Online supplement to "Asymptotically efficient root estimators for spatial autoregressive models with spatial autoregressive disturbances"

Fei Jin*[1] and Lung-fei Lee[2]

[1]School of Economics, Fudan University, and Shanghai Institute of International Finance and Economics, Shanghai 200433, China

[2]Department of Economics, The Ohio State University, Columbus, OH 43210, USA

July 1, 2020

1 Monte Carlo settings and estimation results

The data generating process is the SARAR model (2.1). There are two variables in X_n: an intercept term and a variable randomly drawn from the standard normal distribution. The spatial weights matrix W_n is the circular world matrix as in Arraiz et al. (2010). For this matrix, the spatial units are equally spaced on a circle, one third of spatial units are connected to 4 nearest neighbors and the rest are connected to 10 nearest neighbors. If spatial unit i is connected to spatial unit j, then the (i,j)th element of the connectivity matrix is 1, and it is zero otherwise. The spatial weights matrix W_n is derived by normalizing the connectivity matrix to have row sums equal to one. We set $M_n = W_n$. The disturbances are normally distributed. In the homoskedastic case, $\sigma_0^2 = 1$; in the heteroskedastic case, each σ_{ni}^2 is proportional to the number of nonzero elements in the ith row of W_n, and the mean of σ_{ni}^2's is 1. The true value of $\beta = [\beta_1, \beta_2]'$ is $[1, 1]'$. The true values of λ and ρ are either 0.2 or 0.5. The sample size is either 200 or 400, and the number of Monte Carlo repetitions is $5,000$. For our root estimators, the initial consistent estimate of $[\lambda_0, \beta_0']'$ is a 2SLSE with the IV matrix $[X_n, W_nX_{1n}, W_n^2X_{1n}]$, where X_{1n} is the non-constant variable in X_n, and the initial estimator of ρ_0 is a consistent root estimator for which we use the quadratic matrices $M_n + \kappa M_n^2 + \kappa^2 M_n^3 - \mathrm{diag}(M_n + \kappa M_n^2 + \kappa M_n^3)$ with $\kappa = 0.2$ or 0.6.

We report the following robust measures of central tendency and dispersions: median bias (MB), median absolute deviation (MAD) and interdecile range (IDR), where the IDR is the difference between the 0.9 and 0.1 quantiles in the empirical distribution.[1] We also report coverage probabilities of 95% confidence intervals. The estimation results are reported in Table S.1. RE is the root estimator with the quadratic matrices $\tilde{G}_{nd}$ and $\tilde{T}_{nd}$ defined above (3.7), and RE5 is the root estimator with the quadratic

*Corresponding author. E-mail addresses: jin.fei@live.com (F. Jin), lee.1777@osu.edu (L.-F. Lee).

[1]There is a small proportion of outliers for the MME, so we report robust measures.

Table S.1: MBs, MADs, IDRs and CPs of various estimates

		λ	ρ	β_1	β_2
Homoskedastic case					
$n = 200$, $\lambda_0 = 0.2$, $\rho_0 = 0.2$	MLE	0.000(0.118)0.449[0.907]	-0.036(0.145)0.550[0.905]	-0.002(0.157)0.612[0.912]	-0.004(0.049)0.182[0.956]
	RE	0.008(0.108)0.419[0.918]	-0.038(0.131)0.521[0.930]	-0.010(0.147)0.582[0.925]	-0.003(0.048)0.183[0.955]
	RE5	0.008(0.109)0.419[0.918]	-0.038(0.131)0.521[0.929]	-0.010(0.147)0.583[0.924]	-0.003(0.048)0.182[0.955]
$n = 200$, $\lambda_0 = 0.5$, $\rho_0 = 0.5$	MLE	0.003(0.114)0.441[0.873]	-0.030(0.133)0.496[0.875]	-0.014(0.242)0.963[0.885]	-0.006(0.047)0.181[0.954]
	RE	0.023(0.105)0.445[0.887]	-0.042(0.125)0.506[0.903]	-0.056(0.232)0.961[0.898]	-0.005(0.047)0.184[0.952]
	RE5	0.024(0.106)0.450[0.872]	-0.045(0.129)0.516[0.889]	-0.056(0.233)0.973[0.884]	-0.005(0.047)0.185[0.945]
$n = 400$, $\lambda_0 = 0.2$, $\rho_0 = 0.2$	MLE	0.003(0.086)0.330[0.917]	-0.016(0.104)0.390[0.919]	-0.004(0.117)0.437[0.922]	-0.003(0.033)0.127[0.951]
	RE	0.007(0.081)0.317[0.927]	-0.019(0.099)0.374[0.936]	-0.008(0.111)0.423[0.932]	-0.003(0.033)0.127[0.950]
	RE5	0.007(0.081)0.317[0.927]	-0.019(0.099)0.374[0.935]	-0.008(0.111)0.423[0.932]	-0.003(0.033)0.127[0.950]
$n = 400$, $\lambda_0 = 0.5$, $\rho_0 = 0.5$	MLE	-0.004(0.087)0.334[0.903]	-0.008(0.095)0.362[0.898]	0.005(0.185)0.712[0.915]	-0.004(0.034)0.129[0.947]
	RE	0.011(0.078)0.317[0.914]	-0.017(0.090)0.354[0.922]	-0.021(0.171)0.693[0.922]	-0.003(0.034)0.130[0.944]
	RE5	0.011(0.078)0.318[0.911]	-0.017(0.092)0.355[0.916]	-0.022(0.172)0.695[0.921]	-0.003(0.034)0.130[0.941]
Heteroskedastic case					
$n = 200$, $\lambda_0 = 0.2$, $\rho_0 = 0.2$	MME	-0.005(0.107)0.421[0.917]	-0.035(0.142)0.530[0.915]	0.000(0.145)0.579[0.926]	-0.003(0.048)0.185[0.946]
	RE	0.000(0.102)0.406[0.930]	-0.038(0.132)0.504[0.934]	-0.005(0.141)0.556[0.933]	-0.002(0.049)0.185[0.946]
	RE5	0.000(0.102)0.405[0.929]	-0.038(0.132)0.504[0.933]	-0.005(0.141)0.556[0.933]	-0.002(0.049)0.185[0.946]
$n = 200$, $\lambda_0 = 0.5$, $\rho_0 = 0.5$	MME	-0.000(0.102)0.408[0.895]	-0.031(0.125)0.486[0.898]	-0.001(0.222)0.879[0.909]	-0.004(0.049)0.187[0.941]
	RE	0.016(0.097)0.427[0.900]	-0.040(0.120)0.499[0.915]	-0.040(0.215)0.926[0.913]	-0.003(0.050)0.193[0.937]
	RE5	0.016(0.099)0.429[0.887]	-0.044(0.122)0.507[0.900]	-0.041(0.216)0.926[0.903]	-0.003(0.050)0.193[0.932]
$n = 400$, $\lambda_0 = 0.2$, $\rho_0 = 0.2$	MME	-0.003(0.082)0.304[0.928]	-0.016(0.102)0.380[0.931]	0.004(0.110)0.421[0.930]	-0.002(0.033)0.127[0.948]
	RE	-0.001(0.078)0.293[0.937]	-0.018(0.097)0.369[0.943]	-0.000(0.108)0.407[0.938]	-0.002(0.033)0.127[0.947]
	RE5	-0.001(0.078)0.293[0.937]	-0.018(0.097)0.369[0.943]	-0.001(0.108)0.407[0.938]	-0.002(0.033)0.127[0.947]
$n = 400$, $\lambda_0 = 0.5$, $\rho_0 = 0.5$	MME	0.002(0.075)0.293[0.907]	-0.018(0.090)0.347[0.912]	-0.008(0.164)0.639[0.914]	-0.003(0.033)0.127[0.951]
	RE	0.012(0.070)0.281[0.919]	-0.025(0.086)0.337[0.925]	-0.027(0.156)0.615[0.923]	-0.002(0.033)0.127[0.950]
	RE5	0.012(0.071)0.283[0.917]	-0.025(0.087)0.339[0.922]	-0.028(0.157)0.615[0.920]	-0.002(0.033)0.128[0.948]

(i) RE is the root estimator with the quadratic matrices $\tilde{G}_{nd}$ and $\tilde{T}_{nd}$, and RE5 is the root estimator with the quadratic matrices $\tilde{G}_{nd,5}$ and $\tilde{T}_{nd,5}$.

(ii) The four numbers in each entry of the table are MB(MAD)IDR[CP], where MB is the median bias, MAD is the median absolute deviation, IDR is the interdecile range, i.e., the difference between the 0.9 and 0.1 quantiles in an empirical distribution, and CP is the coverage probability of a 95% confidence interval.

(iii) The true value of $\beta = [\beta_1, \beta_2]'$ is $[1, 1]'$. The mean of ϵ_{ni}'s variances is 1.

matrices $\tilde{G}_{nd,5}$ and $\tilde{T}_{nd,5}$ defined below (3.8). RE and RE5 have similar performance, and RE has slightly smaller MADs and IDRs in some cases. MLE and MME have smaller MBs than those of RE and RE5 in most cases, but they have larger MADs and IDRs in most cases. The CPs of all estimates for β_2 are close to 95%, but the CPs of estimates of other parameters are smaller than 95%. MLE and MME have lower CPs than those of RE and RE5 for parameters other than β_2. As the sample size increases from 200 to 400, the CPs are closer to the nominal 95%.

2 Regularity conditions for the QMLE and MME

Assumption S.1. *(a) The true ϕ_0 is in the interior of a compact parameter space of ϕ. (b) $\{S_n^{-1}(\lambda)\}$ is bounded in either row or column sum norm uniformly on the parameter space Λ of λ, and the same* *holds for $\{R_n^{-1}(\rho)\}$ on the parameter space $\boldsymbol{\varrho}$ of ρ. (c)* $\lim_{n\to\infty}\frac{1}{n}X_n'R_n'(\rho)R_n(\rho)X_n$ *exists and is nonsingular for any $\rho\in\boldsymbol{\varrho}$, and the sequence of smallest eigenvalues of $R_n'(\rho)R_n(\rho)$ is bounded away from zero uniformly on $\boldsymbol{\varrho}$. (d) For the QMLE, either (i)* $\lim_{n\to\infty}\frac{1}{n}[\ln|\sigma_0^2\Upsilon_n\Upsilon_n'| - \ln|\sigma^2(\gamma)\Upsilon_n(\gamma)\Upsilon_n'(\gamma)|]$ *exists and is nonzero for any $\gamma\neq\gamma_0$, where $\Upsilon_n = S_n^{-1}R_n^{-1}$, $\Upsilon_n(\gamma) = S_n^{-1}(\lambda)R_n^{-1}(\rho)$ and $\sigma^2(\gamma) = \frac{\sigma_0^2}{n}\operatorname{tr}[\Upsilon_n' S_n'(\lambda)R_n'(\rho)R_n(\rho)S_n(\lambda)\Upsilon_n]$, or (ii)* $\lim_{n\to\infty}\frac{1}{n}(W_nS_n^{-1}X_n\beta_0, X_n)'(W_nS_n^{-1}X_n\beta_0, X_n)$ *exists and is nonsingular, and* $\lim_{n\to\infty}\frac{1}{n}[\ln|\sigma_0^2\Upsilon_n\Upsilon_n'| - \ln|\sigma^2(\lambda_0,\rho)S_n^{-1}R_n^{-1}(\rho)R_n'^{-1}(\rho)S_n'^{-1}|]$ *exists and is nonzero for any $\rho\neq\rho_0$; for the MME, ϕ_0 is the unique root of* $\lim_{n\to\infty}\frac{1}{n}\operatorname{E}[h_n(\phi)] = 0$ *on the parameter space of ϕ.*

Assumptions S.1(a)–(c) and the identification condition for the QMLE in (d) are from Jin and Lee (2012). For the MME, as $h_n(\phi)$ is nonlinear in ϕ, a primitive identification condition is not obvious. So we maintain the relatively high level identification condition in Assumption S.1(d).

3 Proofs of theorems

This section provides proofs of the two theorems in the main paper. In the following, "MVT" will denote the mean value theorem, and "UB" will denote "uniformly bounded in both row and column sum norms".

Proof of Theorem 1. As $y_n = Z_n\eta_0 + u_n$,

$$\begin{aligned}\tilde{\eta} &= [Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'Z_n]^{-1}Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'(Z_n\eta_0+u_n)\\ &= \eta_0 + [Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'Z_n]^{-1}Z_n'Q_n(Q_n'Q_n)^{-1}Q_n'R_n^{-1}\epsilon_n.\end{aligned}$$

For any $k\times 1$ vector α, where k is the column dimension of Q_n,

$$\operatorname{var}\Big(\frac{1}{\sqrt{n}}\alpha'Q_n'R_n^{-1}\epsilon_n\Big) = \frac{1}{n}\alpha'Q_n'R_n^{-1}\Sigma_nR_n'^{-1}Q_n\alpha \le \frac{c}{n}\alpha'Q_n'R_n^{-1}R_n'^{-1}Q_n\alpha = O(1),$$

for some constant c, under Assumptions 1(a)–(b) and 2(a)–(b). Then $\frac{1}{\sqrt{n}}Q_n'R_n^{-1}\epsilon_n = O_p(1)$. As $\lim_{n\to\infty}\frac{1}{n}Q_n'Z_n$ has full column rank, $\tilde{\eta} = \eta_0 + O_p(n^{-1/2})$.

Let $\bar{\epsilon}_n = (I_n - \rho_0M_n)\tilde{u}_n$. Then $\tilde{u}_n = \rho_0M_n\tilde{u}_n + \bar{\epsilon}_n$ is an estimated SAR process. Using this equation, $b_n = -(\rho_0M_n\tilde{u}_n+\bar{\epsilon}_n)'P_n^{(s)}M_n\tilde{u}_n = -2\rho_0a_n - d_n$ and $c_n = (\rho_0M_n\tilde{u}_n+\bar{\epsilon}_n)'P_n(\rho_0M_n\tilde{u}_n+\bar{\epsilon}_n) = \rho_0^2a_n+\rho_0d_n+e_n$, where $d_n = \bar{\epsilon}_n'P_n^{(s)}M_n\tilde{u}_n$ and $e_n = \bar{\epsilon}_n'P_n\bar{\epsilon}_n$. Note that $\tilde{u}_n = [S_n+(\lambda_0-\tilde{\lambda})W_n]y_n - X_n\beta_0 + X_n(\beta_0-\tilde{\beta}) = u_n + (\lambda_0-\tilde{\lambda})W_ny_n + X_n(\beta_0-\tilde{\beta})$. Then, with $P_n = P_n^*$,

$$\begin{aligned}a_n^* &= [u_n+(\lambda_0-\tilde{\lambda})W_ny_n+X_n(\beta_0-\tilde{\beta})]'M_n'P_n^*M_n[u_n+(\lambda_0-\tilde{\lambda})W_ny_n+X_n(\beta_0-\tilde{\beta})]\\ &= u_n'M_n'P_n^*M_nu_n + (\lambda_0-\tilde{\lambda})^2(W_ny_n)'M_n'P_n^*M_nW_ny_n + (\beta_0-\tilde{\beta})'X_n'M_n'P_n^*M_nX_n(\beta_0-\tilde{\beta})\\ &\quad + (\lambda_0-\tilde{\lambda})u_n'M_n'P_n^{*(s)}M_nW_ny_n + u_n'M_n'P_n^{*(s)}M_nX_n(\beta_0-\tilde{\beta})\\ &\quad + (\lambda_0-\tilde{\lambda})(W_ny_n)'M_n'P_n^{*(s)}M_nX_n(\beta_0-\tilde{\beta}).\end{aligned}$$

Using $u_n = R_n^{-1}\epsilon_n$ and the reduced form $y_n = S_n^{-1}X_n\beta_0 + S_n^{-1}R_n^{-1}\epsilon_n$, we have $\frac{1}{n}(W_ny_n)'M_n'P_n^*M_nW_ny_n = O_p(1)$, $\frac{1}{n}u_n'M_n'P_n^{*(s)}M_nW_ny_n = O_p(1)$, $\frac{1}{n}u_n'M_n'P_n^{*(s)}M_nX_n = o_p(1)$ and $\frac{1}{n}(W_ny_n)'M_n'P_n^{*(s)}M_nX_n = O_p(1)$. Thus, $\frac{1}{n}a_n^* = \frac{1}{n}u_n'M_n'P_n^*M_nu_n + o_p(1) = \frac{1}{n}\operatorname{tr}(R_n'^{-1}M_n'P_n^*M_nR_n^{-1}\Sigma_n) + o_p(1)$, where the second equality follows by Lemma A.3 in Lin and Lee (2010). Similarly, $\frac{1}{n}d_n^* = \frac{1}{n}\operatorname{tr}(P_n^{*(s)}M_nR_n^{-1}\Sigma_n) + o_p(1)$, and $\frac{1}{n}e_n^* =$

$\frac{1}{n}\operatorname{tr}(P_n^*\Sigma_n)+o_p(1)=o_p(1)$ because P_n^* has a zero diagonal. Hence, $\frac{1}{n^2}b_n^{*2}-\frac{4}{n^2}a_n^*c_n^*=(\frac{d_n^*}{n})^2+o_p(1)$. Since $\lim_{n\to\infty}\frac{1}{n}\operatorname{tr}(R_n'^{-1}M_n'P_n^*M_nR_n^{-1}\Sigma_n)\neq 0$, (2.2) is quadratic in the limit. Then the consistent root is $(-b_n^*-\sqrt{b_n^{*2}-4a_n^*c_n^*})/(2a_n^*)$ if $\operatorname{plim}_{n\to\infty}\frac{1}{n}d_n^*\geq 0$, and is $(-b_n^*+\sqrt{b_n^{*2}-4a_n^*c_n^*})/(2a_n^*)$ if $\operatorname{plim}_{n\to\infty}\frac{1}{n}d_n^*<0$. Denote the consistent root by $\tilde{\rho}^*$. By the MVT,

$$0=\tilde{u}_n'R_n'(\tilde{\rho}^*)P_n^*R_n(\tilde{\rho}^*)\tilde{u}_n=\tilde{u}_n'R_n'P_n^*R_n\tilde{u}_n-\tilde{u}_n'R_n'(\bar{\rho})P_n^{*(s)}M_n\tilde{u}_n(\tilde{\rho}^*-\rho_0),$$

where $\bar{\rho}$ lies between $\tilde{\rho}^*$ and ρ_0. Thus, $\sqrt{n}(\tilde{\rho}^*-\rho_0)=(\frac{1}{n}\tilde{u}_n'R_n'(\bar{\rho})P_n^{*(s)}M_n\tilde{u}_n)^{-1}\frac{1}{\sqrt{n}}\tilde{u}_n'R_n'P_n^*R_n\tilde{u}_n$, where $\frac{1}{n}\tilde{u}_n'R_n'(\bar{\rho})P_n^{*(s)}M_n\tilde{u}_n=\frac{1}{n}u_n'R_n'(\bar{\rho})P_n^{*(s)}M_nu_n+o_p(1)=\frac{1}{n}\operatorname{tr}(P_n^{*(s)}M_nR_n^{-1}\Sigma_n)+o_p(1)$ and

$$\begin{aligned}\frac{1}{\sqrt{n}}\tilde{u}_n'R_n'P_n^*R_n\tilde{u}_n&=\frac{1}{\sqrt{n}}u_n'R_n'P_n^*R_nu_n+\frac{1}{n}u_n'R_n'P_n^{*(s)}R_n[\sqrt{n}(\lambda_0-\tilde{\lambda})W_ny_n+X_n\sqrt{n}(\beta_0-\tilde{\beta})]\\&\quad+\frac{1}{n}[(\lambda_0-\tilde{\lambda})W_ny_n+X_n(\beta_0-\tilde{\beta})]'R_n'P_n^*R_n[\sqrt{n}(\lambda_0-\tilde{\lambda})W_ny_n+X_n\sqrt{n}(\beta_0-\tilde{\beta})]\\&=\frac{1}{\sqrt{n}}\epsilon_n'P_n^*\epsilon_n+O_p(1)+o_p(1)=O_p(1).\end{aligned}$$

As $\lim_{n\to\infty}\frac{1}{n}\operatorname{tr}(P_n^{*(s)}M_nR_n^{-1}\Sigma_n)\neq 0$ under Assumption 2, $\sqrt{n}(\tilde{\rho}^*-\rho_0)=O_p(1)$. We have a similar result for the consistent root $\tilde{\rho}^{**}$ of the moment equation with P_n^{**}. Thus the method described in the main text will locate a $\sqrt{n}$-consistent root estimator $\tilde{\rho}$ of ρ_0, since $\tilde{\rho}=\rho_0+O_p(n^{-1/2})$. □

Proof of Theorem 2. We first prove the consistency of $\hat{\rho}$. For that purpose, we now investigate the order of a_{1n} in (3.9). As $\tilde{u}_n=u_n+(\lambda_0-\tilde{\lambda})W_ny_n+X_n(\beta_0-\tilde{\beta})$,

$$\begin{aligned}\frac{1}{n}\tilde{u}_n'M_n'\tilde{T}_{nd,k}M_n\tilde{u}_n&=\frac{1}{n}u_n'M_n'\tilde{T}_{nd,k}M_nu_n+\frac{1}{n}(\lambda_0-\tilde{\lambda})^2y_n'W_n'M_n'\tilde{T}_{nd,k}M_nW_ny_n\\&\quad+\frac{1}{n}(\beta_0-\tilde{\beta})'X_n'M_n'\tilde{T}_{nd,k}M_nX_n(\beta_0-\tilde{\beta})+\frac{1}{n}(\lambda_0-\tilde{\lambda})u_n'M_n'\tilde{T}_{nd,k}^{(s)}M_nW_ny_n\\&\quad+\frac{1}{n}u_n'M_n'\tilde{T}_{nd,k}^{(s)}M_nX_n(\beta_0-\tilde{\beta})+\frac{1}{n}(\lambda_0-\tilde{\lambda})y_n'W_n'M_n'\tilde{T}_{nd,k}^{(s)}M_nX_n(\beta_0-\tilde{\beta}).\end{aligned}$$

Since $\tilde{\rho}=\rho_0+o_p(1)$, in a neighborhood of ρ_0, $\sum_{i=0}^k\tilde{\rho}^iM_n^i$ is UB for any natural number k, and $\sum_{i=0}^\infty\tilde{\rho}^iM_n^i=\tilde{R}_n^{-1}$ is UB by Lemma A.3 in Lee (2004). It follows that $\tilde{T}_{nd,k}$ is UB in a neighborhood of ρ_0, by the sub-multiplicability of the row and column sum matrix norms. Then $\frac{1}{n}X_n'M_n'\tilde{T}_{nd,k}M_nX_n=O_p(1)$ and $\frac{1}{n}u_n'M_n'\tilde{T}_{nd,k}^{(s)}M_nX_n=o_p(1)$. Using $u_n=R_n^{-1}\epsilon_n$ and $y_n=S_n^{-1}X_n\beta_0+S_n^{-1}R_n^{-1}\epsilon_n$, we have $\frac{1}{n}y_n'W_n'M_n'\tilde{T}_{nd,k}M_nW_ny_n=O_p(1)$, $\frac{1}{n}u_n'M_n'\tilde{T}_{nd,k}^{(s)}M_nW_ny_n=O_p(1)$, and $\frac{1}{n}y_n'W_n'M_n'\tilde{T}_{nd,k}^{(s)}M_nX_n=O_p(1)$. Thus,

$$\begin{aligned}\frac{1}{n}\tilde{u}_n'M_n'\tilde{T}_{nd,k}M_n\tilde{u}_n&=\frac{1}{n}u_n'M_n'\tilde{T}_{nd,k}M_nu_n+o_p(1)\\&=\frac{1}{n}u_n'M_n'T_{nd,k}M_nu_n+o_p(1)\\&=\frac{1}{n}\operatorname{tr}(R_n'^{-1}M_n'T_{nd,k}M_nR_n^{-1}\Sigma_n)+o_p(1),\end{aligned}$$

where the second equality follows by the MVT and the third equality follows by Lemma A.3 in Lin and Lee (2010). Similarly, $\frac{1}{n}\tilde{u}_n'M_n'\tilde{G}_{nd,k}M_n\tilde{u}_n=\frac{1}{n}\operatorname{tr}(R_n'^{-1}M_n'G_{nd,k}M_nR_n^{-1}\Sigma_n)+o_p(1)$. Since $\frac{\partial\epsilon_n(\phi)}{\partial\eta'}=-R_n(\rho)[W_ny_n,X_n]$, as in the above proof,

$$\frac{1}{n}\frac{\partial g_{1n}(\tilde{\rho},\tilde{\eta})}{\partial\eta'}=-\frac{1}{n}\tilde{\epsilon}_n'\tilde{T}_{nd,k}^{(s)}\tilde{R}_n[W_ny_n,X_n]$$

$$= -[\frac{1}{n}\operatorname{tr}(T_{nd,k}^{(s)} R_n W_n S_n^{-1} R_n^{-1} \Sigma_n), 0] + o_p(1)$$
$$= -[\frac{1}{n}\operatorname{tr}(T_{nd,k}^{(s)} G_{nd} \Sigma_n), 0] + o_p(1),$$

by recalling that $\tilde{\epsilon}_n = \tilde{R}_n \tilde{u}_n$ and $G_{nd} = R_n W_n S_n^{-1} R_n^{-1} - \frac{1}{n}\operatorname{tr}(W_n S_n^{-1}) I_n$ for the homogeneous variance case, but $G_{nd} = R_n W_n S_n^{-1} R_n^{-1} - \operatorname{diag}(R_n W_n S_n^{-1} R_n^{-1})$ for the heterogeneous variances case, where the last equality holds because $T_{nd,k}^{(s)}$ has a zero trace and $\Sigma_n = \sigma_0^2 I_n$ in the homoskedastic case, and $T_{nd,k}^{(s)}$ has a zero diagonal and Σ_n is a diagonal matrix in the heteroskedastic case; and

$$\frac{1}{n}\frac{\partial g_{2n}(\tilde{\rho}, \tilde{\eta})}{\partial \eta'} = \begin{pmatrix} -\frac{1}{n}\tilde{\epsilon}_n' \tilde{G}_{nd,k}^{(s)} \tilde{R}_n [W_n y_n, X_n] - \frac{1}{n}(\tilde{R}_n W_n \tilde{S}_n^{-1} X_n \tilde{\beta})' \tilde{R}_n [W_n y_n, X_n] \\ -\frac{1}{n} X_n' \tilde{R}_n' \tilde{R}_n [W_n y_n, X_n] \end{pmatrix}$$
$$= -\begin{pmatrix} F_n & \frac{1}{n}(R_n W_n S_n^{-1} X_n \beta_0)' R_n X_n \\ \frac{1}{n} X_n' R_n' R_n W_n S_n^{-1} X_n \beta_0 & \frac{1}{n} X_n' R_n' R_n X_n \end{pmatrix} + o_p(1),$$

where $F_n = \frac{1}{n}\operatorname{tr}(G_{nd,k}^{(s)} G_{nd} \Sigma_n) + \frac{1}{n}(R_n W_n S_n^{-1} X_n \beta_0)'(R_n W_n S_n^{-1} X_n \beta_0)$. For a block matrix $E = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$, where A and D are square matrices and D is invertible, since

$$\begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I & 0 \\ -D^{-1}C & I \end{pmatrix} = \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix}, \tag{S.1}$$

E has full rank if $A - BD^{-1}C$ has full rank. Thus, $\operatorname{plim}_{n\to\infty} \frac{1}{n}\frac{\partial g_{2n}(\tilde{\rho}, \tilde{\eta})}{\partial \eta'}$ has full rank under the conditions that $\lim_{n\to\infty} \frac{1}{n} X_n' R_n' R_n X_n$ is nonsingular and $\lim_{n\to\infty} \frac{1}{n}\Xi_n \neq 0$ in Assumption 3(b). It follows that $\tilde{C}_{n\rho}$ exists for large enough n and $\tilde{C}_{n\rho} = O_p(1)$. Hence, $\frac{1}{n} a_{1n} = \frac{1}{n} a_{1n}^* + o_p(1)$, where a_{1n}^* is given in (3.13) and is derived by the block matrix inverse formula. It follows that $\operatorname{plim}_{n\to\infty} \frac{1}{n} a_{1n} \neq 0$ if $\lim_{n\to\infty} \frac{1}{n} a_{1n}^* \neq 0$. With $\operatorname{plim}_{n\to\infty} \frac{1}{n} a_{1n} \neq 0$, (3.9) is a quadratic equation of ρ for a large enough n.

Using $\tilde{u}_n = \rho_0 M_n \tilde{u}_n + \bar{\epsilon}_n$, which defines $\bar{\epsilon}_n = (I_n - \rho_0 M_n)\tilde{u}_n$, we have

$$b_{1n} = -\tilde{u}_n' M_n' \tilde{T}_{nd,k}^{(s)} (\rho_0 M_n \tilde{u}_n + \bar{\epsilon}_n) + \tilde{C}_{n\rho} \begin{pmatrix} \tilde{u}_n' M_n' \tilde{G}_{nd,k}^{(s)} (\rho_0 M_n \tilde{u}_n + \bar{\epsilon}_n) + \tilde{u}_n' M_n' \tilde{R}_n W_n \tilde{S}_n^{-1} X_n \tilde{\beta} \\ X_n' \tilde{R}_n' M_n \tilde{u}_n \end{pmatrix}$$
$$= -2\rho_0 a_{1n} - d_{1n}^* + e_{1n},$$

and

$$c_{1n} = (\rho_0 M_n \tilde{u}_n + \bar{\epsilon}_n)' \tilde{T}_{nd,k} (\rho_0 M_n \tilde{u}_n + \bar{\epsilon}_n) - \tilde{C}_{n\rho} \begin{pmatrix} (\rho_0 M_n \tilde{u}_n + \bar{\epsilon}_n)' \tilde{G}_{nd,k} (\rho_0 M_n \tilde{u}_n + \bar{\epsilon}_n) + \tilde{u}_n' \tilde{R}_n W_n \tilde{S}_n^{-1} X_n \tilde{\beta} \\ X_n' \tilde{R}_n' \tilde{u}_n \end{pmatrix}$$
$$= \rho_0^2 a_{1n} + \rho_0 d_{1n}^* + f_{1n},$$

where $d_{1n}^* = \tilde{u}_n' M_n' \tilde{T}_{nd,k}^{(s)} \bar{\epsilon}_n - \tilde{C}_{n\rho} \left(\begin{smallmatrix} \tilde{u}_n' M_n' \tilde{G}_{nd,k}^{(s)} \bar{\epsilon}_n \\ 0 \end{smallmatrix}\right)$, $e_{1n} = \tilde{C}_{n\rho} \left(\begin{smallmatrix} \tilde{u}_n' M_n' \tilde{R}_n W_n \tilde{S}_n^{-1} X_n \tilde{\beta} \\ X_n' \tilde{R}_n' M_n \tilde{u}_n \end{smallmatrix}\right)$, and $f_{1n} = \bar{\epsilon}_n' \tilde{T}_{nd,k} \bar{\epsilon}_n - \tilde{C}_{n\rho} \left(\begin{smallmatrix} \bar{\epsilon}_n' \tilde{G}_{nd,k} \bar{\epsilon}_n + \tilde{u}_n' \tilde{R}_n W_n \tilde{S}_n^{-1} X_n \tilde{\beta} \\ X_n' \tilde{R}_n' \tilde{u}_n \end{smallmatrix}\right)$. We have $\frac{1}{n}\bar{\epsilon}_n' \tilde{T}_{nd,k} \bar{\epsilon}_n = \frac{1}{n}\tilde{u}_n' R_n' \tilde{T}_{nd,k} R_n \tilde{u}_n = \frac{1}{n} u_n' R_n' \tilde{T}_{nd,k} R_n u_n + o_p(1) = \frac{1}{n}\epsilon_n' T_{nd,k} \epsilon_n + o_p(1) = o_p(1)$, where the last equality holds since $\frac{1}{n}\operatorname{E}(\epsilon_n' T_{nd,k} \epsilon_n) = 0$. Similarly, $\frac{1}{n}\bar{\epsilon}_n' \tilde{G}_{nd,k} \bar{\epsilon}_n = o_p(1)$. Then $\frac{1}{n} d_{1n}^* = O_p(1)$, $\frac{1}{n} e_{1n} = o_p(1)$ and $\frac{1}{n} f_{1n} = o_p(1)$, by arguments as that for the order of a_{1n}. Thus, $\frac{1}{n} b_{1n} = -\frac{2\rho_0}{n} a_{1n} - \frac{1}{n} d_{1n}^* + o_p(1)$, and $\frac{1}{n} c_{1n} = \frac{\rho_0^2}{n} a_{1n} + \frac{\rho_0}{n} d_{1n}^* + o_p(1)$. It follows that the roots of (3.9) are $\frac{-b_{1n} \pm \sqrt{b_{1n}^2 - 4a_{1n}c_{1n}}}{2a_{1n}} = \rho_0 + \frac{d_{1n}^*/n \pm \sqrt{(d_{1n}^*/n)^2}}{2a_{1n}/n} + o_p(1)$. Hence, the consistent root estimator $\hat{\rho}$ of ρ_0 is $\hat{\rho} = \frac{-b_{1n} - \sqrt{b_{1n}^2 - 4a_{1n}c_{1n}}}{2a_{1n}}$ if $\operatorname{plim}_{n\to\infty} \frac{1}{n} d_{1n}^* \geq 0$, and it is $\hat{\rho} = \frac{-b_{1n} + \sqrt{b_{1n}^2 - 4a_{1n}c_{1n}}}{2a_{1n}}$ if $\operatorname{plim}_{n\to\infty} \frac{1}{n} d_{1n}^* < 0$.

The d_{1n} in (3.10) differs from d_{1n}^* only in that $\tilde{R}_n$ replaces R_n in d_{1n}^*, thus d_{1n} and d_{1n}^* have the same probability limit.

We next prove the consistency of $\hat{\lambda}$. Let $\bar{u}_n = y_n - \lambda_0 W_n y_n - X_n\tilde{\beta}$. Then, $y_n - X_n\tilde{\beta} = \lambda_0 W_n y_n + \bar{u}_n$. Using this equation, we have

$$\begin{aligned} b_{2n} &= -(W_n y_n)'\tilde{R}_n'\tilde{G}_{nd,k}^{(s)}\tilde{R}_n(\lambda_0 W_n y_n + \bar{u}_n) - (W_n y_n)'\tilde{R}_n'\tilde{R}_n W_n \tilde{S}_n^{-1}X_n\tilde{\beta} \\ &\quad + \tilde{C}_{n\lambda}\begin{pmatrix} (W_n y_n)'\tilde{R}_n'\tilde{T}_{nd,k}^{(s)}\tilde{R}_n(\lambda_0 W_n y_n + \bar{u}_n) \\ X_n'\tilde{R}_n'\tilde{R}_n W_n y_n \end{pmatrix} \\ &= -2\lambda_0 a_{2n} - d_{2n}^*, \end{aligned}$$

and

$$\begin{aligned} c_{2n} &= (\lambda_0 W_n y_n + \bar{u}_n)'\tilde{R}_n'\tilde{G}_{nd,k}\tilde{R}_n(\lambda_0 W_n y_n + \bar{u}_n) + (\lambda_0 W_n y_n + \bar{u}_n)'\tilde{R}_n'\tilde{R}_n W_n \tilde{S}_n^{-1}X_n\tilde{\beta} \\ &\quad - \tilde{C}_{n\lambda}\begin{pmatrix} (\lambda_0 W_n y_n + \bar{u}_n)'\tilde{R}_n'\tilde{T}_{nd,k}\tilde{R}_n(\lambda_0 W_n y_n + \bar{u}_n) \\ X_n'\tilde{R}_n'\tilde{R}_n(\lambda_0 W_n y_n + \bar{u}_n) \end{pmatrix} \\ &= \lambda_0^2 a_{2n} + \lambda_0 d_{2n}^* + e_{2n}, \end{aligned}$$

where $d_{2n}^* = (W_n y_n)'\tilde{R}_n'\tilde{G}_{nd,k}^{(s)}\tilde{R}_n\bar{u}_n + (W_n y_n)'\tilde{R}_n'\tilde{R}_n W_n\tilde{S}_n^{-1}X_n\tilde{\beta} - \tilde{C}_{n\lambda}\begin{pmatrix} (W_n y_n)'\tilde{R}_n'\tilde{T}_{nd,k}^{(s)}\tilde{R}_n\bar{u}_n \\ X_n'\tilde{R}_n'\tilde{R}_n W_n y_n \end{pmatrix}$, and $e_{2n} = \bar{u}_n'\tilde{R}_n'\tilde{G}_{nd,k}\tilde{R}_n\bar{u}_n + \bar{u}_n'\tilde{R}_n'\tilde{R}_n W_n\tilde{S}_n^{-1}X_n\tilde{\beta} - \tilde{C}_{n\lambda}\begin{pmatrix} \bar{u}_n'\tilde{R}_n'\tilde{T}_{nd,k}\tilde{R}_n\bar{u}_n \\ X_n'\tilde{R}_n'\tilde{R}_n\bar{u}_n \end{pmatrix}$. We may prove that $\frac{1}{n}a_{2n} = \frac{1}{n}a_{2n}^* + o_p(1) = O_p(1)$, where $\frac{1}{n}a_{2n}^*$ is in (3.14), $\frac{1}{n}d_{2n}^* = O_p(1)$ and $\frac{1}{n}e_{2n} = o_p(1)$. Then the consistency of $\hat{\lambda}$ can be proved similarly as that of $\hat{\rho}$. By (3.5), $\hat{\beta}$ can be explicitly expressed as a function of $\hat{\rho}$ and $\hat{\lambda}$, so its consistency follows by the MVT.

By the MVT, $0 = g_{1n}(\hat{\rho}, \tilde{\eta}) - \tilde{C}_{n\rho}g_{2n}(\hat{\rho}, \tilde{\eta}) = g_{1n}(\rho_0, \eta_0) - \tilde{C}_{n\rho}g_{2n}(\rho_0, \eta_0) + (\frac{\partial g_{1n}(\bar{\rho}, \bar{\eta})}{\partial \rho} - \tilde{C}_{n\rho}\frac{\partial g_{2n}(\bar{\rho}, \bar{\eta})}{\partial \rho})(\hat{\rho} - \rho_0) + (\frac{\partial g_{1n}(\bar{\rho}, \bar{\eta})}{\partial \eta'} - \tilde{C}_{n\rho}\frac{\partial g_{2n}(\bar{\rho}, \bar{\eta})}{\partial \eta'})(\tilde{\eta} - \eta_0)$, where $(\bar{\rho}, \bar{\eta}')'$ lies between $(\hat{\rho}, \tilde{\eta}')'$ and ϕ_0. Then, $\sqrt{n}(\hat{\rho} - \rho_0) = -(\frac{1}{n}\frac{\partial g_{1n}(\bar{\rho}, \bar{\eta})}{\partial \rho} - \frac{1}{n}\tilde{C}_{n\rho}\frac{\partial g_{2n}(\bar{\rho}, \bar{\eta})}{\partial \rho})^{-1}[\frac{1}{\sqrt{n}}g_{1n}(\rho_0, \eta_0) - \frac{1}{\sqrt{n}}\tilde{C}_{n\rho}g_{2n}(\rho_0, \eta_0) + (\frac{1}{n}\frac{\partial g_{1n}(\bar{\rho}, \bar{\eta})}{\partial \eta'} - \frac{1}{n}\tilde{C}_{n\rho}\frac{\partial g_{2n}(\bar{\rho}, \bar{\eta})}{\partial \eta'})\sqrt{n}(\tilde{\eta} - \eta_0)]$. As $\tilde{C}_{n\rho} = \frac{\partial g_{1n}(\tilde{\rho}, \tilde{\eta})}{\partial \eta'}(\frac{\partial g_{2n}(\tilde{\rho}, \tilde{\eta})}{\partial \eta'})^{-1}$, $\frac{1}{n}\frac{\partial g_{1n}(\bar{\rho}, \bar{\eta})}{\partial \eta'} - \frac{1}{n}\tilde{C}_{n\rho}\frac{\partial g_{2n}(\bar{\rho}, \bar{\eta})}{\partial \eta'} = o_p(1)$ by the MVT. Hence, $\sqrt{n}(\hat{\rho} - \rho_0) = -(\frac{1}{n}\frac{\partial g_{1n}(\bar{\rho}, \bar{\eta})}{\partial \rho} - \frac{1}{n}\tilde{C}_{n\rho}\frac{\partial g_{2n}(\bar{\rho}, \bar{\eta})}{\partial \rho})^{-1}[\frac{1}{\sqrt{n}}g_{1n}(\rho_0, \eta_0) - \frac{1}{\sqrt{n}}\tilde{C}_{n\rho}g_{2n}(\rho_0, \eta_0)] + o_p(1)$. Let $g_n^*(\phi)$ be the vector derived by replacing $\tilde{G}_{nd,k}$, $\tilde{T}_{nd,k}$, $\tilde{R}_n$, $\tilde{S}_n$ and $\tilde{\beta}$ in $g_n(\phi)$, which is in (3.8), with, respectively, $G_{nd,k}$, $T_{nd,k}$, R_n, S_n and β_0. By the the block matrix inverse formula, the second element of $-(\frac{1}{n}\frac{\partial g_n^*(\phi_0)}{\partial \phi'})^{-1}\frac{1}{\sqrt{n}}g_n^*(\phi_0)$ has the same leading order term as that of $\sqrt{n}(\hat{\rho} - \rho_0)$. Similarly, the first element of $-(\frac{1}{n}\frac{\partial g_n^*(\phi_0)}{\partial \phi'})^{-1}\frac{1}{\sqrt{n}}g_n^*(\phi_0)$ has the same leading order term as that of $\sqrt{n}(\hat{\lambda} - \lambda_0)$, and the last k_x elements of $-(\frac{1}{n}\frac{\partial g_n^*(\phi_0)}{\partial \phi'})^{-1}\frac{1}{\sqrt{n}}g_n^*(\phi_0)$ have the same leading order term as that of $\sqrt{n}(\hat{\beta} - \beta_0)$. Note that $\frac{1}{n}\,\mathrm{E}\,\frac{\partial g_n^*(\phi_0)}{\partial \phi'} = -\Gamma_{nd,k}$, and $\mathrm{var}[\frac{1}{\sqrt{n}}g_n^*(\phi_0)] = \Omega_{nd,k} + \Delta_{nd,k}$ can be derived by using, e.g., Lemma 2 in Jin and Lee (2012). Using (S.1) twice, we see that $\lim_{n\to\infty}\Gamma_{nd,k}$ has full rank under Assumption 3(d). Then, $\sqrt{n}(\hat{\phi} - \phi_0) \xrightarrow{d} N(0, \lim_{n\to\infty}\Gamma_{nd,k}^{-1}(\Omega_{nd,k} + \Delta_{nd,k})\Gamma_{nd,k}'^{-1})$, by Theorem 1 in Kelejian and Prucha (2001).

When ϵ_{ni}'s are homoskedastic, the consistency of the QMLE ϕ^* can be seen from Jin and Lee (2013). The ϕ^* is characterized by $m_n(\phi^*) = 0$, where $m_n(\phi)$ is defined above (3.7). When $k = \infty$, since $m_n(\phi_0) = g_n^*(\phi_0)$, an MVT expansion show that $\sqrt{n}(\phi^* - \phi_0)$ is asymptotically equivalent to $\sqrt{n}(\phi_{g^*} - \phi_0)$.

When ϵ_{ni}'s are heteroskedastic, denote the MME of ϕ_0 from solving $h_n(\phi) = 0$ by ϕ_h. As $\epsilon_n(\phi)$ is linear in λ, ρ and β, using the reduced form of y_n, it is straightforward to prove that $\frac{1}{n}h_n(\phi) - \frac{1}{n}\,\mathrm{E}[h_n(\phi)] = o_p(1)$. Under Assumption S.1($b$), we may prove by the MVT that $\frac{1}{n}h_n(\phi) - \frac{1}{n}\,\mathrm{E}[h_n(\phi)]$ is stochastically equicontinuous. Then $\frac{1}{n}h_n(\phi) - \frac{1}{n}\,\mathrm{E}[h_n(\phi)] = o_p(1)$ uniformly on the parameter space of ϕ. With a compact parameter space of ϕ and the identification condition in Assumption S.1(d), the consistency of ϕ_h follows. When $k = \infty$, an MVT expansion shows that $\frac{1}{\sqrt{n}}h_n(\phi_0) - \frac{1}{\sqrt{n}}g_n^*(\phi_0) = o_p(1)$, thus $\sqrt{n}(\phi_h - \phi_0)$ is asymptotically equivalent to $\sqrt{n}(\hat{\phi} - \phi_0)$. □

References

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.

Jin, F., Lee, L.F., 2012. Approximated likelihood and root estimators for spatial interaction in spatial autoregressive models. Regional Science and Urban Economics 42, 446–458.

Jin, F., Lee, L.F., 2013. Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances. Regional Science and Urban Economics 43, 590–616.

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.

Lee, L.F., 2004. A supplement to "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models". Department of Economics, The Ohio State University, Columbus, OH.

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