n} \equiv (Z_n\theta_0)' A_n (Z_n\theta_0) = O(n)$, $b_{1n} \equiv (Z_n\theta_0)' B_n (Z_n\theta_0) = O(n)$, and $c_{1n} \equiv (Z_n\theta_0)' C_n (Z_n\theta_0) = O(n)$. The fact $M_{Z_n} Z_n = 0$ can be used to simplify the expressions for b_{1n} and c_{1n}.

As elements of matrices bounded in either row or column sum norms are uniformly bounded, $a_{2n} \equiv E(\epsilon_n' A_n \epsilon_n) = \sigma_0^2 tr(A_n) = O(n)$, $b_{2n} \equiv E(\epsilon_n' B_n \epsilon_n) = \sigma_0^2 tr(B_n) = O(n)$, and $c_{2n} \equiv E(\epsilon_n' C_n \epsilon_n) = \sigma_0^2 tr(C_n) = O(n)$. We can simplify the expression for c_{2n} by using $tr(P_n M_{Z_n}) = 0$. In addition, $\frac{1}{n}\epsilon_n' A_n \epsilon_n = \frac{1}{n} a_{2n} + o_P(1) = O_P(1)$, $\frac{1}{n}\epsilon_n' B_n \epsilon_n = \frac{1}{n} b_{2n} + o_P(1) = O_P(1)$ and $\frac{1}{n}\epsilon_n' C_n \epsilon_n = \frac{1}{n} c_{2n} + o_P(1) = O_P(1)$, by Lemma 3.

Then we have

$$
\begin{aligned}
\frac{1}{n^2} b_n^2 - \frac{4}{n^2} a_n c_n &= \big[\frac{1}{n} b_{1n} + \frac{1}{n} b_{2n} + o_P(1)\big]^2 - 4\big[\frac{1}{n} a_{1n} + \frac{1}{n} a_{2n} + o_P(1)\big]\big[\frac{1}{n} c_{1n} + \frac{1}{n} c_{2n} + o_P(1)\big] \\
&= \big[\frac{1}{n}(Z_n\theta_0)' P_n M_{Z_n} G_n (Z_n\theta_0) + \frac{1}{n}\sigma_0^2 tr(G_n^s P_n M_{Z_n})\big]^2 + o_P(1) \\
&= \big[\frac{1}{n}(Z_n\theta_0)' P_n M_{Z_n} G_n (Z_n\theta_0) + \frac{1}{2n}\sigma_0^2 tr(P_n^s G_n^s)\big]^2 + o_P(1),
\end{aligned}
$$

where the last equation follows by Lemma 1.

1) When $\lim_{n\to\infty}\frac{1}{n}[(Z_n\theta_0)'G_n'P_nM_{Z_n}G_n(Z_n\theta_0)+\sigma_0^2tr(G_n'P_nG_n)]\neq 0$, i.e., $\frac{1}{n}a_n$ does not converge to zero in probability,

$$
\begin{aligned}
\hat\rho_n &\equiv \big[b_n/n-\big((Z_n\theta_0)'P_nM_{Z_n}G_n(Z_n\theta_0)/n+\sigma_0^2tr(P_n^sG_n^s)/(2n)\big)\big]/(2a_n/n)\\
&=\big[(b_{1n}+b_{2n})/n-\big((Z_n\theta_0)'P_nM_{Z_n}G_n(Z_n\theta_0)/n+\sigma_0^2tr(P_n^sG_n^s)/(2n)\big)+o_P(1)\big]/[2(a_{1n}+a_{2n})/n+o_P(1)]\\
&=\rho_0+o_P(1).
\end{aligned}
$$

If $(Z_n\theta_0)'P_nM_{Z_n}G_n(Z_n\theta_0)+\frac{1}{2}\sigma_0^2tr(P_n^sG_n^s)$ were non-negative, $\hat\rho_{1n}=\hat\rho_n+o_P(1)$, thus $\hat\rho_{1n}$ is the consistent root; if $(Z_n\theta_0)'P_nM_{Z_n}G_n(Z_n\theta_0)+\frac{1}{2}\sigma_0^2tr(P_n^sG_n^s)$ were negative, $\hat\rho_{2n}=\hat\rho_n+o_P(1)$, thus $\hat\rho_{2n}$ is the consistent root.

2) When $\frac{1}{n}a_n=o_P(1)$, Eq. (19) over n is linear in ρ asymptotically. As $\lim_{n\to\infty}\frac{1}{n}[(Z_n\theta_0)'P_nM_{Z_n}G_n(Z_n\theta_0)+\frac{1}{2}\sigma_0^2tr(P_n^sG_n^s)]\neq 0$,

$$
\begin{aligned}
\hat\rho_{3n}&=c_n/b_n\\
&=[(c_{1n}+c_{2n})/n+o_P(1)]/[(b_{1n}+b_{2n})/n+o_P(1)]\\
&=\big[\rho_0\big((Z_n\theta_0)'P_nM_{Z_n}G_n(Z_n\theta_0)/n+\sigma_0^2tr(P_n^sG_n^s)/(2n)\big)+\rho_0^2(a_{1n}+a_{2n})/n\big]/\big[\big((Z_n\theta_0)'P_nM_{Z_n}G_n(Z_n\theta_0)/n\\
&\quad+\sigma_0^2tr(P_n^sG_n^s)/(2n)\big)+2\rho_0(a_{1n}+a_{2n})/n\big]+o_P(1)\\
&=\rho_0+o_P(1).
\end{aligned}
$$

□

Proof of Proposition 2. We still use the notations in the proof of Proposition 1. As $g_n(\rho)=y_n'(I_n-\rho W_n')P_nM_{Z_n}(I_n-\rho W_n)y_n=a_n\rho^2-b_n\rho+c_n$, $\frac{\partial}{\partial\rho}g_n(\rho)=2a_n\rho-b_n$. $g_n(\rho_0)=(Z_n\theta_0+\epsilon_n)'P_nM_{Z_n}(Z_n\theta_0+\epsilon_n)=(Z_n\theta_0)'P_nM_{Z_n}\epsilon_n+\epsilon_n'P_nM_{Z_n}\epsilon_n$. According to the mean value theorem, $0=g_n(\hat\rho_n)=g_n(\rho_0)+\frac{\partial g_n(\bar\rho_n)}{\partial\rho}(\hat\rho_n-\rho_0)$, where $\bar\rho_n$ is between ρ_0 and $\hat\rho_n$. Then $\sqrt{n}(\hat\rho_n-\rho_0)=-[\frac{1}{n}(2a_n\bar\rho_n-b_n)]^{-1}\frac{1}{\sqrt{n}}[(Z_n\theta_0)'P_nM_{Z_n}\epsilon_n+\epsilon_n'P_nM_{Z_n}\epsilon_n]$.

From the proof of Proposition 1, $\frac{1}{n}(2a_n\bar\rho_n-b_n)=\frac{1}{n}(2\rho_0a_n-b_n)+2(\bar\rho_n-\rho_0)\frac{1}{n}a_n=-\frac{1}{n}[(Z_n\theta_0)'P_nM_{Z_n}G_n(Z_n\theta_0)+\frac{1}{2}\sigma_0^2tr(P_n^sG_n^s)]+o_P(1)=-\Sigma_{\rho,n}+o_P(1)$.

As $E[\frac{1}{\sqrt{n}}(Z_n\theta_0)'P_nM_{Z_n}\epsilon_n+\frac{1}{\sqrt{n}}\epsilon_n'P_nM_{Z_n}\epsilon_n]=\sigma_0^2\frac{1}{\sqrt{n}}tr(P_nM_{Z_n})=0$, by Lemma 5,

$$\big[(Z_n\theta_0)'P_nM_{Z_n}\epsilon_n/\sqrt{n}+\epsilon_n'P_nM_{Z_n}\epsilon_n/\sqrt{n}\big]/V_{\rho,n}^{\frac{1}{2}}\xrightarrow{D}N(0,1),$$

where, by Lemmas 1 and 2,

$$
\begin{aligned}
V_{\rho,n}&=var\big(\frac{1}{\sqrt{n}}(Z_n\theta_0)'P_nM_{Z_n}\epsilon_n+\frac{1}{\sqrt{n}}\epsilon_n'P_nM_{Z_n}\epsilon_n\big)\\
&=\frac{1}{n}\sigma_0^2(Z_n\theta_0)'P_nM_{Z_n}P_n'(Z_n\theta_0)+\frac{2}{n}E(\epsilon_{ni}^3)(Z_n\theta_0)'P_nM_{Z_n}Diag(P_nM_{Z_n})l_n\\
&\quad+\frac{1}{n}[E(\epsilon_{ni}^4)-3\sigma_0^4]\sum_{i=1}^n p_{n,ii}^2+\frac{1}{2n}\sigma_0^4tr(P_n^sP_n^s)+o(1).
\end{aligned}
$$

Therefore, $\sqrt{n}(\hat{\rho}_n - \rho_0) \xrightarrow{D} N(0, \Omega)$ with

$$\Omega = \lim_{n\to\infty}(V_{\rho,n}\Sigma_{\rho,n}^{-2}) = V_\rho \Sigma_\rho^{-2}. \quad \square$$

Proof of Proposition 3. For any $n \times n$ symmetric matrix $A_n = [A_{1n}, \ldots, A_{nn}]$, where A_{in}'s are column vectors, $tr(A_n^2) = tr(A_n' A_n) = \sum_{i=1}^n A_{in}' A_{in}$. In addition, $(Z_n\theta_0)' P_n M_{Z_n} P_n'(Z_n\theta_0) = [M_{Z_n} P_n'(Z_n\theta_0)]' M_{Z_n} P_n'(Z_n\theta_0)$ and $(Z_n\theta_0)' H_n M_{Z_n} H_n' Z_n\theta_0 = (M_{Z_n} H_n' Z_n\theta_0)' M_{Z_n} H_n' Z_n\theta_0$. Therefore, by the Cauchy inequality,

$$\begin{aligned}
&[(Z_n\theta_0)' P_n M_{Z_n} P_n'(Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(P_n^s P_n^s)][(Z_n\theta_0)' H_n M_{Z_n} H_n'(Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(H_n^s H_n^s)] \\
&\geq [(Z_n\theta_0)' P_n M_{Z_n} H_n'(Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(P_n^s H_n^s)]^2,
\end{aligned}$$

i.e.,

$$\frac{(Z_n\theta_0)' P_n M_{Z_n} P_n'(Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(P_n^s P_n^s)}{[(Z_n\theta_0)' P_n M_{Z_n} G_n(Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(P_n^s G_n^s)]^2} \geq \frac{(Z_n\theta_0)' H_n M_{Z_n} H_n'(Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(H_n^s H_n^s)}{[(Z_n\theta_0)' H_n M_{Z_n} G_n(Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(H_n^s G_n^s)]^2},$$

using the facts $M_{Z_n} H_n'(Z_n\theta_0) = M_{Z_n} G_n(Z_n\theta_0)$, $tr(P_n^s G_n^s) = tr(P_n^s H_n^s)$, and $tr(H_n^s G_n^s) = tr(H_n^s H_n^s)$. Hence, using H_n as a quadratic matrix in the moment equation can generate a consistent root that has the smallest asymptotic variance. It follows from Proposition 1 that the consistent root is Eq. (20) when $P_n = H_n$, if $\lim_{n\to\infty} \frac{1}{n}[(Z_n\theta_0)' G_n' H_n M_{Z_n} G_n Z_n\theta_0 + \sigma_0^2 tr(G_n' H_n G_n)] \neq 0$, since $(Z_n\theta_0)' H_n M_{Z_n} G_n(Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(H_n^s G_n^s) = (Z_n\theta_0)' H_n M_{Z_n} H_n'(Z_n\theta_0) + \frac{1}{2}\sigma_0^2 tr(H_n^s H_n^s) \geq 0$. In addition, $\frac{1}{n} tr(H_n^s H_n^s) = \frac{1}{n} tr(G_n^s G_n^s) - 4\frac{tr^2(G_n)}{n^2} + o_P(1)$, by Lemma 1. Thus we have the expression for Ω_b. $\square$

Proof of Proposition 4. Let a_n, b_n and c_n be as given in Proposition 3. According to the proof of Proposition 1, $\hat{a}_n = (Z_n\theta_0 + \epsilon_n)' \hat{A}_n (Z_n\theta_0 + \epsilon_n)$, $\hat{b}_n = (Z_n\theta_0 + \epsilon_n)' \hat{B}_n (Z_n\theta_0 + \epsilon_n)$ and $\hat{c}_n = (Z_n\theta_0 + \epsilon_n)' \hat{C}_n (Z_n\theta_0 + \epsilon_n)$ with $\hat{A}_n = G_n' \hat{H}_n M_{Z_n} G_n$, $\hat{B}_n = G_n'(\hat{H}_n M_{Z_n})^s S_n^{-1}$ and $\hat{C}_n = S_n'^{-1} \hat{H}_n M_{Z_n} S_n^{-1}$. By the mean value theorem, $\hat{H}_n - H_n = \left(G_n'^2(\bar{\rho}_n) - \frac{tr[G_n'^2(\bar{\rho}_n) M_{Z_n}]}{n-d} M_{Z_n}\right)(\hat{\rho}_n - \rho_0)$, where $\bar{\rho}_n$ is between $\hat{\rho}_n$ and ρ_0. Writing $\hat{A}_n = G_n' H_n M_{Z_n} G_n + G_n'(\hat{H}_n - H_n) M_{Z_n} G_n$ and substituting the expression for $\hat{H}_n - H_n$ into $\hat{A}_n$, we have $\frac{1}{n}\hat{a}_n = \frac{1}{n} a_n + \frac{1}{n}(\hat{\rho}_n - \rho_0)(Z_n\theta_0 + \epsilon_n)' G_n' \cdot [G_n'^2(\bar{\rho}_n) - \frac{tr[G_n'^2(\bar{\rho}_n) M_{Z_n}]}{n-d} M_{Z_n}] M_{Z_n} G_n (Z_n\theta_0 + \epsilon_n)$. By Lemma 6, $G_n(\bar{\rho}_n)$ is bounded in both row and column sum norms for large enough n. As in the proof of Proposition 1, expanding the second term in the expression for $\frac{1}{n}\hat{a}_n$, we obtain $\frac{1}{n}\hat{a}_n = \frac{1}{n} a_n + o_P(1)$. Similarly, $\frac{1}{n}\hat{b}_n = \frac{1}{n} b_n + o_P(1)$ and $\frac{1}{n}\hat{c}_n = \frac{1}{n} c_n + o_P(1)$. Then, it follows by the continuous mapping theorem (see, e.g., Proposition 2.30 in White (1984)) that $\frac{\hat{b}_n - \sqrt{\hat{b}_n^2 - 4\hat{a}_n\hat{c}_n}}{2\hat{a}_n} = \frac{b_n - \sqrt{b_n^2 - 4a_n c_n}}{2a_n} + o_P(1)$ if $\lim_{n\to\infty} \frac{a_n}{n} \neq 0$.

For the asymptotic distribution, by the mean value theorem, $\sqrt{n}(\tilde{\rho}_{b,n} - \rho_0) = -(\frac{1}{n}\frac{\partial \hat{g}_n(\bar{\rho}_{b,n})}{\partial \rho})^{-1} \frac{1}{\sqrt{n}} \hat{g}_n(\rho_0)$, where $\bar{\rho}_{b,n}$ is between $\tilde{\rho}_{b,n}$ and ρ_0 and $\hat{g}_n(\rho) = \hat{a}_n \rho^2 - \hat{b}_n \rho + \hat{c}_n$. As $\frac{\partial \hat{g}_n(\rho)}{\partial \rho} = 2\hat{a}_n \rho - \hat{b}_n$, the above argument suggests that $\frac{1}{n}\frac{\partial \hat{g}_n(\bar{\rho}_{b,n})}{\partial \rho} = \frac{1}{n}\frac{\partial g_n(\bar{\rho}_{b,n})}{\partial \rho} + o_P(1)$, where $g_n(\rho) = a_n \rho^2 - b_n \rho + c_n$. In addition, $\frac{1}{\sqrt{n}}\hat{g}_n(\rho_0) = \frac{1}{\sqrt{n}}(Z_n\theta_0 + \epsilon_n)' \hat{H}_n M_{Z_n}(Z_n\theta_0 + \epsilon_n) = \frac{1}{\sqrt{n}} g_n(\rho_0) + \frac{1}{\sqrt{n}}(Z_n\theta_0 + \epsilon_n)'(\hat{H}_n - H_n) M_{Z_n}\epsilon_n = \frac{1}{\sqrt{n}} g_n(\rho_0) + \sqrt{n}(\hat{\rho}_n - \rho_0)\frac{1}{n}(Z_n\theta_0)'\left(G_n'^2(\bar{\rho}_n) M_{Z_n} - \frac{tr(G_n'^2(\bar{\rho}_n) M_{Z_n})}{n-d} M_{Z_n}\right)\epsilon_n + \sqrt{n}(\hat{\rho}_n - \rho_0)\frac{1}{n}\epsilon_n'\left(G_n'^2(\bar{\rho}_n) M_{Z_n} - \frac{tr[G_n'^2(\bar{\rho}_n) M_{Z_n}]}{n-d} M_{Z_n}\right)\epsilon_n = \frac{1}{\sqrt{n}} g_n(\rho_0) + o_P(1)$. It follows that $\tilde{\rho}_{b,n}$ has the same asymptotic distribution as the consistent root estimator $\hat{\rho}_{b,n}$ in Proposition 3, which is derived from solving $g_n(\rho) = 0$. $\square$

Proof of Proposition 5. To prove this proposition, we only need to slightly modify the proof of Proposition 1 to take into account the presence of unknown heteroskedasticity. A_n, B_n, C_n, a_{1n}, b_{1n} and c_{1n} have the same expressions as in Proposition 1. Because of heteroskedastic disturbances, a_{2n}, b_{2n} and c_{2n} now have different forms: $a_{2n} \equiv E(\epsilon_n' A_n \epsilon_n) = tr(\Sigma_n A_n) = O(n)$, $b_{2n} \equiv E(\epsilon_n' B_n \epsilon_n) = tr(\Sigma_n B_n) = O(n)$, and $c_{2n} \equiv E(\epsilon_n' C_n \epsilon_n) = tr(\Sigma_n C_n) = O(n)$. The expression for c_{2n} can be simplified by using $tr(\Sigma_n P_n M_{Z_n}) = 0$ as $P_n M_{Z_n}$ has a zero diagonal. As a result,

$$\frac{1}{n^2} b_n^2 - \frac{4}{n^2} a_n c_n = \big[\frac{1}{n}(Z_n\theta_0)' P_n M_{Z_n} G_n (Z_n\theta_0) + \frac{1}{n} tr(\Sigma_n P_n^s G_n)\big]^2 + o_P(1).$$

The rest of the proof is the same as the corresponding part of the proof of Proposition 1 except that different expressions for a_{2n}, b_{2n}, c_{2n} and $\frac{1}{n^2} b_n^2 - \frac{4}{n^2} a_n c_n$ are used. □

Proof of Proposition 6. We modify the proof of Proposition 2 to account for unknown heteroskedasticity. Since the error terms are heteroskedastic, $\sqrt{n}(\hat{\rho}_n - \rho_0) = -[\frac{1}{n}(2a_n\bar{\rho}_n - b_n)]^{-1} \frac{1}{\sqrt{n}}[(Z_n\theta_0)' P_n M_{Z_n} \epsilon_n + \epsilon_n' P_n M_{Z_n} \epsilon_n]$, where $\frac{1}{n}(2a_n\bar{\rho}_n - b_n) = -\frac{1}{n}[(Z_n\theta_0)' P_n M_{Z_n} G_n (Z_n\theta_0) + tr(\Sigma_n P_n^s G_n)] + o_P(1) = -\Sigma_{\rho,n} + o_P(1)$.

As $P_n M_{Z_n}$ has a zero diagonal, $E[\frac{1}{\sqrt{n}}(Z_n\theta_0)' P_n M_{Z_n} \epsilon_n + \frac{1}{\sqrt{n}} \epsilon_n' P_n M_{Z_n} \epsilon_n] = 0$, then by Lemma 5,

$$\big[(Z_n\theta_0)' P_n M_{Z_n} \epsilon_n / \sqrt{n} + \epsilon_n' P_n M_{Z_n} \epsilon_n / \sqrt{n}\big] / V_{\rho,n}^{\frac{1}{2}} \xrightarrow{D} N(0,1),$$

where, by Lemmas 1 and 2,

$$\begin{aligned} V_{\rho,n} &= var\big[\frac{1}{\sqrt{n}}(Z_n\theta_0)' P_n M_{Z_n} \epsilon_n + \frac{1}{\sqrt{n}} \epsilon_n' P_n M_{Z_n} \epsilon_n\big] \\ &= \frac{1}{n}(Z_n\theta_0)' P_n M_{Z_n} \Sigma_n M_{Z_n} P_n' (Z_n\theta_0) + \frac{1}{n} tr(\Sigma_n P_n \Sigma_n P_n^s) + o(1). \end{aligned}$$

Therefore, $\sqrt{n}(\hat{\rho}_n - \rho_0) \xrightarrow{D} N(0, \Omega)$, where

$$\Omega = \lim_{n\to\infty} (V_{\rho,n} \Sigma_{\rho,n}^{-2}) = V_\rho \Sigma_\rho^{-2}.$$ □

References

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Boston: Kluwer Academic Publishers.

Arraiz, I., Drukker, D.M., Kelejian, H.H., Prucha, I.R., 2010. A spatial Cliff-Ord-type model with heteroskedastic innovations: Small and large sample results. Regional Science and Urban Economics 50, 592–614.

Barry, R., Pace, R.K., 1999. Monte Carlo estimates of the log determinant of large sparse matrices. Linear Algebra and Its Applications 289, 41–54.

Griffith, D., 2000. Eigenfunction properties and approximations of selected incidence matrices employed in spatial analysis. Linear Algebra and Its Applications 321, 95–112.

Griffith, D., Sone, A., 1995. Trade-offs associated with normalizing constant computational simplifications for estimating spatial statistic models. Journal of Statistical Computation and Simulation 51, 165–183.

Helms, A.C., 2012. Keeping up with the Joneses: Neighborhood effects in housing renovation. Regional Science and Urban Economics 42, 303–313.

Horn, R., Johnson, C., 1985. Matrix Analysis. New York: Cambridge University Press.

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

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

Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics 104, 219–257.

Kelejian, H.H., Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157, 53–67.

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., 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. Department of Economics, The Ohio State University, Columbus, OH.

Lee, L.F., 2007a. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics 137, 489–514.

Lee, L.F., 2007b. The method of elimination and substitution in the GMM estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics 140, 155–189.

Lee, L.F., Liu, X., 2010. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. Econometric Theory 26, 187–230.

LeSage, J., Pace, R.K., 2009. Introduction to spatial econometrics. volume 196. Chapman & Hall/CRC.

Li, H., Calder, C.A., Cressie, N., 2007. Beyond Moran's I: Testing for spatial dependence based on the spatial autoregressive model. Geographical Analysis 39, 357–375.

Li, H., Calder, C.A., Cressie, N., 2012. One-step estimation of spatial dependence parameters: properties and extensions of the $APLE$ statistic. Journal of Multivariate Analysis 105, 68–84.

Lin, X., Lee, L.F., 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. Journal of Econometrics 157, 34–52.

Liu, X., Lee, L.F., Bollinger, C.R., 2010. An efficient GMM estimator of spatial autoregressive models. Journal of Econometrics 159, 303–319.

Martellosio, F., 2010. Power properties of invariant tests for spatial autocorrelation in linear regression. Econometric Theory 26, 152–186.

Martin, R.J., 1993. Approximations to the determinant term in Gaussian maximum likelihood estimation of some spatial models. Communications in Statistics—Theory and Methods 22, 189–205.

Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 35, 255–260.

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

Pace, R.K., 1997. Performing large spatial regressions and autoregressions. Economics Letters 54, 283–291.

Pace, R.K., Barry, R., 1997a. Quick computation of spatial autoregressive estimators. Geographical Analysis 29, 232–246.

Pace, R.K., Barry, R., 1997b. Sparse spatial autoregressions. Statistics and Probability Letters 33, 291–297.

Pace, R.K., LeSage, J., 2004. Techniques for improved approximation of the determinant term in the spatial likelihood function. Computational Statistics and Data Analysis 45, 179–196.

Pace, R.K., LeSage, J.P., 2009. A sampling approach to estimate the log determinant used in spatial likelihood problems.

Journal of Geographical Systems 11, 209–225.

Smirnov, O., Anselin, L., 2001. Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach. Computational Statistics and Data Analysis 35, 301–319.

Smirnov, O., Anselin, L., 2009. An $O(N)$ parallel method of computing the log-Jacobian of the variable transformation for models with spatial interaction on a lattice. Computational Statistics and Data Analysis 53, 2980–2988.

White, H., 1984. Asymptotic Theory for Econometricians. Orlando: Academic Press.