GMM estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors

Fei Jin[1] and Yuqin Wang[*2]

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

[2]Institute for Advanced Research, Shanghai University of Finance and Economics, Shanghai 200433, China

October 21, 2021

Abstract

This paper considers the generalized method of moments (GMM) estimation of a spatial autoregressive (SAR) model with SAR disturbances, where we allow for endogenous regressors in addition to a spatial lag of the dependent variable. We do not assume any reduced form of the endogenous regressors, thus we allow for spatial dependence and heterogeneity in endogenous regressors, and allow for nonlinear relations between endogenous regressors and their instruments. Innovations in the model can be homoskedastic or heteroskedastic with unknown forms. We prove that GMM estimators with linear and quadratic moments are consistent and asymptotically normal. In the homoskedastic case, we derive the best linear and quadratic moments that can generate an optimal GMM estimator with the minimum asymptotic variance.

Keywords: Spatial autoregression, endogeneity, GMM, heteroskedasticity, efficiency

JEL classification: C13, C21, C31, C36, R15

1 Introduction

Spatial autoregressive (SAR) models are popular spatial econometric models in empirical research. Various estimation methods for SAR models have been considered, including, among others, the

*Corresponding author. E-mail addresses: jin.fei@live.com (F. Jin); wang.yuqin@sufe.edu.cn (Y. Wang).

two stage least squares (2SLS) (Lee, 2003), the quasi maximum likelihood (QML) (Ord, 1975; Lee, 2004), and the generalized method of moments (GMM) (Lee, 2007). QML is relatively computationally intensive, since it involves the computation of the determinants of square matrices with their dimensions equal to the sample size. GMM can employ both linear moments characterizing instrumental variables (IV) and quadratic moments capturing spatial dependence, which are motivated from the QML estimation. Thus, GMM estimators are generally more efficient than 2SLS estimators. They are also computationally simpler than QML estimators, because they avoid the computation of determinants. The generalized spatial two stage least squares (GS2SLS) in Kelejian and Prucha (1998) is a multiple-step method specially designed for SAR models with SAR disturbances (SARAR models). It is computationally simple, but parameters in the equation for the dependent variable are estimated using only linear moments. When innovations in SAR models are heteroskedastic with unknown forms, the GMM and GS2SLS estimators are studied in, respectively, Lin and Lee (2010) and Kelejian and Prucha (2010). Liu and Yang (2015) propose a modified QML method, where QML first order conditions are modified to be valid under unknown heteroskedasticity and consistent estimators are derived by solving the modified first order conditions.

SAR models that allow for endogenous regressors in addition to spatial lags of the dependent variable have also been studied in the literature. Fingleton and Le Gallo (2008) and Drukker et al. (2013) investigate the GS2SLS estimation, Liu (2012) considers the limited information maximum likelihood (LIML) estimation, Liu and Lee (2013) study the 2SLS estimation, and Liu and Saraiva (2015) propose the GMM estimation. As for SAR models without endogenous regressors, likelihood based methods are relatively intensive in computation, and GMM can be computationally simple and relatively efficient asymptotically. Gupta and Robinson (2015) and Gupta (2019) have also considered the 2SLS estimation of SAR models with endogenous regressors, in the context of increasingly many parameters or stochastic spatial weights matrices.

We note that Liu and Saraiva (2015) assume a linear reduced form for endogenous regressors when considering GMM estimators.[1] The reduced form excludes any spatial dependence and heterogeneity in the endogenous regressors, which might arise for spatial variables. It also excludes any nonlinear relation between the endogenous regressors and their IVs. As the reduced form is used to form moment conditions, if it is misspecified, then the GMM estimators will no longer be consistent in general. Fingleton and Le Gallo (2008) and Drukker et al. (2013) have not assumed reduced forms for endogenous regressors when they consider the GS2SLS estimation, but the GS2SLS estimator is asymptotically less efficient than the GMM estimator that employs both linear and quadratic moments, as mentioned above.

[1]A reduced form is also assumed in Liu (2012) when the LIML estimator is considered. It is needed to form a proper likelihood function.

In this paper, we consider the GMM estimation of an SARAR model with endogenous regressors, by employing both linear and quadratic moments. Firstly, we do not impose any reduced form of endogenous regressors. Secondly, we study both the cases with homoskedastic and heteroskedastic model innovations. Thus, our paper extends the study on SAR models with unknown heteroskedasticity to the case with both unknown heteroskedasticity and endogenous regressors. We prove that GMM estimators are consistent and asymptotically normal under regularity conditions. Lastly, in the homoskedastic case, among a class of GMM estimators with linear and quadratic moments, we derive the best one with a minimum asymptotic variance. The best GMM estimator can guide our selection of linear and quadratic moments.

This paper is organized as follows. Section 2 studies large sample properties of our proposed GMM estimators, Section 3 reports some Monte Carlo results on the finite sample performance of our GMM estimators, and Section 4 concludes. Proofs are collected in an Appendix.

2 GMM estimation

We consider the following SAR model with SAR disturbances and endogenous regressors:

$$Y_n = \lambda W_n Y_n + Z_n \gamma + X_n \beta + u_n, \; u_n = \rho M_n u_n + \epsilon_n, \tag{1}$$

where n is the sample size, Y_n is an $n \times 1$ vector of observations on the dependent variable, W_n and M_n are $n \times n$ spatial weights matrices with zero diagonals, Z_n is an $n \times k_z$ matrix of observations on endogenous regressors, X_n is an $n \times k_x$ matrix of observations on exogenous regressors, $\epsilon_n = [\epsilon_{n1}, \cdots, \epsilon_{nn}]'$ is an $n \times 1$ vector of independent disturbances with zero means, λ and ρ are scalar spatial dependence parameters, γ is a $k_z \times 1$ parameter vector, and β is a $k_x \times 1$ parameter vector. The exogenous variable matrix is assumed to be nonrandom for simplicity, and Z_n is stochastic. The spatial weights matrices W_n and M_n may or may not be the same in practice. We shall consider both the case where ϵ_{ni}'s are i.i.d. with mean zero and variance σ_0^2, and the case where ϵ_{ni}'s are independent with mean zero but they may have different variances σ_{ni}^2's. Let $\theta = [\lambda, \rho, \gamma', \beta']'$ and $\theta_0 = [\lambda_0, \rho_0, \gamma_0', \beta_0']'$ be the true value of θ. Denote $S_n(\lambda) = I_n - \lambda W_n$, $R_n(\rho) = I_n - \rho M_n$, $S_n = S_n(\lambda_0)$ and $R_n = R_n(\rho_0)$, where I_n is the $n \times n$ identity matrix.[2] Provided that S_n and R_n are invertible, $Y_n = S_n^{-1}(Z_n \gamma_0 + X_n \beta_0 + R_n^{-1} \epsilon_n)$.

Fingleton and Le Gallo (2008) and Drukker et al. (2013) have considered the GS2SLS estimation of model (1) with homoskedastic ϵ_{ni}'s. The GS2SLS estimation has several steps, which makes the computation easy, but the derivation of the joint asymptotic distribution of final estimators is relatively complicated. For the estimation of the parameters λ, γ and β in the equation for Y_n, only linear moments are used, but quadratic moments are not.

[2]A list of notations is provided in Appendix A for convenient reference.

Liu (2012) and Liu and Saraiva (2015) consider the estimation of model (1) with no SAR process on u_n, where $u_n = \epsilon_n$ and Z_n contains a single endogenous regressor. They assume that Z_n has a reduced form

$$Z_n = F_n \delta + v_n, \tag{2}$$

where F_n is an $n \times k_f$ matrix of exogenous variables, δ is a $k_f \times 1$ vector of coefficients, and v_n is a vector of i.i.d. disturbances which may correlate with ϵ_n in (1). The reduced form (2) specifies a fixed and restrictive linear relation between Z_n and F_n. It does not allow for spatial dependence in Z_n, unless variables in F_n show spatial dependence. The i.i.d. disturbances also exclude heterogeneity in Z_n. With (2) imposed, Liu (2012) considers the LIML estimation, and Liu and Saraiva (2015) consider the GMM estimation with moment conditions that are linear and quadratic forms of $[\epsilon_n', v_n']'$ at the true parameter values. Thus, if (2) is misspecified, then their estimators will be inconsistent in general.

We consider the estimation of model (1) without imposing a reduced form of Z_n. As no reduced form is imposed, a likelihood or quasi likelihood function might not be formulated. We investigate a GMM estimator with the following moment vector:

$$g_n(\theta) = \frac{1}{n}[\epsilon_n'(\theta) P_{1n} \epsilon_n(\theta), \cdots, \epsilon_n'(\theta) P_{k_p n} \epsilon_n(\theta), \epsilon_n'(\theta) Q_n]', \tag{3}$$

where $\epsilon_n(\theta) = R_n(\rho)[S_n(\lambda) Y_n - Z_n \gamma - X_n \beta]$, P_{jn}'s are $n \times n$ matrices, and Q_n is an $n \times k_q$ IV matrix. The total number of moments $k_g = k_p + k_q$ is non-smaller than the total number of parameters $k_\theta = 2 + k_z + k_x$. In the homoskedastic case, P_{jn}'s are required to have zero traces, as $\mathrm{E}(\epsilon_n' P_{jn} \epsilon_n) = \sigma_0^2 \operatorname{tr}(P_{jn})$; in the heteroskedastic case, P_{jn}'s are required to have zero diagonals, as $\mathrm{E}(\epsilon_n' P_{jn} \epsilon_n) = \operatorname{tr}(P_{jn} \Sigma_n)$ is a weighted sum of the diagonal elements of P_{jn}, where $\Sigma_n = \operatorname{diag}(\sigma_{n1}^2, \ldots, \sigma_{nn}^2)$ is a diagonal matrix of σ_{ni}^2's. The P_{jn}'s can be, e.g., W_n, M_n, $W_n^2 - I_n \operatorname{tr}(W_n^2)/n$ and $M_n^2 - I_n \operatorname{tr}(M_n^2)/n$ in the homoskedastic case, and they can be W_n, M_n, $W_n^2 - \operatorname{diag}(W_n^2)$ and $M_n^2 - \operatorname{diag}(M_n^2)$ in the heteroskedastic case, where $\operatorname{diag}(A)$ for a square matrix A denotes a diagonal matrix formed by the diagonal elements of A. The quadratic moments for SAR models are motivated from the QML estimation, which may significantly improve the estimation efficiency for SAR models (Lee, 2007; Lin and Lee, 2010). If Z_n has an IV F_n, then the IV matrix Q_n can be the matrix formed by the independent columns of $[X_n, F_n, W_n X_n, W_n F_n, W_n^2 X_n, W_n^2 F_n]$. The GMM estimator $\hat{\theta}_{\text{GMM}}$ with the moment vector $g_n(\theta)$ and a weighting matrix $a_n' a_n$ has the objective function

$$\min_{\theta \in \Theta} g_n'(\theta) a_n' a_n g_n(\theta), \tag{4}$$

where Θ is the parameter space of θ, and a_n is a $k_a \times k_g$ matrix with a limit a_0 by design. Here the row dimension k_a can be greater or non-greater than k_g for generality.

Some basic regularity conditions are summarized in the following assumptions.

Assumption 1. *Either (a) ϵ_{ni}'s in $\epsilon_n = [\epsilon_{n1}, \ldots, \epsilon_{nn}]'$ are i.i.d. $(0, \sigma_0^2)$ and the moment* $\mathrm{E}(|\epsilon_{ni}|^{4+\iota})$ *exists for some $\iota > 0$, or (b) ϵ_{ni}'s are independent with mean zero and variances σ_{ni}^2's, and* $\sup_n \sup_{1\leq i\leq n} \mathrm{E}(|\epsilon_{ni}|^{4+\iota}) < \infty$ *for some $\iota > 0$.*[3]

Assumption 2. *The W_n and M_n have zero diagonals, and $\{W_n\}$, $\{M_n\}$, $\{S_n^{-1}\}$ and $\{R_n^{-1}\}$ are bounded in both row and column sum matrix norms.*

Assumption 3. *Elements of X_n and $\mathrm{E}(Z_n)$ are uniformly bounded constants, and* $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ *exists and is positive definite.*

Assumption 4. *Elements of Q_n are uniformly bounded constants, and $\{P_{jn}\}$ for $j = 1, \ldots, k_p$ are bounded in both row and column sum matrix norms.*

Assumption 5. *The parameter space Θ of θ is a compact subset of $\mathbb{R}^{k_\theta}$.*

In Assumption 1, the existence of moments of disturbances with an order higher than four is for the applicability of a central limit theorem for linear and quadratic forms in Kelejian and Prucha (2001). In Assumption 2, the diagonal elements of W_n and M_n are required to be zero to exclude self-influence. The boundedness condition on spatial weights matrices, originated in Kelejian and Prucha (1998, 1999), is a standard condition in the spatial econometric literature that limits the degree of spatial dependence to be manageable. Since the analysis involves the matrix inverses S_n^{-1} and R_n^{-1}, they are also assumed to be bounded in both row and column sum matrix norms. In Delgado and Robinson (2015) and Gupta and Robinson (2018), the consistency of relevant estimators is proved under the assumption of boundedness in the spectral norm of related matrices, although asymptotic distributions are still proved under the assumption of boundedness in the row and column sum norms. Since the spectral norm is non-greater than the row and column sum norms, their assumption is weaker. We also provide in Appendix D a proof of the consistency of our GMM estimator under the weaker assumption. As in Lee (2004), elements of X_n are assumed to be nonstochastic for simplicity in Assumption 3. In Assumption 4, elements of Q_n are also assumed to be constants, as Q_n can be functions of X_n, W_n, M_n and other IVs; and the boundedness condition on P_{jn}'s is similar to that on W_n and M_n, as P_{jn}'s relate to W_n and M_n. The compactness of the parameter space in Assumption 5 is standard for an extremum estimator.

We now discuss the identification condition for $\lim_{n\to\infty} \mathrm{E}[g_n(\theta)]$ to be uniquely zero at $\theta = \theta_0$. Denote $\bar{Z}_n = \mathrm{E}(Z_n)$ and $\check{Z}_n = Z_n - \bar{Z}_n$. Due to the endogeneity of Z_n, $\mathrm{cov}(\check{Z}_n, \epsilon_n) \neq 0$. Using $Y_n = S_n^{-1}(Z_n\gamma_0 + X_n\beta_0 + R_n^{-1}\epsilon_n)$, we have

$$\mathrm{E}[Q_n'\epsilon_n(\theta)] = Q_n'\bar{\epsilon}_n(\theta), \tag{5}$$

[3]In the homoskedastic case, as pointed out by an anonymous referee, we may omit the subscript n of ϵ_{ni}'s, e.g., denote $\epsilon_n = [\varepsilon_1, \cdots, \varepsilon_n]'$. But in the heteroskedastic case, we need the subscript n for ϵ_{ni}'s, in order to show that ϵ_{ni} can have different variances for different n.

$$\mathrm{E}[\epsilon_n'(\theta)P_{jn}\epsilon_n(\theta)] = \bar{\epsilon}_n'(\theta)P_{jn}\bar{\epsilon}_n(\theta) + \mathrm{E}[\check{\epsilon}_n'(\theta)P_{jn}\check{\epsilon}_n(\theta)], \tag{6}$$

where

$$\bar{\epsilon}_n(\theta) = R_n(\rho)[(\lambda_0 - \lambda)T_n(\bar{Z}_n\gamma_0 + X_n\beta_0) + \bar{Z}_n(\gamma_0 - \gamma) + X_n(\beta_0 - \beta)], \tag{7}$$

$$\check{\epsilon}_n(\theta) = R_n(\rho)[R_n^{-1}\epsilon_n + (\lambda_0 - \lambda)T_n(\check{Z}_n\gamma_0 + R_n^{-1}\epsilon_n) + \check{Z}_n(\gamma_0 - \gamma)], \tag{8}$$

with $T_n = W_n S_n^{-1}$. When $(\lambda, \gamma, \beta) = (\lambda_0, \gamma_0, \beta_0)$, as $\bar{\epsilon}_n(\theta) = 0$, $\mathrm{E}[Q_n'\epsilon_n(\theta)] = 0$ for any ρ, so the linear moments alone are not enough to identify the parameter ρ_0 in the disturbance process. But it is possible to identify other parameters from the linear moments. If $\lim_{n\to\infty} \frac{1}{n}Q_n'R_n(\rho)[T_n(\bar{Z}_n\gamma_0 + X_n\beta_0), \bar{Z}_n, X_n]$ has full column rank for any ρ in its parameter space $\boldsymbol{\rho}$, then the linear moment part of $\lim_{n\to\infty} \mathrm{E}[g_n(\theta)] = 0$, i.e. $\lim_{n\to\infty} \frac{1}{n}\mathrm{E}[Q_n'\epsilon_n(\theta)] = 0$, implies that $(\lambda, \gamma, \beta) = (\lambda_0, \gamma_0, \beta_0)$. It is possible that $\lim_{n\to\infty} \frac{1}{n}Q_n'R_n(\rho)[T_n(\bar{Z}_n\gamma_0 + X_n\beta_0), \bar{Z}_n, X_n]$ has reduced column rank for some $\rho \in \boldsymbol{\rho}$, then the identification of some parameters in (λ, γ, β) would reply on the quadratic moments. Sufficient conditions for $\lim_{n\to\infty} \mathrm{E}[g_n(\theta)]$ to be uniquely zero at $\theta = \theta_0$ is presented in the following assumption.

Assumption 6. *Either (i)* $\lim_{n\to\infty} \frac{1}{n}Q_n'R_n(\rho)[T_n(\bar{Z}_n\gamma_0 + X_n\beta_0), \bar{Z}_n, X_n]$ *has full column rank for any* $\rho \in \boldsymbol{\rho}$*, and (C.2) or (C.3) holds; or (ii)* $\lim_{n\to\infty} \frac{1}{n}Q_n'R_n(\rho)X_n$ *has full column rank for any* $\rho \in \boldsymbol{\rho}$*, and (C.4) holds.*

Lemma 1. *Under Assumption 6, for* $\theta \in \Theta$*,* $\lim_{n\to\infty} \mathrm{E}[g_n(\theta)]$ *is uniquely zero at* $\theta = \theta_0$*.*

The identification conditions in Assumption 6 can be compared with the corresponding Assumptions 7–8 in Drukker et al. (2013). The estimation in Drukker et al. (2013) is carried out in several steps, where the parameters (λ, γ, β) are estimated by 2SLS and the parameter ρ is estimated with moments quadratic in residuals computed using the first step estimate, thus their identification conditions are more similar to the conditions in Assumption 6(i), where (λ, γ, β) is identified from the linear moments and ρ is identified from the quadratic moments. Since our GMM approach estimates ρ jointly with (λ, γ, β), unlike that in Drukker et al. (2013), our rank condition on $\lim_{n\to\infty} \frac{1}{n}Q_n'R_n(\rho)[T_n(\bar{Z}_n\gamma_0 + X_n\beta_0), \bar{Z}_n, X_n]$ involves $\rho \in \boldsymbol{\rho}$. Assumption 8 in Drukker et al. (2013) is in terms of the minimum eigenvalue of a relevant matrix, while our corresponding condition (C.3) requires some matrices to have full rank in the limit, which seems to be more explicit. Furthermore, as (λ, γ, β) is estimated jointly with ρ in our approach, some of the parameters in (λ, γ, β) can be identified from the quadratic moments, thus we have the identification condition in Assumption 6(ii).

While Assumption 6 guarantees $\lim_{n\to\infty} \mathrm{E}[g_n(\theta)]$ to be uniquely zero at $\theta = \theta_0$, it is necessary but may not be sufficient for $\lim_{n\to\infty} a_n \mathrm{E}[g_n(\theta)]$ to be uniquely zero at $\theta = \theta_0$. Note that the dimension of a_n is $k_a \times k_g$, where k_a can be smaller than k_g. For example, $[\kappa - \kappa_0, \tau - \tau_0]' = 0$

implies that $[\kappa, \tau] = [\kappa_0, \tau_0]$, but $(\kappa - \kappa_0) + (\tau - \tau_0) = 0$ does not have the implication. If $\lim_{n\to\infty} a_n$ has full column rank, where $k_a \geq k_g$, then Assumption 6 is sufficient for $\lim_{n\to\infty} a_n \mathrm{E}[g_n(\theta)]$ to be uniquely zero at $\theta = \theta_0$. As we would like to consider a general class of GMM estimators with the weighting matrix $a_n' a_n$ and investigate the optimal choice of $a_n' a_n$ as in Hansen (1982), the following assumption is imposed.

Assumption 7. $\lim_{n\to\infty} a_n \mathrm{E}[g_n(\theta)]$ *is uniquely zero at* $\theta = \theta_0$.

The following assumption contains some technical conditions needed for the consistency of the GMM estimator $\hat{\theta}_{\text{GMM}}$.

Assumption 8. $\frac{1}{n} Q_n' A_n \check{Z}_n = o_p(1)$, *and for each* $j = 1, \ldots, m$, $\frac{1}{n} \Upsilon_n' P_{jn}^s A_n \check{Z}_n = o_p(1)$, $\frac{1}{n} \check{Z}_n' B_n' P_{jn}^s A_n \check{Z}_n - \frac{1}{n} \mathrm{E}(\check{Z}_n' B_n' P_{jn}^s A_n \check{Z}_n) = o_p(1)$, $\frac{1}{n} \epsilon_n' C_n' P_{jn}^s A_n \check{Z}_n - \frac{1}{n} \mathrm{E}(\epsilon_n' C_n' P_{jn}^s A_n \check{Z}_n) = o_p(1)$, $\frac{1}{n} \mathrm{E}(\check{Z}_n' B_n' P_{jn}^s A_n \check{Z}_n) = O(1)$ *and* $\frac{1}{n} \mathrm{E}(\epsilon_n' C_n' P_{jn}^s A_n \check{Z}_n) = O(1)$, *where* $\Upsilon_n = [\Upsilon_{1n}, M_n \Upsilon_{1n}]$ *with* $\Upsilon_{1n} = [T_n \bar{Z}_n, T_n X_n, \bar{Z}_n, X_n]$; A_n *and* B_n *are either* I_n, M_n, T_n *or* $M_n T_n$; *and* $C_n = I_n$, H_n, $T_n R_n^{-1}$ *or* $M_n T_n R_n^{-1}$.

If the endogenous regressors Z_n have the reduced form (2), it is straightforward to verify Assumption 8. Since we do not assume a reduced form of Z_n, we impose the convergence and order conditions in Assumption 8. As mentioned above, we may allow for spatial dependence and heterogeneity in Z_n. In those situations, the conditions in the above assumption can be verified by the law of large numbers for spatial near-epoch processes, which are processes with weak spatial dependence developed by Jenish and Prucha (2012). Based on spatial near-epoch processes, the supplementary file of Jin and Lee (2018) provides some primitive conditions for orders of terms similar to those in the above assumption. Thus, we maintain the relatively high level assumption above for simplicity. The consistency of $\hat{\theta}_{\text{GMM}}$ holds under the above assumptions.

Proposition 1. *Under Assumptions 1–5 and 7–8, the GMM estimator* $\hat{\theta}_{\text{GMM}}$ *is consistent.*

As usual, to derive the asymptotic distribution of $\hat{\theta}_{\text{GMM}}$, we require θ_0 to be in the interior $\mathrm{int}(\Theta)$ of the parameter space Θ.

Assumption 9. $\theta_0 \in \mathrm{int}(\Theta)$.

The variance matrix Ω_n of $\sqrt{n} g_n(\theta_0)$ can be derived by, e.g., Lemma 2 in Jin and Lee (2012) on the covariances of linear and quadratic forms. For any square matrix A, let $A^s = A + A'$, $\mathrm{vec}(A)$ be the vectorization of A, and d_A be a column vector formed by the diagonal elements of A. When ϵ_{ni}'s are i.i.d., Ω_n has the expression

$$\Omega_n = \frac{1}{n} \begin{pmatrix} (\mu_{40} - 3\sigma_0^4) \omega_{nd}' \omega_{nd} + \frac{\sigma_0^4}{2} \omega_n' \omega_n & \mu_{30} \omega_{nd}' Q_n \\ \mu_{30} Q_n' \omega_{nd} & \sigma_0^2 Q_n' Q_n \end{pmatrix}, \tag{9}$$

where $\mu_{30} = \mathrm{E}(\epsilon_{ni}^3)$, $\mu_{40} = \mathrm{E}(\epsilon_{ni}^4)$, $\omega_{nd} = [d_{P_{1n}}, \cdots, d_{P_{k_pn}}]$, and $\omega_n = [\mathrm{vec}(P_{1n}^s), \cdots, \mathrm{vec}(P_{k_pn}^s)]$; when ϵ_{ni}'s are heteroskedastic, as P_{jn}'s have zero diagonals,

$$\Omega_n = \frac{1}{n}\begin{pmatrix} \frac{1}{2}\omega_{nh}'\omega_{nh} & \\ & Q_n'\Sigma_n Q_n \end{pmatrix}, \tag{10}$$

where $\omega_{nh} = [\mathrm{vec}(\Sigma_n^{1/2}P_{1n}^s\Sigma_n^{1/2}), \cdots, \mathrm{vec}(\Sigma_n^{1/2}P_{k_pn}^s\Sigma_n^{1/2})]$ with $\Sigma_n^{1/2} = \mathrm{diag}(\sigma_{n1}, \ldots, \sigma_{nn})$. As $\mathrm{tr}(AB) = \mathrm{vec}'(A')\,\mathrm{vec}(B)$ for two conformable square matrices A and B, the (j,k)th element of $\omega_n'\omega_n$ is $\mathrm{tr}(P_{jn}^sP_{kn}^s)$, and the (j,k)th element of $\omega_{nh}'\omega_{nh}$ is $\mathrm{tr}(\Sigma_nP_{jn}^s\Sigma_nP_{kn}^s)$.

The gradient matrix $G_n = \mathrm{E}(\frac{\partial g_n(\theta_0)}{\partial\theta'})$ has the following expression:

$$G_n = -\frac{1}{n}\begin{pmatrix} \mathrm{E}(\epsilon_n'P_{1n}^sR_n\zeta_n) & \mathrm{E}(\epsilon_n'P_{1n}^sH_n\epsilon_n) & \mathrm{E}(\epsilon_n'P_{1n}^sR_n\check{Z}_n) & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{E}(\epsilon_n'P_{k_pn}^sR_n\zeta_n) & \mathrm{E}(\epsilon_n'P_{k_pn}^sH_n\epsilon_n) & \mathrm{E}(\epsilon_n'P_{k_pn}^sR_n\check{Z}_n) & 0 \\ Q_n'R_nT_n(\bar{Z}_n\gamma_0 + X_n\beta_0) & 0 & Q_n'R_n\bar{Z}_n & Q_n'R_nX_n \end{pmatrix}, \tag{11}$$

where $\zeta_n = T_n(\check{Z}_n\gamma_0 + R_n^{-1}\epsilon_n)$. For $\hat{\theta}_{\mathrm{GMM}}$ to be $\sqrt{n}$-consistent, $\lim_{n\to\infty} a_nG_n$ needs to have full column rank, for which a necessary condition is that $\lim_{n\to\infty} G_n$ has full column rank. The following Assumption 10 guarantees that $\lim_{n\to\infty} G_n$ has full column rank. Let

$$G_{1n} = -\frac{1}{n}\begin{pmatrix} \mathrm{E}(\epsilon_n'P_{1n}^sR_n\zeta_n) & \mathrm{E}(\epsilon_n'P_{1n}^sH_n\epsilon_n) & \mathrm{E}(\epsilon_n'P_{1n}^sR_n\check{Z}_n) \\ \vdots & \vdots & \vdots \\ \mathrm{E}(\epsilon_n'P_{k_pn}^sR_n\zeta_n) & \mathrm{E}(\epsilon_n'P_{k_pn}^sH_n\epsilon_n) & \mathrm{E}(\epsilon_n'P_{k_pn}^sR_n\check{Z}_n) \end{pmatrix}. \tag{12}$$

Assumption 10. *In the case of Assumption 6(i),* $\lim_{n\to\infty}\frac{1}{n}\mathrm{E}(\epsilon_n'P_{jn}^sH_n\epsilon_n) \neq 0$ *for some* $1 \le j \le k_p$*; in the case of Assumption 6(ii),* $\lim_{n\to\infty} G_{1n}$ *has full column rank.*

In the case that $\lim_{n\to\infty}\frac{1}{n}Q_n'R_n(\rho)[T_n(\bar{Z}_n\gamma_0+X_n\beta_0), \bar{Z}_n, X_n]$ has full column rank for any $\rho \in \boldsymbol{\rho}$, which is a condition in Assumption 6(i), the above assumption excludes the second situation in (C.3), where each $\lim_{n\to\infty}\frac{1}{n}\mathrm{E}(\epsilon_n'P_{jn}^sH_n\epsilon_n)$ is zero.

Proposition 2. *Under Assumptions 1–5 and 7–9, if* $\lim_{n\to\infty} a_nG_n$ *has full column rank, the GMM estimator* $\hat{\theta}_{\mathrm{GMM}}$ *has the asymptotic distribution*

$$\sqrt{n}(\hat{\theta}_{\mathrm{GMM}} - \theta_0) \xrightarrow{d} N\Big(0, \lim_{n\to\infty}(G_n'a_n'a_nG_n)^{-1}G_n'a_n'a_n\Omega_na_n'a_nG_n(G_n'a_n'a_nG_n)^{-1}\Big).$$

The gradient matrix G_n in (11) involves the correlation between $\check{Z}_n$ and ϵ_n with an unknown form, thus the explicit expression of G_n cannot be used to estimate G_n using a plug-in approach. However, G_n can be estimated by $\frac{\partial g_n(\hat{\theta}_{\mathrm{GMM}})}{\partial\theta'}$. Since each element of $g_n(\theta)$ is a polynomial function of θ, by the proof of Proposition 2, $\frac{\partial g_n(\hat{\theta}_{\mathrm{GMM}})}{\partial\theta'}$ is a consistent estimator of $\lim_{n\to\infty} G_n$.

As in Hansen (1982), the optimal weighting matrix $a_n'a_n$ is the matrix inverse Ω_n^{-1} of the variance matrix. To formulate an optimal GMM (OGMM) estimator, $\lim_{n\to\infty}\Omega_n$ needs to be invertible, which implies that Ω_n is also invertible for a large enough n. Let $\xi = \frac{\sqrt{2}}{2}(\frac{\mu_{40}}{\sigma_0^4}-1-\frac{\mu_{30}^2}{\sigma_0^6})^{1/2}$, $P_{jn,\xi}^s = \xi\,\mathrm{diag}(P_{jn}^s)+[P_{jn}^s-\mathrm{diag}(P_{jn}^s)]$ and $\omega_{n\xi}=[\mathrm{vec}(P_{1n,\xi}^s),\cdots,\mathrm{vec}(P_{k_pn,\xi}^s)]$.[4] By Jin et al. (2020), Ω_n in (9) in the homoskedastic case can be rewritten as

$$\Omega_n = \frac{1}{n}\begin{pmatrix} \frac{\mu_{30}^2}{\sigma_0^2}\omega_{nd}'\omega_{nd} + \frac{\sigma_0^4}{2}\omega_{n\xi}'\omega_{n\xi} & \mu_{30}\omega_{nd}'Q_n \\ \mu_{30}Q_n'\omega_{nd} & \sigma_0^2 Q_n'Q_n \end{pmatrix}. \tag{13}$$

Then the following assumption guarantees that $\lim_{n\to\infty}\Omega_n$ is positive definite.

Assumption 11. *In the case of Assumption 1(a) with homoskedastic disturbances,* $\lim_{n\to\infty}\frac{1}{n}Q_n'Q_n$ *and* $\lim_{n\to\infty}\frac{1}{n}\{\frac{2\mu_{30}^2}{\sigma_0^6}\omega_{nd}'[I_n-Q_n(Q_n'Q_n)^{-1}Q_n']\omega_{nd}+\omega_{n\xi}'\omega_{n\xi}\}$ *are nonsingular; in the case of Assumption 1(b) with heteroskedastic disturbances,* $\lim_{n\to\infty}\frac{1}{n}Q_n'\Sigma_nQ_n$ *and* $\lim_{n\to\infty}\frac{1}{n}\omega_{nh}'\omega_{nh}$ *are nonsingular.*

For a block matrix $E = (\begin{smallmatrix} A & B \\ C & D\end{smallmatrix})$, where A and D are square matrices, if D is invertible, then

$$\begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} I & 0 \\ -D^{-1}C & I \end{pmatrix} = \begin{pmatrix} A-BD^{-1}C & 0 \\ 0 & D \end{pmatrix}. \tag{14}$$

Thus, if D is invertible, then E is invertible if and only if $A-BD^{-1}C$ is invertible. In Assumption 11, for the homoskedastic case, as the condition that $\lim_{n\to\infty}\Omega_n$ is positive definite implies that $\lim_{n\to\infty}\frac{1}{n}Q_n'Q_n$ is nonsingular, we can see by (14) that the nonsingularity of $\lim_{n\to\infty}\frac{1}{n}\{\frac{2\mu_{30}^2}{\sigma_0^6}\omega_{nd}'[I_n-Q_n(Q_n'Q_n)^{-1}Q_n']\omega_{nd}+\omega_{n\xi}'\omega_{n\xi}\}$ implies that of $\lim_{n\to\infty}\Omega_n$. Furthermore, (14) implies that the nonsingularity of $\lim_{n\to\infty}\frac{1}{n}\omega_{nd}'[I_n-Q_n(Q_n'Q_n)^{-1}Q_n']\omega_{nd}$ is guaranteed by that of $\lim_{n\to\infty}\frac{1}{n}[Q_n,\omega_{nd}]'[Q_n,\omega_{nd}]$. Thus, in the homoskedastic case of Assumption 11, when $\mu_{30}\neq 0$, either (i) the nonsingularity of $\lim_{n\to\infty}\frac{1}{n}[Q_n,\omega_{nd}]'[Q_n,\omega_{nd}]$ or (ii) the nonsingularity of both $\lim_{n\to\infty}\frac{1}{n}Q_n'Q_n$ and $\lim_{n\to\infty}\frac{1}{n}\omega_{n\xi}'\omega_{n\xi}$ guarantees the nonsingularity of $\lim_{n\to\infty}\Omega_n$; when $\mu_{30}=0$, (ii) is required. In the heteroskedastic case, as Ω_n is block diagonal, the conditions in Assumption 11 are straightforward. By the definitions of ω_{nd}, $\omega_{n\xi}$ and ω_{nh}, if P_{jn}'s are linearly dependent, then ω_{nd}, $\omega_{n\xi}$ and ω_{nh} do not have full column rank and Assumption 11 is not satisfied. Thus P_{jn}'s should be linearly independent under Assumption 11.

Let $\hat{\Omega}_n$ be a consistent estimator of $\lim_{n\to\infty}\Omega_n$. In the homoskedastic case, $\hat{\Omega}_n$ can be obtained by plugging consistent estimators of involved unknown parameters into Ω_n; in the heteroskedastic case, as in the approach of White (1980), $\hat{\Omega}_n$ can be obtained by replacing each Σ_n in Ω_n with the diagonal matrix $\hat{\Sigma}_n = \mathrm{diag}(\epsilon_{n1}^2(\hat{\theta}_{\text{GMM}}),\ldots,\epsilon_{nn}^2(\hat{\theta}_{\text{GMM}}))$, where $\epsilon_{nj}(\theta)$ is the jth element of $\epsilon_n(\theta)$. Lin and Lee (2010) prove the consistency of a White-type variance estimator in the heteroskedastic case for

[4] As $(\mu_{40}-\sigma_0^4)\sigma_0^2 = \mathrm{E}[(\epsilon_{ni}^2-\sigma_0^2)^2]\cdot\mathrm{E}(\epsilon_{ni}^2)\geq\mu_{30}^2$ by the Cauchy-Schwarz inequality, $\frac{\mu_{40}}{\sigma_0^4}-1-\frac{\mu_{30}^2}{\sigma_0^6}\geq 0$.

SAR models with no endogenous regressors. For SAR models with endogenous regressors, the proof is similar. With $\hat{\Omega}_n$, the feasible OGMM estimator $\hat{\theta}_{\text{OGMM}}$ is $\hat{\theta}_{\text{OGMM}} = \arg\min_{\theta\in\Theta} g_n'(\theta)\hat{\Omega}_n^{-1}g_n(\theta)$. Under regularity conditions, $\hat{\theta}_{\text{OGMM}}$ is consistent and asymptotically normal, and its objective function can be used to test for over-identification.

Proposition 3. *Under Assumptions 1–6 and 8–11, if $\hat{\Omega}_n = \Omega_n + o_p(1)$, the feasible OGMM estimator $\hat{\theta}_{\text{OGMM}}$ is consistent and follows the asymptotic distribution*

$$\sqrt{n}(\hat{\theta}_{\text{OGMM}} - \theta_0) \xrightarrow{d} N(0, \lim_{n\to\infty}(G_n'\Omega_n^{-1}G_n)^{-1}).$$

Besides, $ng_n'(\hat{\theta}_{\text{OGMM}})\hat{\Omega}_n^{-1}g_n(\hat{\theta}_{\text{OGMM}}) \xrightarrow{d} \chi^2(k_g - k_\theta)$.

Our results above are based on a given set of linear and quadratic moments. When disturbances are homoskedastic, there is an issue on the best selection of linear and quadratic moments (Liu et al., 2010).[5] We use the analytical method in Jin et al. (2020) to derive the best linear and quadratic moments. In their method, the variance matrix Ω_n is rewritten in a form $\frac{1}{n}\Delta_n'\Delta_n$ and the gradient matrix is rewritten in a form $-\frac{1}{n}\Delta_n'\Gamma_n$, where Γ_n is properly reformulated, so that the Cauchy-Schwarz inequality can be applied to derive a lower bound for the asymptotic variance $(G_n'\Omega_n^{-1}G_n)^{-1}$ in Proposition 3 and the lower bound can be attained by using some IV matrix Q_n and quadratic matrices P_{jn}'s. Let $\Psi_n = R_nT_n(\bar{Z}_n\gamma_0 + X_n\beta_0)$, l_n be an $n\times 1$ vector of ones, $\tilde{a}_n = a_n - \frac{1}{n}l_nl_n'a_n$ for any $n\times 1$ vector a_n, $B_{n,\cdot j}$ be the jth column of an $n\times k_b$ matrix B_n, $C_{1n} = R_n\,\mathrm{E}(\zeta_n\epsilon_n') - I_n\,\mathrm{tr}[R_n\,\mathrm{E}(\zeta_n\epsilon_n')]/n$, $C_{2n} = H_n - I_n\,\mathrm{tr}(H_n)/n$, and $C_{2+j,n} = R_n\,\mathrm{E}(\check{Z}_{n,\cdot j}\epsilon_n') - I_n\,\mathrm{tr}[R_n\,\mathrm{E}(\check{Z}_{n,\cdot j}\epsilon_n')]/n$ for $j = 1,\ldots,k_z$. Note that the sum of all elements in $\tilde{a}_n$ is zero, and C_{jn}'s have zero traces.

Proposition 4. *Suppose that Assumptions 1(a), 2–6 and 8–11 are satisfied.*

(a) The best Q_n and P_{jn}'s that can generate an OGMM estimator with the minimum asymptotic variance are

$$\begin{aligned} Q_n^* = \Big[&\Psi_n - \frac{\mu_{30}}{\xi^2\sigma_0^4}d_{C_{1n}} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\tilde{\Psi}_n,\ d_{C_{2n}},\\ &R_n\bar{Z}_{n,\cdot 1} - \frac{\mu_{30}}{\xi^2\sigma_0^4}d_{C_{3n}} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{R_n\bar{Z}_{n,\cdot 1}},\ \cdots,\ R_n\bar{Z}_{n,\cdot k_z} - \frac{\mu_{30}}{\xi^2\sigma_0^4}d_{C_{k_z+2,n}} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{R_n\bar{Z}_{n,\cdot k_z}},\\ &R_nX_{n,\cdot 1} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{R_nX_{n,\cdot 1}},\ \cdots,\ R_nX_{n,\cdot k_x} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{R_nX_{n,\cdot k_x}}\Big], \end{aligned}$$

[5] When disturbances are heteroskedastic, the best selection of linear and quadratic moments might exist (see Debarsy et al., 2015, for the theoretically best moments of the matrix exponential spatial specification and SAR models with no endogenous regressors), but a best GMM estimator would not be feasible due to the unknown Σ_n (Lin and Lee, 2010).

$P_{1n}^{*}=[C_{1n}-\operatorname{diag}(C_{1n})]+\frac{1}{\xi^2}\operatorname{diag}(C_{1n})-\frac{\mu_{30}}{2\xi^2\sigma_0^2}\operatorname{diag}\big(\tilde{\Psi}_n\big)$, $P_{2n}^{*}=[C_{2n}-\operatorname{diag}(C_{2n})]+\frac{1}{\xi^2}\operatorname{diag}(C_{2n})$, $P_{2+j,n}^{*}=[C_{2+j,n}-\operatorname{diag}(C_{2+j,n})]+\frac{1}{\xi^2}\operatorname{diag}(C_{2+j,n})-\frac{\mu_{30}}{2\xi^2\sigma_0^2}\operatorname{diag}\big(\widetilde{R_n\bar{Z}_{n,\cdot j}}\big)$ *for* $j=1,...,k_z$, *and* $P_{2+k_z+j,n}^{*}=\operatorname{diag}\big(\widetilde{R_nX_{n,\cdot j}}\big)$ *for* $j=1,\ldots,k_x$.

(b) The OGMM estimator with the above Q_n^* *and* P_{jn}^**'s,*[6] *denoted by* $\hat{\theta}_{\text{BGMM}}$, *has the asymptotic distribution* $\sqrt{n}(\hat{\theta}_{\text{BGMM}}-\theta_0)\xrightarrow{d}N(0,\lim_{n\to\infty}(\frac{1}{n}\Gamma_n'\Gamma_n)^{-1})$, *where*

$$\Gamma_n=\begin{pmatrix}\Gamma_{n,11} & -\frac{\mu_{30}}{\sqrt{2}\xi\sigma_0^4}\big[\operatorname{vec}\big(\operatorname{diag}(\widetilde{R_nX_{n,\cdot 1}})\big),\cdots,\operatorname{vec}\big(\operatorname{diag}(\widetilde{R_nX_{n,\cdot k_x}})\big)\big]\\ \frac{1}{\sigma_0}[\Psi_n,0_{n\times 1},R_n\bar{Z}_n] & \frac{1}{\sigma_0}R_nX_n\end{pmatrix}$$

with

$$\Gamma_{n,11}=\frac{1}{\sqrt{2}\sigma_0^2}\Big[\operatorname{vec}\Big(C_{1n,1/\xi}^s-\frac{\mu_{30}}{\xi\sigma_0^2}\operatorname{diag}(\tilde{\Psi}_n)\Big),\operatorname{vec}(C_{2n,1/\xi}^s),$$
$$\operatorname{vec}\Big(C_{3n,1/\xi}^s-\frac{\mu_{30}}{\xi\sigma_0^2}\operatorname{diag}\big(\widetilde{R_n\bar{Z}_{n,\cdot 1}}\big)\Big),\cdots,\operatorname{vec}\Big(C_{k_z+2,n,1/\xi}^s-\frac{\mu_{30}}{\xi\sigma_0^2}\operatorname{diag}\big(\widetilde{R_n\bar{Z}_{n,\cdot k_z}}\big)\Big)\Big].$$

(c) When $\mu_{30}=0$, *the IV* $d_{C_{2n}}$ *is redundant and the quadratic matrices* $\operatorname{diag}\big(\widetilde{R_nX_{n,\cdot j}}\big)$ *for* $j=1,\ldots,k_x$ *are redundant, so* Q_n^* *reduces to* $Q_n^*=[\Psi_n,R_n\bar{Z}_n,R_nX_n]$ *and* P_{jn}^**'s are* $P_{jn}^*=[C_{jn}-\operatorname{diag}(C_{jn})]+\frac{1}{\xi^2}\operatorname{diag}(C_{jn})$ *for* $j=1,\ldots,k_z+2$.

(d) For Q_n^* *and* P_{jn}^**'s in (a), if we use the IVs in each column of* Q_n^* *and the quadratic matrices in each* P_{jn}^* *separately, then* $Q_n^*=[\Psi_n,R_n\bar{Z}_n,R_nX_n,d_{C_{1n}},\cdots,d_{C_{k_z+2,n}},l_n]$,[7] *and* P_{jn}^**'s are* $C_{jn}-\operatorname{diag}(C_{jn})$ *for* $j=1,\ldots,k_z+2$, $\operatorname{diag}(C_{jn})$ *for* $j=1,\ldots,k_z+2$, $\operatorname{diag}(\tilde{\Psi}_n)$, $\operatorname{diag}(\widetilde{R_n\bar{Z}_{n,\cdot j}})$ *for* $j=1,\ldots,k_z$, *and* $\operatorname{diag}(\widetilde{R_nX_{n,\cdot j}})$ *for* $j=1,\ldots,k_x$.

When $\mu_{30}=0$, $Q_n^*=[\Psi_n,R_n\bar{Z}_n,R_nX_n]$, *and* P_{jn}^**'s are* $P_{jn}^*=C_{jn}-\operatorname{diag}(C_{jn})$ *for* $j=1,\ldots,k_z+2$, *and* $P_{k_z+2+j,n}^*=\operatorname{diag}(C_{jn})$ *for* $j=1,\ldots,k_z+2$.

Proposition $4(a)$ gives the best combined IVs and quadratic matrices, and Proposition $4(b)$ provides the corresponding asymptotic distribution. Note that the asymptotic variance is given by $(\frac{1}{n}\Gamma_n'\Gamma_n)^{-1}$ and we do not need to compute it with the sandwich form $(G_n'\Omega_n^{-1}G_n)^{-1}$ as in Proposition 3. The combined IVs and quadratic matrices in Proposition $4(a)$ are more complicated than those separate IVs and quadratic matrices in Proposition $4(d)$, but they avoid the use of more moments. Generally, the presence of endogenous regressors Z_n affects both the best IV matrix

[6]For this estimator, in Assumptions 4, 6, 8, 10 and 11, Q_n and P_{jn} become, respectively, Q_n^* and P_{jn}^*. Some of the assumptions can be directly verified, e.g., some elements of Q_n^* can be shown to be uniformly bounded, but some are not, e.g., the orders of terms involving $\check{Z}_n$.

[7]When X_n contains l_n as a column, which corresponds to the intercept term, and M_n is normalized to have row sums equal to one, R_nX_n generates a column of constants. In this situation, l_n should be removed from Q_n^* to avoid multicolinearity.

Q_n^* and the best quadratic matrices P_{jn}^*'s. But when $\mu_{30} = 0$, the endogeneity of Z_n, i.e., the correlation between $\check{Z}_n$ and ϵ_n, does not affect the best IV matrix Q_n^*, although it affects the best quadratic matrices. While $\bar{Z}_n$ and the correlation between $\check{Z}_n$ and ϵ_n are unknown, we can choose IVs and quadratic moments according to the implications of the above proposition. If we have an IV matrix F_n for Z_n, the 2SLS estimate of Z_n is $\hat{Z}_n = F_n(F_n'F_n)^{-1}F_n'Z_n$, which is an estimate of $\bar{Z}_n$. By Proposition 4(a), as $\zeta_n = T_n(\check{Z}_n\gamma_0 + R_n^{-1}\epsilon_n)$ and $\Psi_n = R_nT_n(\bar{Z}_n\gamma_0 + X_n\beta_0)$, P_{jn}'s can be taken as

$$\begin{aligned}
&\sigma_0^2[R_nT_nR_n^{-1} - \operatorname{diag}(R_nT_nR_n^{-1})] + \frac{\sigma_0^2}{\xi^2}\operatorname{diag}\big(R_nT_nR_n^{-1} - \frac{\operatorname{tr}(T_n)}{n}I_n\big) - \frac{\mu_{30}}{2\xi^2\sigma_0^2}\operatorname{diag}(\tilde{\Pi}_n),\\
&H_n - \operatorname{diag}(H_n) + \frac{1}{\xi^2}\operatorname{diag}\big(H_n - \frac{\operatorname{tr}(H_n)}{n}I_n\big), \operatorname{diag}(\widetilde{R_n\hat{Z}_{n,\cdot 1}}), \cdots, \operatorname{diag}(\widetilde{R_n\hat{Z}_{n,\cdot k_z}}),\\
&\operatorname{diag}(\widetilde{R_nX_{n,\cdot 1}}), \cdots, \operatorname{diag}(\widetilde{R_nX_{n,\cdot k_x}}),
\end{aligned} \tag{15}$$

where $\Pi_n = R_nT_n(\hat{Z}_n\gamma_0 + X_n\beta_0)$, and the IV matrix Q_n can be taken as

$$\begin{aligned}
&\Big[\Pi_n - \frac{\mu_{30}}{\xi^2\sigma_0^2}d_{R_nT_nR_n^{-1} - I_n\operatorname{tr}(T_n)/n} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\tilde{\Pi}_n,\ d_{H_n - I_n\operatorname{tr}(H_n)/n},\\
&\quad R_n\hat{Z}_{n,\cdot 1} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{R_n\hat{Z}_{n,\cdot 1}}, \cdots, R_n\hat{Z}_{n,\cdot k_z} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{R_n\hat{Z}_{n,\cdot k_z}},\\
&\quad R_nX_{n,\cdot 1} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{R_nX_{n,\cdot 1}}, \cdots, R_nX_{n,\cdot k_x} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{R_nX_{n,\cdot k_x}}\Big].
\end{aligned} \tag{16}$$

When $\mu_{30} = 0$, the quadratic matrices $\operatorname{diag}(\widetilde{R_n\hat{Z}_{n,\cdot 1}})$, ..., $\operatorname{diag}(\widetilde{R_n\hat{Z}_{n,\cdot k_z}})$, $\operatorname{diag}(\widetilde{R_nX_{n,\cdot 1}})$, ..., $\operatorname{diag}(\widetilde{R_nX_{n,\cdot k_x}})$ in (15) are redundant, and the IV $d_{H_n - I_n\operatorname{tr}(H_n)/n}$ in (16) is redundant. The unknown parameters in P_{jn}'s and Q_n can be replaced by their consistent estimators, which will not affect the asymptotic distribution of the corresponding OGMM estimator, as in Liu et al. (2010) for SAR models without endogenous regressors.

Model (1) nests the special case of an SARAR model with no endogenous regressors, for which the best IV matrix Q_n and the best P_{jn}'s can be deduced from Proposition 4. We can see that the results are the same as those in Lee and Liu (2010) for the SARAR model. Another special model of interest nested in model (1) is the SAR model with endogenous regressors and without SAR disturbances. We present the best linear and quadratic moments for such a model in Appendix B.

Note that $\xi = 1$ when $\mu_{30} = 0$ and $\mu_{40} = 3\sigma_0^4$, e.g., when ϵ_{ni}'s are normally distributed. In such a situation, by Proposition 4(c), $Q_n^* = [\Psi_n, R_n\bar{Z}_n, R_nX_n]$ and P_{jn}^*'s are $P_{jn}^* = C_{jn}$ for $j = 1, \ldots, k_z + 2$.

3 Monte Carlo

In this section, we conduct some Monte Carlo experiments to study the finite sample performance of the proposed GMM estimators.

We first generate data from model (1) with no SAR process on disturbances and with one endogenous regressor in Z_n, i.e.,

$$Y_n = \lambda W_n Y_n + Z_n \gamma + X_n \beta + \epsilon_n, \tag{17}$$

where elements ϵ_{ni}'s of ϵ_n are independent, W_n is a block diagonal matrix with each block being a row-normalized matrix for the study in Anselin (1988) on crime activities in 49 districts of Columbus, OH, X_n contains a variable randomly drawn from the standard normal distribution, $\lambda_0 = 0.5$, $\gamma_0 = 1$, and $\beta_0 = 1$. The endogenous regressor Z_n in (17) is generated from the following model:

$$Z_n = \kappa W_n Z_n + F_n \delta + v_n, \tag{18}$$

where F_n and $v_n = [v_{n1}, \dots, v_{nn}]'$ are independent, elements of F_n and v_n are random draws from the standard normal distribution, $\delta = 1$, and κ is either 0 or 0.5. Each element ϵ_{ni} of ϵ_n in (17) is equal to $\frac{1}{2} v_{ni} + \frac{\sqrt{3}}{2} \tau_{ni}$, where τ_{ni}'s in the homoskedastic case are randomly drawn from either the standard normal distribution or the gamma$(1, 1)$ distribution with its mean adjusted to be zero, which has unit variance, skewness 2 and excess kurtosis 6; and τ_{ni}'s in the heteroskedastic case are further multiplied by $\sqrt{c_{ni}}$, where c_{ni} is proportional to the number of nonzero elements in the ith row of W_n and the mean of c_{ni}'s is 1. Thus, the mean of ϵ_{ni}'s variances is 1.

We consider three OGMM estimators in the homoskedastic case: the first estimator BGMM is the theoretically best GMM estimator with linear and quadratic moments, with moments given in Corollary 1(a); for the second estimator GMM2, the IV matrix is $[X_n, F_n, W_n X_n, W_n F_n, W_n^2 X_n, W_n^2 F_n]$, and the quadratic moments have square matrices W_n and $W_n^2 - I_n \operatorname{tr}(W_n^2)/n$; and for the third estimator GMM3, the square matrices for quadratic moments and the IV matrix are in, respectively, (B.2) and (B.3), which are implied from the theoretically best P_{jn}'s and Q_n. In the first steps of BGMM, GMM2 and GMM3, identity matrices are used as weighting matrices in the GMM objective functions. BGMM provides a basis for comparisons. For GMM3, the unknown parameters in Q_n and P_{jn}'s are replaced by their first step estimates for GMM2, and the redundant IVs and quadratic matrices in the case of normally distributed τ_{ni}'s are excluded. In the heteroskedastic case, the quadratic moments with diagonal quadratic matrices are removed, and the quadratic matrices for other quadratic moments are modified to have zero diagonals. Corresponding to GMM2, the GMM estimator in Liu and Saraiva (2015) with the same IV matrix and quadratic matrices is computed for comparison purposes, which is denoted by GMM2-LS. The moment conditions in Liu and Saraiva (2015) at true parameters are linear and quadratic in ϵ_n and v_n, so there are 2 linear moments corresponding to each IV and 4 quadratic moments corresponding to each quadratic matrix. We also report results on the 2SLS estimator, for which the IV matrix is the same as that for GMM2. The sample size is either 196 or 392, and the number of Monte Carlo repetitions is $5,000$.

Table 1 reports the estimation results for the case with homoskedastic disturbances ϵ_{ni}'s. Note that GMM2-LS is consistent when $\kappa = 0$, but it is not when $\kappa = 0.5$. We observe that all estimators have relatively small biases when $\kappa = 0$, but GMM2-LS has relatively large biases when $\kappa = 0.5$ and the biases do not decrease as the sample size doubles from 196 to 392, while other estimators still have small biases for the case with $\kappa = 0.5$. When τ_{ni}'s are normally distributed, BGMM, GMM2 and GMM3 have similar standard deviations (SD); when τ_{ni}'s are gamma-distributed, BGMM and GMM3 have similar SDs, which are smaller than those of GMM2. In particular, with gamma-distributed τ_{ni}'s, BGMM and GMM3 show very significant efficiency improvement upon GMM2 for the parameters γ and β. For all estimators, as the SDs dominate biases, the root mean squared errors (RMSE) are similar to the SDs. For the case with $\kappa = 0$, GMM2-LS does not always have smaller SDs and RMSEs than GMM2. This is the case since GMM2-LS also estimates the parameter κ in the reduced form of Z_n, although GMM2-LS employs more moment conditions. 2SLS has significantly larger SDs for the spatial dependence parameter than other estimators. It also has the largest SD for β, and the second largest SDs for γ, which are only smaller than those of GMM2-LS. As the sample size increases from $n = 196$ to $n = 392$, the SDs and RMSEs of BGMM, GMM2 and GMM3 decrease.

Table 2 reports the estimation results for the case with heteroskedastic disturbances. GMM2-LS still has relatively small biases when $\kappa = 0$ and relatively large biases when $\kappa = 0.5$, while other estimators have smaller biases for both $\kappa = 0$ and $\kappa = 0.5$. BGMM, GMM2 and GMM3 have similar SDs and RMSEs.[8] The patterns for the performance of 2SLS are similar to those in the homoskedastic case.

We also generate data from model (1), where the spatial weights matrix M_n is based on the queen criterion and normalized to have row sums equal to one, the true ρ_0 is 0.2, and other settings are the same as those for model (17). For the theoretically best GMM estimator BGMM with linear and quadratic moments in the homoskedastic case, the moments are given in Proposition 4. For the second estimator GMM2, the IV matrix is $[X_n, F_n, W_nX_n, W_nF_n, W_n^2X_n, W_n^2F_n]$, and the quadratic moments have square matrices W_n, $W_n^2 - I_n \operatorname{tr}(W_n^2)/n$, M_n and $M_n^2 - I_n \operatorname{tr}(M_n^2)/n$. For the third estimator GMM3, the square matrices for quadratic moments and the IV matrix are in, respectively, (15) and (16), and the involved unknown parameters in the square matrices and IV matrix are replaced by their first step estimates for GMM2. GMM2-LS is not considered since it has not taken into account the SAR process in disturbances and thus it is not expected to perform well. The GS2SLS estimator in Drukker et al. (2013) is also considered, which uses

[8]Note that BGMM in the heteroskedastic case is no longer the theoretically best GMM estimator with linear and quadratic moments. In the heteroskedastic case, BGMM uses moment conditions modified from the best linear and quadratic moments for the homoskedastic case, where the IV matrix is the same and the quadratic matrices are modified to have zero diagonals.

Table 1: Estimation results of model (17) with homoskedastic disturbances

		λ	γ	β
$n = 196$				
Normally distributed τ_{ni}'s, $\kappa = 0$	BGMM	-0.002[0.042]0.042	-0.002[0.072]0.072	-0.003[0.071]0.071
	GMM2	0.003[0.042]0.042	0.010[0.070]0.071	-0.004[0.071]0.071
	GMM2-LS	0.004[0.038]0.039	-0.006[0.075]0.075	-0.002[0.062]0.062
	GMM3	-0.002[0.042]0.042	-0.002[0.072]0.072	-0.003[0.071]0.071
	2SLS	0.003[0.072]0.073	0.002[0.072]0.072	-0.004[0.072]0.072
Normally distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	0.001[0.037]0.037	-0.000[0.071]0.071	-0.002[0.073]0.073
	GMM2	0.000[0.038]0.038	0.006[0.070]0.070	-0.002[0.072]0.072
	GMM2-LS	-0.065[0.038]0.075	0.048[0.079]0.092	0.011[0.065]0.066
	GMM3	-0.000[0.038]0.038	-0.003[0.073]0.073	-0.003[0.072]0.072
	2SLS	0.000[0.062]0.062	-0.000[0.074]0.074	-0.002[0.073]0.073
Gamma-distributed τ_{ni}'s, $\kappa = 0$	BGMM	-0.001[0.039]0.039	-0.003[0.062]0.062	-0.002[0.060]0.060
	GMM2	0.002[0.042]0.042	0.009[0.072]0.073	-0.003[0.073]0.073
	GMM2-LS	0.003[0.039]0.039	-0.007[0.077]0.077	-0.002[0.063]0.063
	GMM3	-0.001[0.040]0.040	-0.003[0.062]0.062	-0.002[0.060]0.060
	2SLS	0.002[0.075]0.075	0.002[0.074]0.074	-0.003[0.074]0.074
Gamma-distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	-0.002[0.035]0.035	-0.001[0.060]0.060	-0.003[0.061]0.061
	GMM2	-0.000[0.038]0.038	0.008[0.071]0.071	-0.003[0.073]0.073
	GMM2-LS	-0.066[0.039]0.077	0.051[0.080]0.094	0.011[0.066]0.067
	GMM3	-0.001[0.036]0.036	-0.001[0.061]0.061	-0.003[0.061]0.061
	2SLS	0.000[0.061]0.061	0.002[0.074]0.074	-0.003[0.074]0.074
$n = 392$				
Normally distributed τ_{ni}'s, $\kappa = 0$	BGMM	-0.001[0.029]0.029	-0.001[0.050]0.050	-0.002[0.051]0.051
	GMM2	0.002[0.029]0.029	0.005[0.050]0.050	-0.002[0.051]0.051
	GMM2-LS	0.002[0.026]0.026	-0.003[0.051]0.051	-0.001[0.044]0.044
	GMM3	-0.001[0.028]0.028	-0.001[0.050]0.050	-0.002[0.051]0.051
	2SLS	0.002[0.051]0.051	0.001[0.051]0.051	-0.002[0.052]0.052
Normally distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	-0.000[0.026]0.026	0.000[0.050]0.050	-0.002[0.051]0.051
	GMM2	-0.001[0.026]0.026	0.003[0.050]0.050	-0.002[0.051]0.051
	GMM2-LS	-0.066[0.027]0.071	0.052[0.056]0.076	0.011[0.046]0.047
	GMM3	-0.001[0.026]0.026	-0.002[0.051]0.051	-0.002[0.051]0.051
	2SLS	-0.001[0.043]0.043	0.000[0.053]0.053	-0.002[0.051]0.051
Gamma-distributed τ_{ni}'s, $\kappa = 0$	BGMM	-0.001[0.027]0.027	-0.001[0.043]0.043	-0.001[0.043]0.043
	GMM2	0.001[0.029]0.029	0.004[0.051]0.051	-0.001[0.052]0.052
	GMM2-LS	0.001[0.026]0.026	-0.004[0.052]0.053	-0.001[0.045]0.045
	GMM3	-0.001[0.027]0.027	-0.001[0.043]0.043	-0.001[0.043]0.043
	2SLS	0.001[0.050]0.050	0.001[0.052]0.052	-0.001[0.052]0.052
Gamma-distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	-0.001[0.024]0.024	-0.002[0.042]0.042	-0.000[0.043]0.043
	GMM2	-0.001[0.027]0.027	0.004[0.050]0.050	-0.002[0.051]0.051
	GMM2-LS	-0.067[0.027]0.072	0.052[0.056]0.077	0.012[0.046]0.048
	GMM3	-0.001[0.024]0.024	-0.002[0.042]0.042	-0.001[0.043]0.043
	2SLS	-0.001[0.044]0.044	0.001[0.052]0.052	-0.001[0.051]0.051

The three numbers in each cell are bias[SD]RMSE. $[\lambda_0, \gamma_0, \beta_0] = [0.5, 1, 1]$.

Table 2: Estimation results of model (17) with heteroskedastic disturbances

		λ	γ	β
$n = 196$				
Normally distributed τ_{ni}'s, $\kappa = 0$	BGMM	-0.002[0.041]0.041	-0.002[0.072]0.072	-0.003[0.073]0.074
	GMM2	0.003[0.041]0.041	0.010[0.071]0.072	-0.003[0.074]0.074
	GMM2-LS	0.004[0.036]0.036	-0.006[0.074]0.074	-0.003[0.064]0.064
	GMM3	-0.002[0.040]0.040	-0.002[0.072]0.072	-0.003[0.073]0.073
	2SLS	0.003[0.070]0.070	0.002[0.072]0.072	-0.003[0.073]0.074
Normally distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	0.000[0.036]0.036	0.001[0.071]0.071	-0.002[0.072]0.072
	GMM2	-0.000[0.036]0.036	0.008[0.071]0.071	-0.002[0.073]0.073
	GMM2-LS	-0.064[0.036]0.074	0.046[0.080]0.093	0.012[0.065]0.067
	GMM3	-0.001[0.036]0.036	-0.002[0.072]0.072	-0.002[0.072]0.072
	2SLS	-0.000[0.058]0.058	0.001[0.074]0.074	-0.001[0.073]0.073
Gamma-distributed τ_{ni}'s, $\kappa = 0$	BGMM	-0.002[0.039]0.039	-0.002[0.073]0.073	-0.004[0.072]0.072
	GMM2	0.002[0.040]0.040	0.010[0.071]0.072	-0.004[0.071]0.071
	GMM2-LS	0.003[0.035]0.035	-0.007[0.076]0.076	-0.003[0.063]0.063
	GMM3	-0.002[0.039]0.039	-0.002[0.073]0.073	-0.004[0.072]0.072
	2SLS	0.004[0.069]0.069	0.002[0.073]0.073	-0.005[0.072]0.073
Gamma-distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	-0.000[0.036]0.036	0.001[0.070]0.070	-0.004[0.073]0.073
	GMM2	-0.001[0.036]0.036	0.008[0.070]0.070	-0.004[0.072]0.072
	GMM2-LS	-0.064[0.037]0.074	0.046[0.080]0.092	0.009[0.066]0.067
	GMM3	-0.001[0.036]0.036	-0.002[0.072]0.072	-0.004[0.072]0.072
	2SLS	-0.000[0.059]0.059	0.001[0.074]0.074	-0.004[0.073]0.073
$n = 392$				
Normally distributed τ_{ni}'s, $\kappa = 0$	BGMM	-0.001[0.028]0.028	-0.001[0.051]0.051	-0.002[0.050]0.050
	GMM2	0.002[0.028]0.028	0.005[0.051]0.051	-0.002[0.050]0.050
	GMM2-LS	0.002[0.025]0.025	-0.004[0.052]0.052	-0.001[0.044]0.044
	GMM3	-0.001[0.028]0.028	-0.001[0.051]0.051	-0.002[0.050]0.050
	2SLS	0.002[0.048]0.048	0.001[0.051]0.051	-0.002[0.051]0.051
Normally distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	0.000[0.025]0.025	0.001[0.050]0.050	-0.001[0.051]0.051
	GMM2	-0.000[0.025]0.025	0.004[0.050]0.050	-0.001[0.052]0.052
	GMM2-LS	-0.065[0.026]0.070	0.050[0.057]0.076	0.013[0.046]0.048
	GMM3	-0.000[0.025]0.025	-0.001[0.051]0.051	-0.001[0.051]0.051
	2SLS	-0.000[0.041]0.041	0.001[0.052]0.052	-0.001[0.051]0.051
Gamma-distributed τ_{ni}'s, $\kappa = 0$	BGMM	-0.001[0.028]0.028	0.001[0.051]0.051	-0.002[0.052]0.052
	GMM2	0.001[0.028]0.028	0.007[0.050]0.051	-0.002[0.051]0.051
	GMM2-LS	0.002[0.025]0.025	-0.002[0.052]0.052	-0.001[0.045]0.045
	GMM3	-0.001[0.028]0.028	0.001[0.051]0.051	-0.002[0.051]0.051
	2SLS	0.001[0.048]0.048	0.003[0.051]0.051	-0.002[0.052]0.052
Gamma-distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	-0.000[0.025]0.025	-0.001[0.050]0.050	-0.002[0.050]0.050
	GMM2	-0.001[0.026]0.026	0.003[0.049]0.049	-0.002[0.050]0.050
	GMM2-LS	-0.065[0.026]0.070	0.047[0.057]0.075	0.012[0.045]0.047
	GMM3	-0.001[0.025]0.025	-0.002[0.051]0.051	-0.002[0.050]0.050
	2SLS	0.000[0.040]0.040	-0.001[0.052]0.052	-0.002[0.050]0.050

The three numbers in each cell are bias[SD]RMSE. $[\lambda_0, \gamma_0, \beta_0] = [0.5, 1, 1]$.

the same quadratic moments as those of GMM2 in estimating the spatial dependence parameter in disturbances. It corresponds to the 2SLS estimator for model (17). Tables 3–4 report the estimation results in the homoskedastic and heteroskedastic cases respectively. We observe similar patterns for BGMM, GMM2 and GMM3 as those in Tables 1–2 for model (17). GS2SLS is observed to have the largest SDs and RMSEs.

In sum, our proposed GMM estimators perform well in finite samples for both homoskedastic and heteroskedastic cases, while the 2SLS or GS2SLS estimator has larger SDs and RMSEs, and the GMM estimator in Liu and Saraiva (2015) can have a bad performance if the endogenous regressors have a reduced form different from the assumed one. In the homoskedastic case, the feasible moments implied by the theoretically best linear and quadratic moments can generate a GMM estimator that has similar performance to that of the theoretically best GMM estimator. When the disturbances follow a distribution with nonzero skew and nonzero excess kurtosis, this feasible best GMM estimator can have significant efficiency improvement upon the GMM estimator with some linear and quadratic moments that are commonly used in the literature. We thus suggest the use of this feasible best GMM estimator in practice.

4 Conclusion

This paper considers the GMM estimation of SAR models with SAR disturbances and endogenous regressors. We do not assume any reduced form of the endogenous regressors, thus we allow for spatial dependence and heterogeneity in endogenous regressors, and also allow for nonlinear relations among endogenous regressors and their instruments. Both linear and quadratic moments are employed for estimation. We prove that GMM estimators are consistent and asymptotically normal in both the homoskedastic and heteroskedastic cases. In the homoskedastic case, we derive the best linear and quadratic moments that can generate an OGMM estimator with the minimum asymptotic variance.

Acknowledgement. We are grateful to the Editor Esfandiar Maasoumi, an associate editor and two anonymous referees for helpful comments. Fei Jin gratefully acknowledges the financial support from the National Natural Science Foundation of China (No. 71973030 and No. 71833004). Yuqin Wang is also grateful for the financial support from the National Natural Science Foundation of China (No. 72103122).

Appendix A List of notations

I_n is the $n \times n$ identity matrix, $S_n(\lambda) = I_n - \lambda W_n$, $R_n(\rho) = I_n - \rho M_n$, $S_n = S_n(\lambda_0)$, $R_n = R_n(\rho_0)$, $\bar{Z}_n = \mathrm{E}(Z_n)$, $\check{Z}_n = Z_n - \bar{Z}_n$, $T_n = W_n S_n^{-1}$, $H_n = M_n R_n^{-1}$, $\Omega_n = \mathrm{var}[\sqrt{n} g_n(\theta_0)]$ and $G_n = \mathrm{E}(\frac{\partial g_n(\theta_0)}{\partial \theta'})$.

Table 3: Estimation results of model (1) with homoskedastic disturbances

		λ	ρ	γ	β
$n = 196$					
Normally distributed τ_{ni}'s, $\kappa = 0$	BGMM	-0.002[0.044]0.044	-0.001[0.127]0.127	0.003[0.072]0.072	-0.003[0.072]0.072
	GMM2	0.003[0.045]0.045	0.001[0.131]0.131	0.014[0.070]0.072	-0.003[0.071]0.071
	GMM3	-0.003[0.044]0.044	-0.008[0.127]0.127	-0.001[0.072]0.072	-0.003[0.072]0.072
	GS2SLS	0.003[0.076]0.076	-0.002[0.141]0.141	0.002[0.073]0.073	-0.004[0.072]0.072
Normally distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	-0.000[0.039]0.039	-0.004[0.133]0.133	0.000[0.071]0.071	-0.002[0.072]0.072
	GMM2	0.000[0.040]0.040	-0.002[0.133]0.134	0.010[0.071]0.071	-0.003[0.072]0.072
	GMM3	-0.001[0.040]0.040	-0.010[0.130]0.130	-0.003[0.073]0.073	-0.003[0.072]0.072
	GS2SLS	0.001[0.065]0.065	-0.002[0.145]0.145	-0.001[0.075]0.075	-0.002[0.073]0.073
Gamma-distributed τ_{ni}'s, $\kappa = 0$	BGMM	-0.000[0.042]0.042	0.004[0.131]0.132	-0.000[0.061]0.061	-0.003[0.061]0.061
	GMM2	0.004[0.045]0.045	-0.002[0.130]0.130	0.013[0.072]0.073	-0.004[0.073]0.073
	GMM3	-0.000[0.042]0.042	0.004[0.130]0.130	-0.000[0.062]0.062	-0.003[0.061]0.061
	GS2SLS	0.005[0.077]0.077	-0.003[0.145]0.145	0.001[0.074]0.074	-0.004[0.074]0.074
Gamma-distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	-0.002[0.038]0.038	0.000[0.133]0.133	0.000[0.060]0.060	-0.002[0.061]0.061
	GMM2	-0.000[0.040]0.040	-0.004[0.132]0.132	0.013[0.070]0.072	-0.002[0.073]0.073
	GMM3	-0.001[0.039]0.039	0.000[0.132]0.132	0.000[0.062]0.062	-0.002[0.062]0.062
	GS2SLS	-0.000[0.065]0.065	-0.004[0.143]0.143	0.002[0.075]0.075	-0.002[0.073]0.073
$n = 392$					
Normally distributed τ_{ni}'s, $\kappa = 0$	BGMM	-0.000[0.030]0.030	0.001[0.091]0.091	0.001[0.051]0.051	-0.001[0.051]0.051
	GMM2	0.002[0.031]0.031	0.001[0.092]0.092	0.007[0.050]0.051	-0.001[0.051]0.051
	GMM3	-0.001[0.030]0.030	-0.003[0.091]0.091	-0.001[0.051]0.051	-0.001[0.051]0.051
	GS2SLS	0.001[0.054]0.054	0.000[0.100]0.100	0.001[0.051]0.051	-0.001[0.052]0.052
Normally distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	-0.000[0.027]0.027	-0.003[0.092]0.092	0.001[0.051]0.051	-0.002[0.050]0.051
	GMM2	-0.000[0.027]0.027	-0.001[0.092]0.092	0.006[0.051]0.051	-0.002[0.051]0.051
	GMM3	-0.001[0.027]0.027	-0.005[0.090]0.091	-0.001[0.052]0.052	-0.002[0.050]0.051
	GS2SLS	-0.002[0.044]0.044	0.000[0.098]0.098	0.001[0.053]0.053	-0.002[0.051]0.051
Gamma-distributed τ_{ni}'s, $\kappa = 0$	BGMM	-0.000[0.028]0.028	0.004[0.092]0.092	0.001[0.042]0.042	-0.001[0.042]0.042
	GMM2	0.002[0.031]0.031	0.001[0.092]0.092	0.008[0.051]0.052	-0.001[0.050]0.050
	GMM3	-0.000[0.029]0.029	0.004[0.091]0.091	0.001[0.042]0.042	-0.001[0.042]0.042
	GS2SLS	-0.000[0.054]0.054	0.001[0.099]0.099	0.002[0.052]0.052	-0.001[0.051]0.051
Gamma-distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	-0.001[0.025]0.025	0.002[0.091]0.091	0.000[0.042]0.042	-0.001[0.042]0.042
	GMM2	-0.001[0.027]0.027	-0.001[0.091]0.091	0.007[0.050]0.051	-0.001[0.051]0.051
	GMM3	-0.000[0.025]0.025	0.002[0.090]0.090	0.000[0.042]0.042	-0.001[0.042]0.042
	GS2SLS	-0.001[0.044]0.044	-0.000[0.097]0.097	0.002[0.053]0.053	-0.001[0.052]0.052

The three numbers in each cell are bias[SD]RMSE. $[\lambda_0, \rho_0, \gamma_0, \beta_0] = [0.5, 0.2, 1, 1]$.

Table 4: Estimation results of model (1) with heteroskedastic disturbances

		λ	ρ	γ	β
$n = 196$					
Normally distributed τ_{ni}'s, $\kappa = 0$	BGMM	-0.001[0.042]0.042	-0.002[0.128]0.128	0.005[0.074]0.074	-0.003[0.073]0.073
	GMM2	0.003[0.043]0.043	0.000[0.131]0.131	0.017[0.073]0.075	-0.003[0.073]0.073
	GMM3	-0.002[0.043]0.043	-0.009[0.127]0.128	0.001[0.074]0.074	-0.003[0.073]0.073
	GS2SLS	0.002[0.073]0.073	-0.003[0.143]0.143	0.005[0.074]0.074	-0.003[0.074]0.074
Normally distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	-0.001[0.038]0.038	-0.003[0.134]0.134	0.003[0.071]0.071	-0.001[0.072]0.072
	GMM2	-0.001[0.039]0.039	0.000[0.133]0.133	0.013[0.071]0.072	-0.001[0.073]0.073
	GMM3	-0.002[0.038]0.038	-0.009[0.130]0.130	-0.001[0.073]0.073	-0.001[0.072]0.072
	GS2SLS	0.001[0.060]0.060	-0.001[0.143]0.143	0.002[0.074]0.074	-0.001[0.073]0.073
Gamma-distributed τ_{ni}'s, $\kappa = 0$	BGMM	-0.000[0.043]0.043	0.011[0.130]0.131	0.006[0.072]0.072	-0.004[0.072]0.072
	GMM2	0.003[0.043]0.043	0.003[0.132]0.132	0.015[0.071]0.072	-0.005[0.072]0.072
	GMM3	-0.001[0.044]0.044	0.005[0.130]0.130	0.001[0.072]0.072	-0.004[0.072]0.072
	GS2SLS	0.005[0.071]0.071	-0.003[0.143]0.143	0.003[0.073]0.073	-0.004[0.073]0.073
Gamma-distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	-0.000[0.038]0.038	0.008[0.132]0.132	0.002[0.070]0.070	-0.003[0.073]0.073
	GMM2	-0.001[0.038]0.038	-0.001[0.131]0.131	0.011[0.070]0.070	-0.002[0.072]0.072
	GMM3	-0.001[0.038]0.038	0.003[0.128]0.128	-0.001[0.072]0.072	-0.003[0.073]0.073
	GS2SLS	0.001[0.060]0.060	-0.004[0.141]0.141	-0.001[0.074]0.074	-0.003[0.073]0.073
$n = 392$					
Normally distributed τ_{ni}'s, $\kappa = 0$	BGMM	-0.000[0.029]0.029	-0.000[0.092]0.092	0.001[0.050]0.050	-0.003[0.051]0.051
	GMM2	0.002[0.030]0.030	0.001[0.093]0.093	0.007[0.050]0.051	-0.003[0.052]0.052
	GMM3	-0.001[0.029]0.029	-0.004[0.092]0.092	-0.001[0.050]0.050	-0.003[0.051]0.051
	GS2SLS	0.001[0.051]0.051	-0.001[0.100]0.100	0.000[0.051]0.051	-0.003[0.052]0.052
Normally distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	-0.000[0.026]0.026	-0.002[0.094]0.094	0.002[0.049]0.049	-0.002[0.050]0.050
	GMM2	-0.000[0.027]0.027	0.000[0.094]0.094	0.007[0.049]0.050	-0.002[0.050]0.050
	GMM3	-0.001[0.026]0.026	-0.005[0.092]0.093	0.000[0.050]0.050	-0.002[0.050]0.050
	GS2SLS	-0.001[0.042]0.042	0.000[0.100]0.100	0.002[0.051]0.052	-0.002[0.050]0.050
Gamma-distributed τ_{ni}'s, $\kappa = 0$	BGMM	0.001[0.029]0.029	0.003[0.092]0.092	0.004[0.050]0.050	-0.003[0.052]0.052
	GMM2	0.002[0.029]0.029	-0.002[0.092]0.092	0.009[0.050]0.051	-0.003[0.052]0.052
	GMM3	0.000[0.029]0.029	0.000[0.092]0.092	0.001[0.050]0.050	-0.003[0.052]0.052
	GS2SLS	0.002[0.050]0.050	-0.004[0.099]0.099	0.002[0.051]0.051	-0.003[0.053]0.053
Gamma-distributed τ_{ni}'s, $\kappa = 0.5$	BGMM	-0.001[0.026]0.026	0.003[0.092]0.092	0.001[0.050]0.050	-0.001[0.050]0.050
	GMM2	-0.001[0.026]0.026	-0.002[0.092]0.092	0.006[0.050]0.050	-0.001[0.050]0.050
	GMM3	-0.001[0.026]0.026	0.000[0.091]0.091	-0.001[0.051]0.051	-0.002[0.050]0.050
	GS2SLS	-0.001[0.042]0.042	-0.003[0.098]0.098	-0.000[0.052]0.052	-0.001[0.051]0.051

The three numbers in each cell are bias[SD]RMSE. $[\lambda_0, \rho_0, \gamma_0, \beta_0] = [0.5, 0.2, 1, 1]$.

For a square matrix A, $\text{diag}(A)$ is a diagonal matrix formed by the diagonal elements of A; for a vector a, $\text{diag}(a)$ is a diagonal matrix formed by the elements of a. $\sigma_{ni}^2 = \text{E}(\epsilon_{ni}^2)$ and $\Sigma_n = \text{diag}(\sigma_{n1}^2, \cdots, \sigma_{nn}^2)$.

For any square matrix A, $A^s = A + A'$, d_A is a column vector formed by the diagonal elements of A, and $\text{vec}(A)$ is the vectorization of A.

$\omega_n = [\text{vec}(P_{1n}^s), \cdots, \text{vec}(P_{k_p n}^s)]$, $\omega_{nd} = [d_{P_{1n}}, \cdots, d_{P_{k_p n}}]$, $\Sigma_n^{1/2} = \text{diag}(\sigma_{n1}, \cdots, \sigma_{nn})$, and $\omega_{nh} = [\text{vec}(\Sigma_n^{1/2} P_{1n}^s \Sigma_n^{1/2}), \cdots, \text{vec}(\Sigma_n^{1/2} P_{k_p n}^s \Sigma_n^{1/2})]$.

In the case that ϵ_{ni}'s are i.i.d., $\sigma_0^2 = \text{E}(\epsilon_{ni}^2)$, $\mu_{30} = \text{E}(\epsilon_{ni}^3)$, $\mu_{40} = \text{E}(\epsilon_{ni}^4)$, $\xi = \frac{\sqrt{2}}{2}(\frac{\mu_{40}}{\sigma_0^4} - 1 - \frac{\mu_{30}^2}{\sigma_0^6})^{1/2}$, $P_{jn,\xi}^s = \xi\,\text{diag}(P_{jn}^s) + [P_{jn}^s - \text{diag}(P_{jn}^s)]$, $\omega_{n\xi} = [\text{vec}(P_{1n,\xi}^s), \cdots, \text{vec}(P_{k_p n,\xi}^s)]$.

$\zeta_n = T_n(\check{Z}_n\gamma_0 + R_n^{-1}\epsilon_n)$, $\Psi_n = R_n T_n(\bar{Z}_n\gamma_0 + X_n\beta_0)$, l_n is an $n \times 1$ vector of ones, and $\tilde{a}_n = a_n - \frac{1}{n} l_n l_n' a_n$ for any $n \times 1$ vector a_n.

Appendix B Best linear and quadratic moments for a special model with no SAR disturbances

In this section, we consider the best linear and quadratic moments for the following model:

$$Y_n = \lambda W_n Y_n + Z_n\gamma + X_n\beta + \epsilon_n, \tag{B.1}$$

where elements of ϵ_n are i.i.d. with mean zero and variance σ_0^2, and the notations are the same as those in the main text. Model (B.1) is a special model nested in model (1). The GMM estimation of (B.1) still has the objective function (4), where the moment vector $g_n(\theta)$ has the form (3), but θ reduces to $[\lambda, \gamma, \beta']'$ and $\epsilon_n(\theta) = S_n(\lambda)Y_n - Z_n\gamma - X_n\beta$. The best linear and quadratic moments for model (B.1) are presented in the following Corollary 1. Let $\Phi_n = T_n(\bar{Z}_n\gamma_0 + X_n\beta_0)$, $\varsigma_n = T_n(\check{Z}_n\gamma_0 + \epsilon_n)$, $\mathbb{C}_{1n} = \text{E}(\varsigma_n\epsilon_n') - I_n\,\text{tr}[\text{E}(\varsigma_n\epsilon_n')]/n$, and $\mathbb{C}_{1+j,n} = \text{E}(\check{Z}_{n,\cdot j}\epsilon_n') - I_n\,\text{tr}[\text{E}(\check{Z}_{n,\cdot j}\epsilon_n')]/n$ for $j = 1, \ldots, k_z$.

Corollary 1. *The following results hold for the GMM estimation of model (B.1) with i.i.d. disturbances.*

(a) The best Q_n and P_{jn}'s that can generate an OGMM estimator with the minimum asymptotic variance are

$$\begin{aligned} Q_n^* = \Big[&\Phi_n - \frac{\mu_{30}}{\xi^2\sigma_0^4} d_{\mathbb{C}_{1n}} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\tilde{\Phi}_n, \\ &\bar{Z}_{n,\cdot 1} - \frac{\mu_{30}}{\xi^2\sigma_0^4} d_{\mathbb{C}_{2n}} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{\bar{Z}_{n,\cdot 1}}, \cdots, \bar{Z}_{n,\cdot k_z} - \frac{\mu_{30}}{\xi^2\sigma_0^4} d_{\mathbb{C}_{k_z+1,n}} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{\bar{Z}_{n,\cdot k_z}}, \\ &X_{n,\cdot 1} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{X_{n,\cdot 1}}, \cdots, X_{n,\cdot k_x} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{X_{n,\cdot k_x}}\Big], \end{aligned}$$

$P^*_{1n} = [\mathbb{C}_{1n} - \text{diag}(\mathbb{C}_{1n})] + \frac{1}{\xi^2}\,\text{diag}(\mathbb{C}_{1n}) - \frac{\mu_{30}}{2\xi^2\sigma_0^2}\,\text{diag}\big(\tilde{\Phi}_n\big)$, $P^*_{1+j,n} = [\mathbb{C}_{1+j,n} - \text{diag}(\mathbb{C}_{1+j,n})] + \frac{1}{\xi^2}\,\text{diag}(\mathbb{C}_{1+j,n}) - \frac{\mu_{30}}{2\xi^2\sigma_0^2}\,\text{diag}\big(\widetilde{\bar{Z}_{n,\cdot j}}\big)$ *for* $j = 1,...,k_z$, *and* $P^*_{1+k_z+j,n} = \text{diag}\big(\widetilde{X_{n,\cdot j}}\big)$ *for* $j = 1,\ldots,k_x$.

(b) The OGMM estimator with the above Q^*_n *and* P^*_{jn}*'s, denoted by* $\hat{\theta}_{\text{BGMM}}$*, has the asymptotic distribution* $\sqrt{n}(\hat{\theta}_{\text{BGMM}} - \theta_0) \xrightarrow{d} N(0, \lim_{n\to\infty}(\frac{1}{n}\Gamma'_n\Gamma_n)^{-1})$*, where*

$$\Gamma_n = \begin{pmatrix} \Gamma_{n,11} & -\frac{\mu_{30}}{\sqrt{2}\xi\sigma_0^4}\big[\text{vec}\big(\text{diag}(\widetilde{X_{n,\cdot 1}})\big), \cdots, \text{vec}\big(\text{diag}(\widetilde{X_{n,\cdot k_x}})\big)\big] \\ \frac{1}{\sigma_0}[\Phi_n, 0_{n\times 1}, \bar{Z}_n] & \frac{1}{\sigma_0}X_n \end{pmatrix}$$

with

$$\Gamma_{n,11} = \frac{1}{\sqrt{2}\sigma_0^2}\Big[\text{vec}\Big(\mathbb{C}^s_{1n,1/\xi} - \frac{\mu_{30}}{\xi\sigma_0^2}\,\text{diag}(\tilde{\Phi}_n)\Big), \\ \text{vec}\Big(\mathbb{C}^s_{2n,1/\xi} - \frac{\mu_{30}}{\xi\sigma_0^2}\,\text{diag}\big(\widetilde{\bar{Z}_{n,\cdot 1}}\big)\Big), \cdots, \text{vec}\Big(\mathbb{C}^s_{k_z+1,n,1/\xi} - \frac{\mu_{30}}{\xi\sigma_0^2}\,\text{diag}\big(\widetilde{\bar{Z}_{n,\cdot k_z}}\big)\Big)\Big].$$

(c) When $\mu_{30} = 0$*, the quadratic matrices* $\text{diag}\big(\widetilde{X_{n,\cdot j}}\big)$ *for* $j = 1,\ldots,k_x$ *are redundant, so* Q^*_n *reduces to* $Q^*_n = [\Phi_n, \bar{Z}_n, X_n]$ *and* P^*_{jn}*'s are* $P^*_{jn} = [\mathbb{C}_{jn} - \text{diag}(\mathbb{C}_{jn})] + \frac{1}{\xi^2}\,\text{diag}(\mathbb{C}_{jn})$ *for* $j = 1,\ldots,k_z+1$.

(d) For Q^*_n *and* P^*_{jn}*'s in (a), if we use the IVs in each column of* Q^*_n *and the quadratic matrices in each* P^*_{jn} *separately, then* $Q^*_n = [\Phi_n, \bar{Z}_n, X_n, d_{C_{1n}}, \cdots, d_{C_{k_z+1,n}}, l_n]$*, and* P^*_{jn}*'s are* $\mathbb{C}_{jn} - \text{diag}(\mathbb{C}_{jn})$ *for* $j = 1,\ldots,k_z+1$, $\text{diag}(\mathbb{C}_{jn})$ *for* $j = 1,\ldots,k_z+1$, $\text{diag}(\tilde{\Phi}_n)$, $\text{diag}(\widetilde{\bar{Z}_{n,\cdot j}})$ *for* $j = 1,\ldots,k_z$*, and* $\text{diag}(\widetilde{X_{n,\cdot j}})$ *for* $j = 1,\ldots,k_x$.

When $\mu_{30} = 0$, $Q^*_n = [\Phi_n, \bar{Z}_n, X_n]$*, and* P^*_{jn}*'s are* $P^*_{jn} = \mathbb{C}_{jn} - \text{diag}(\mathbb{C}_{jn})$ *for* $j = 1,\ldots,k_z+1$*, and* $P^*_{1+k_z+j,n} = \text{diag}(\mathbb{C}_{jn})$ *for* $j = 1,\ldots,k_z+1$.

If we have an IV matrix F_n for Z_n, the 2SLS estimate of Z_n is $\hat{Z}_n = F_n(F'_nF_n)^{-1}F'_nZ_n$, which is an estimate of $\bar{Z}_n$. By Corollary 1(a), as $\varsigma_n = T_n(\check{Z}_n\gamma_0 + \epsilon_n)$ and $\Phi_n = T_n(\bar{Z}_n\gamma_0 + X_n\beta_0)$, P_{jn}'s can be taken as

$$\sigma_0^2[T_n - \text{diag}(T_n)] + \frac{\sigma_0^2}{\xi^2}\,\text{diag}\big(T_n - \frac{\text{tr}(T_n)}{n}I_n\big) - \frac{\mu_{30}}{2\xi^2\sigma_0^2}\,\text{diag}(\tilde{\mathbf{\Pi}}_n), \\ \text{diag}(\widetilde{\hat{Z}_{n,\cdot 1}}), \cdots, \text{diag}(\widetilde{\hat{Z}_{n,\cdot k_z}}), \text{diag}(\widetilde{X_{n,\cdot 1}}), \cdots, \text{diag}(\widetilde{X_{n,\cdot k_x}}); \tag{B.2}$$

where $\mathbf{\Pi}_n = T_n(\hat{Z}_n\gamma_0 + X_n\beta_0)$, and the IV matrix Q_n can be taken as

$$\Big[\mathbf{\Pi}_n - \frac{\mu_{30}}{\xi^2\sigma_0^2}d_{T_n - I_n\,\text{tr}(T_n)/n} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\tilde{\mathbf{\Pi}}_n,\ \hat{Z}_{n,\cdot 1} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{\hat{Z}_{n,\cdot 1}}, \cdots, \hat{Z}_{n,\cdot k_z} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{\hat{Z}_{n,\cdot k_z}} \\ X_{n,\cdot 1} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{X_{n,\cdot 1}}, \cdots, X_{n,\cdot k_x} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\widetilde{X_{n,\cdot k_x}}\Big]. \tag{B.3}$$

When $\mu_{30} = 0$, the quadratic matrices $\text{diag}(\widetilde{\hat{Z}_{n,\cdot 1}})$, ..., $\text{diag}(\widetilde{\hat{Z}_{n,\cdot k_z}})$, $\text{diag}(\widetilde{X_{n,\cdot 1}})$, ..., $\text{diag}(\widetilde{X_{n,\cdot k_x}})$ in (B.2) are redundant.

Appendix C Proofs

Proof of Lemma 1. The discussion in the main text shows that, if $\lim_{n\to\infty}\frac{1}{n}Q_n'R_n(\rho)[T_n(\bar{Z}_n\gamma_0 + X_n\beta_0), \bar{Z}_n, X_n]$ has full column rank for any ρ in its parameter space $\boldsymbol{\rho}$, the linear moment part of $\lim_{n\to\infty}\mathrm{E}[g_n(\theta)] = 0$ implies that $(\lambda, \gamma, \beta) = (\lambda_0, \gamma_0, \beta_0)$. With $(\lambda, \gamma, \beta) = (\lambda_0, \gamma_0, \beta_0)$, by (6), $\mathrm{E}[\epsilon_n'(\theta)P_{jn}\epsilon_n(\theta)]$ reduces to $\mathrm{E}[\epsilon_n'R_n'^{-1}R_n'(\rho)P_{jn}R_n(\rho)R_n^{-1}\epsilon_n]$. Then a sufficient identification condition for ρ_0 is that[9]

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{E}[\epsilon_n'R_n'^{-1}R_n'(\rho)P_{jn}R_n(\rho)R_n^{-1}\epsilon_n] = 0 \text{ for } j = 1, \ldots, k_p \text{ have a unique solution at } \rho = \rho_0 \text{ for } \rho \in \boldsymbol{\rho}. \tag{C.1}$$

This identification condition corresponds to that of a pure SAR process $u_n = \rho M_n u_n + \epsilon_n$ as if u_n were observable, which is the same as that in Liu et al. (2010). Let $H_n = M_n R_n^{-1}$, and $\Xi = [\Xi_{jk}]$ be a $k_p \times 2$ matrix with $\Xi_{j1} = \lim_{n\to\infty}\frac{1}{n}\mathrm{E}(\epsilon_n'P_{jn}^sH_n\epsilon_n)$ and $\Xi_{j2} = \lim_{n\to\infty}\frac{1}{n}\mathrm{E}(\epsilon_n'H_n'P_{jn}^sH_n\epsilon_n)$ for $j = 1, \ldots, k_p$, where $A^s = A + A'$ for any square matrix A. Since $\mathrm{E}(\epsilon_n'P_{jn}\epsilon_n) = 0$ and $R_n(\rho) = R_n + (\rho_0 - \rho)M_n$ is linear in ρ, $\lim_{n\to\infty}\mathrm{E}[\epsilon_n'R_n'^{-1}R_n'(\rho)P_{jn}R_n(\rho)R_n^{-1}\epsilon_n] = 0$ for $j = 1, \ldots, k_p$ can be written as $(\rho_0 - \rho)\Xi\binom{1}{(\rho_0-\rho)/2} = 0$. Let Ξ_1 and Ξ_2 be, respectively, the first column and the second column of Ξ. Then (C.1) is equivalent to the condition that

$$\Xi_1 + \frac{\rho_0 - \rho}{2}\Xi_2 \neq 0 \text{ when } \rho \in \boldsymbol{\rho} \text{ and } \rho \neq \rho_0. \tag{C.2}$$

If Ξ has full column rank, then (C.2) holds. Even if Ξ has reduced column rank, the linear combination $\Xi_1 + \frac{\rho_0-\rho}{2}\Xi_2$ may still be nonzero for any $\rho \in \boldsymbol{\rho}$ and $\rho \neq \rho_0$, since the combination $\Xi_1 + \frac{\rho_0-\rho}{2}\Xi_2$ may be only zero for some $\rho \notin \boldsymbol{\rho}$.[10] We may derive some sufficient conditions for (C.2) to hold in the case that Ξ has reduced column rank. If $\Xi_1 = 0$, then $\Xi_2 \neq 0$ is sufficient; if $\Xi_2 = 0$, then $\Xi_1 \neq 0$ is sufficient. Thus, the following condition is sufficient for (C.2):

$$\text{Either } (i)\, \Xi \text{ has full column rank; or } (ii)\, \Xi_1 = 0, \text{ and } \Xi_{k2} \neq 0 \text{ for some } k; \text{ or } (iii)\, \Xi_{j1} \neq 0 \text{ for some } j, \text{ and } \Xi_2 = 0. \tag{C.3}$$

If $\lim_{n\to\infty}\frac{1}{n}Q_n'R_n(\rho)[T_n(\bar{Z}_n\gamma_0 + X_n\beta_0), \bar{Z}_n, X_n]$ has reduced column rank for some $\rho \in \boldsymbol{\rho}$, as Q_n usually includes X_n, we maintain in Assumption 6(ii) that $\lim_{n\to\infty}\frac{1}{n}Q_n'R_n(\rho)X_n$ has full column rank for any $\rho \in \boldsymbol{\rho}$. Assume that $\lim_{n\to\infty}\frac{1}{n}Q_n'R_n(\rho)[T_n(\bar{Z}_n\gamma_0 + X_n\beta_0), \bar{Z}_n, X_n]$ has column rank $(k_x + k_0)$ for some $0 \leq k_0 < k_z + 1$. Let $\bar{Z}_n = [\bar{Z}_{1n}, \bar{Z}_{2n}]$, where $\bar{Z}_{1n}$ is $n \times (k_z - k_0)$ and $\bar{Z}_{2n}$ is $n \times k_0$. Without loss of generality, assume that $\lim_{n\to\infty}\frac{1}{n}Q_n'R_n(\rho)[\bar{Z}_{2n}, X_n]$ has full column rank for any

[9] For an $n \times n$ matrix A_n, note that $\mathrm{E}(\epsilon_n'A_n\epsilon_n) = \sigma_0^2\,\mathrm{tr}(A_n)$ if ϵ_{ni}'s are homoskedastic, and $\mathrm{E}(\epsilon_n'A_n\epsilon_n) = \mathrm{tr}(A_n\Sigma_n)$ if ϵ_{ni}'s are heteroskedastic. We keep the expectation in (C.1) for simplicity.

[10] We thank an anonymous referee for pointing out this.

$\rho \in \boldsymbol{\rho}$.[11] Then, for a large enough n, there is some $(k_x+k_0)\times 1$ vector c_1 and some $(k_x+k_0)\times(k_z-k_0)$ matrix c_2 such that $T_n(\bar{Z}_n\gamma_0 + X_n\beta_0) = [\bar{Z}_{2n}, X_n]c_1$ and $\bar{Z}_{1n} = [\bar{Z}_{2n}, X_n]c_2$. Denote $\gamma = [\gamma_1', \gamma_2']'$, where γ_1 is $(k_z - k_0)\times 1$ and γ_2 is $k_0 \times 1$. Correspondingly, denote $\gamma_0 = [\gamma_{10}', \gamma_{20}']'$. Thus,

$$\mathrm{E}[Q_n'\epsilon_n(\theta)] = Q_n'R_n(\rho)[\bar{Z}_{2n}, X_n]\big[(\lambda_0-\lambda)c_1 + c_2(\gamma_{10}-\gamma_1) + \big(\begin{smallmatrix}\gamma_{20}-\gamma_2\\ \beta_0-\beta\end{smallmatrix}\big)\big].$$

Hence, $\lim_{n\to\infty}\frac{1}{n}\mathrm{E}[Q_n'\epsilon_n(\theta)] = 0$ implies that $(\lambda_0-\lambda)c_1 + c_2(\gamma_{10}-\gamma_1) + \big(\begin{smallmatrix}\gamma_{20}-\gamma_2\\ \beta_0-\beta\end{smallmatrix}\big) = 0$, and thus $\bar{\epsilon}_n(\theta) = 0$. Then, as long as λ_0 and γ_{10} are identified, γ_{20} and β_0 are identified. The identification of λ_0 and γ_{10} can be from the quadratic moments. As $\bar{\epsilon}_n(\theta) = 0$ for a large enough n, $\mathrm{E}[\epsilon_n'(\theta)P_{jn}\epsilon_n(\theta)] = \mathrm{E}[\check{\epsilon}_n'(\theta)P_{jn}\check{\epsilon}_n(\theta)]$. Therefore, we have the following sufficient identification condition:

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\mathrm{E}[\check{\epsilon}_n'(\theta)P_{jn}\check{\epsilon}_n(\theta)] = 0 \text{ for } j=1,\ldots,k_p, \text{ have a unique solution at}\\ &\qquad(\lambda,\rho,\gamma) = (\lambda_0,\rho_0,\gamma_0) \text{ for } \theta\in\Theta.\end{aligned} \tag{C.4}$$

Since $\check{\epsilon}_n(\theta)$ is linear in each element of θ, we can expand each $\mathrm{E}[\check{\epsilon}_n'(\theta)P_{jn}\check{\epsilon}_n(\theta)]$ as a polynomial function of $\theta-\theta_0$. Correspondingly, (C.4) can be written in an equivalent way where each element is a polynomial of $\theta-\theta_0$. ☐

Proof of Proposition 1. We first prove that $g_n(\theta) - \mathrm{E}[g_n(\theta)] = o_p(1)$. With $\check{\epsilon}_n(\theta)$ in (8), $\frac{1}{n}Q_n'\epsilon_n(\theta) - \frac{1}{n}\mathrm{E}[Q_n'\epsilon_n(\theta)] = \frac{1}{n}Q_n'\check{\epsilon}_n(\theta) = \frac{1}{n}Q_n'R_n(\rho)[R_n^{-1}\epsilon_n + (\lambda_0-\lambda)T_nR_n^{-1}\epsilon_n] + \frac{1}{n}Q_n'R_n(\rho)[(\lambda_0-\lambda)T_n\check{Z}_n\gamma_0 + \check{Z}_n(\gamma_0-\gamma)]$. For simplicity, we abbreviate "bounded in both row and column sum matrix norms" as UB. By Lemma A.4 in Lin and Lee (2010), under Assumptions 1 and 4, $\frac{1}{n}Q_n'K_n\epsilon_n = o_p(1)$, where K_n is an $n\times n$ UB matrix. Thus, with $R_n(\rho) = I_n - \rho M_n$, we have $\frac{1}{n}Q_n'R_n(\rho)[R_n^{-1}\epsilon_n + (\lambda_0-\lambda)T_nR_n^{-1}\epsilon_n] = o_p(1)$ under Assumptions 1, 2 and 4. By Assumption 8, $\frac{1}{n}Q_n'A_n\check{Z}_n = o_p(1)$ for $A_n = I_n$, M_n, T_n and M_nT_n. Thus, $\frac{1}{n}Q_n'R_n(\rho)[(\lambda_0-\lambda)T_n\check{Z}_n\gamma_0 + \check{Z}_n(\gamma_0-\gamma)] = o_p(1)$. Hence, $\frac{1}{n}Q_n'\epsilon_n(\theta) - \frac{1}{n}\mathrm{E}[Q_n'\epsilon_n(\theta)] = o_p(1)$, where $\frac{1}{n}\mathrm{E}[Q_n'\epsilon_n(\theta)] = \frac{1}{n}Q_n'\bar{\epsilon}_n(\theta) = O(1)$ under Assumptions 2–4. With $\bar{\epsilon}_n(\theta)$ in (7) and $\check{\epsilon}_n(\theta)$ in (8),

$$\frac{1}{n}\mathrm{E}[\epsilon_n'(\theta)P_{jn}\epsilon_n(\theta)] = \frac{1}{n}\bar{\epsilon}_n'(\theta)P_{jn}\bar{\epsilon}_n(\theta) + \frac{1}{2n}\mathrm{E}[\check{\epsilon}_n'(\theta)P_{jn}^s\check{\epsilon}_n(\theta)], \tag{C.5}$$

$$\begin{aligned}\frac{1}{n}\epsilon_n'(\theta)P_{jn}\epsilon_n(\theta) - \frac{1}{n}\mathrm{E}[\epsilon_n'(\theta)P_{jn}\epsilon_n(\theta)] &= \frac{1}{n}\bar{\epsilon}_n'(\theta)P_{jn}^s\check{\epsilon}_n(\theta)\\ &\quad + \frac{1}{2n}\{\check{\epsilon}_n'(\theta)P_{jn}^s\check{\epsilon}_n(\theta) - \mathrm{E}[\check{\epsilon}_n'(\theta)P_{jn}^s\check{\epsilon}_n(\theta)]\},\end{aligned} \tag{C.6}$$

where $\frac{1}{n}\bar{\epsilon}_n'(\theta)P_{jn}\bar{\epsilon}_n(\theta) = O(1)$, and

$$\frac{1}{n}\bar{\epsilon}_n'(\theta)P_{jn}^s\check{\epsilon}_n(\theta) = \frac{1}{n}[(\lambda_0-\lambda)\gamma_0', (\lambda_0-\lambda)\beta_0', (\gamma_0-\gamma)', (\beta_0-\beta)']\Upsilon_{1n}'P_{jn}^s\check{\epsilon}_n(\theta)$$

[11] We may permute the columns of $[T_n(\bar{Z}_n\gamma_0 + X_n\beta_0), \bar{Z}_n]$ to derive a new matrix A_n, and assume that the submatrix formed by the last (k_x+k_0) columns of $\lim_{n\to\infty}\frac{1}{n}Q_n'R_n(\rho)[A_n, X_n]$ has full column rank for any ρ in its parameter space. Then the argument below is similar.

$$- \frac{1}{n}\rho[(\lambda_0-\lambda)\gamma_0', (\lambda_0-\lambda)\beta_0', (\gamma_0-\gamma)', (\beta_0-\beta)']\Upsilon_{1n}' M_n' P_{jn}^s \check{\epsilon}_n(\theta)$$

with $\Upsilon_{1n} = [T_n\bar{Z}_n, T_n X_n, \bar{Z}_n, X_n]$. By an argument similar to that for $\frac{1}{n}Q_n'\epsilon_n(\theta) - \frac{1}{n}\mathrm{E}[Q_n'\epsilon_n(\theta)] = o_p(1)$, $\frac{1}{n}\bar{\epsilon}_n'(\theta)P_{jn}^s\check{\epsilon}_n(\theta) = o_p(1)$. By (8),

$$\begin{aligned}\check{\epsilon}_n(\theta) &= [R_n + (\rho_0-\rho)M_n][R_n^{-1}\epsilon_n + (\lambda_0-\lambda)\zeta_n + \check{Z}_n(\gamma_0-\gamma)] \\ &= \epsilon_n + (\rho_0-\rho)H_n\epsilon_n + (\lambda_0-\lambda)R_n\zeta_n + (\rho_0-\rho)(\lambda_0-\lambda)M_n\zeta_n \\ &\quad + R_n\check{Z}_n(\gamma_0-\gamma) + M_n\check{Z}_n(\gamma_0-\gamma)(\rho_0-\rho).\end{aligned} \tag{C.7}$$

where $\zeta_n = T_n(\check{Z}_n\gamma_0 + R_n^{-1}\epsilon_n)$ and $H_n = M_nR_n^{-1}$. Then we may expand $\frac{1}{n}\check{\epsilon}_n'(\theta)P_{jn}^s\check{\epsilon}_n(\theta)$ as a multivariate polynomial of θ. Under Assumption 1, for an $n\times n$ UB matrix K_n, $\frac{1}{n}\mathrm{E}(\epsilon_n'K_n\epsilon_n) = O(1)$ and $\frac{1}{n}\epsilon_n'K_n\epsilon_n - \frac{1}{n}\mathrm{E}(\epsilon_n'K_n\epsilon_n) = o_p(1)$ by Lemma A.3 in Lin and Lee (2010). Let A_n and B_n be either I_n, M_n, T_n or M_nT_n; and C_n be either I_n, H_n, $T_nR_n^{-1}$ or $M_nT_nR_n^{-1}$. Under Assumption 8, terms with the forms $\frac{1}{n}\check{Z}_n'B_n'P_{jn}^sA_n\check{Z}_n$ and $\frac{1}{n}\epsilon_n'C_n'P_{jn}^sA_n\check{Z}_n$ in the expression of $\frac{1}{n}\check{\epsilon}_n'(\theta)P_{jn}^s\check{\epsilon}_n(\theta)$ satisfy $\frac{1}{n}\mathrm{E}(\check{Z}_n'B_n'P_{jn}^sA_n\check{Z}_n) = O(1)$, $\frac{1}{n}\mathrm{E}(\epsilon_n'C_n'P_{jn}^sA_n\check{Z}_n) = O(1)$, $\frac{1}{n}\check{Z}_n'B_n'P_{jn}^sA_n\check{Z}_n - \frac{1}{n}\mathrm{E}(\check{Z}_n'B_n'P_{jn}^sA_n\check{Z}_n) = o_p(1)$ and $\frac{1}{n}\epsilon_n'C_n'P_{jn}^sA_n\check{Z}_n - \frac{1}{n}\mathrm{E}(\epsilon_n'C_n'P_{jn}^sA_n\check{Z}_n) = o_p(1)$. Hence, as $\zeta_n = T_n(\check{Z}_n\gamma_0 + R_n^{-1}\epsilon_n)$, by (C.7), $\frac{1}{n}\mathrm{E}[\check{\epsilon}_n'(\theta)P_{jn}^s\check{\epsilon}_n(\theta)] = O(1)$ and $\frac{1}{n}\check{\epsilon}_n'(\theta)P_{jn}^s\check{\epsilon}_n(\theta) - \frac{1}{n}\mathrm{E}[\check{\epsilon}_n'(\theta)P_{jn}^s\check{\epsilon}_n(\theta)] = o_p(1)$. Therefore, by (C.5)–(C.6), $\frac{1}{n}\mathrm{E}[\epsilon_n'(\theta)P_{jn}\epsilon_n(\theta)] = O(1)$ and $\frac{1}{n}\epsilon_n'(\theta)P_{jn}\epsilon_n(\theta) - \frac{1}{n}\mathrm{E}[\epsilon_n'(\theta)P_{jn}\epsilon_n(\theta)] = o_p(1)$. It follows that $\mathrm{E}[g_n(\theta)] = O(1)$ and $g_n(\theta) - \mathrm{E}[g_n(\theta)] = o_p(1)$. Note that $\epsilon_n(\theta)$ is linear in each element of θ and $g_n(\theta)$ is quadratic in $\epsilon_n(\theta)$. Then, as the parameter space Θ of θ is compact, we have the uniform convergence $\sup_{\theta\in\Theta}\|g_n(\theta) - \mathrm{E}[g_n(\theta)]\| = o_p(1)$. It follows that $\sup_{\theta\in\Theta}\|g_n'(\theta)a_n'a_ng_n(\theta) - \mathrm{E}[g_n'(\theta)]a_n'a_n\mathrm{E}[g_n(\theta)]\| = o_p(1)$.

The identification condition for $\lim_{n\to\infty} a_n\mathrm{E}[g_n(\theta)]$ to be uniquely zero at $\theta = \theta_0$ is maintained in Assumption 7. As each element of $g_n(\theta)$ is a polynomial function of θ, under Assumption 8, $\mathrm{E}[g_n(\theta)]$ is uniformly equicontinuous. So is $\mathrm{E}[g_n'(\theta)]a_n'a_n\mathrm{E}[g_n(\theta)]$. Hence, the identification uniqueness condition holds. With a compact parameter space Θ, the consistency of $\hat{\theta}$ follows from the uniform convergence that $\sup_{\theta\in\Theta}\|g_n'(\theta)a_n'a_ng_n(\theta) - \mathrm{E}[g_n'(\theta)]a_n'a_n\mathrm{E}[g_n(\theta)]\| = o_p(1)$ and the identification uniqueness condition (White, 1994). □

Proof of Proposition 2. The first order condition of $\hat{\theta}_{\mathrm{GMM}}$ is $D_n'(\hat{\theta}_{\mathrm{GMM}})a_n'a_ng_n(\hat{\theta}_{\mathrm{GMM}}) = 0$, where $D_n(\theta) = \frac{\partial g_n(\theta)}{\partial\theta'}$. By the mean value theorem, $0 = D_n'(\hat{\theta}_{\mathrm{GMM}})a_n'a_n[g_n(\theta_0) + D_n(\bar{\theta})(\hat{\theta}_{\mathrm{GMM}} - \theta_0)]$, where $\bar{\theta}$ lies between $\hat{\theta}_{\mathrm{GMM}}$ and θ_0. Thus,

$$\sqrt{n}(\hat{\theta}_{\mathrm{GMM}} - \theta_0) = -[D_n'(\hat{\theta}_{\mathrm{GMM}})a_n'a_nD_n(\bar{\theta})]^{-1}D_n'(\hat{\theta}_{\mathrm{GMM}})a_n'a_n\sqrt{n}g_n(\theta_0). \tag{C.8}$$

As each element of $g_n(\theta)$ is a polynomial function of θ, so is each element of $D_n(\theta)$. In addition, every coefficient for the polynomial functions of $D_n(\theta)$ is $O_p(1)$, and is $o_p(1)$ if its mean is deducted from it, by the proof of Proposition 1. Since $\hat{\theta}_{\mathrm{GMM}} = \theta_0 + o_p(1)$ by Proposition 1, $D_n(\bar{\theta}) =$

$D_n(\theta_0) + o_p(1) = G_n + o_p(1)$, where $G_n = \mathrm{E}[D_n(\theta_0)]$. Since $\frac{\partial \epsilon_n(\theta)}{\partial \theta'} = -[R_n(\rho)W_nY_n, M_n(S_n(\lambda)Y_n - Z_n\gamma - X_n\beta), R_n(\rho)Z_n, R_n(\rho)X_n]$ and $Y_n = S_n^{-1}(Z_n\gamma_0 + X_n\beta_0 + R_n^{-1}\epsilon_n)$, we have

$$\mathrm{E}\Big(\frac{\partial \epsilon_n'(\theta_0)P_{jn}\epsilon_n(\theta_0)}{\partial \theta'}\Big) = -\,\mathrm{E}[\epsilon_n'P_{jn}^sR_n\zeta_n, \epsilon_n'P_{jn}^sH_n\epsilon_n, \epsilon_n'P_{jn}^sR_n\check{Z}_n, 0_{1\times k_x}], \tag{C.9}$$

where $\zeta_n = T_n(\check{Z}_n\gamma_0 + R_n^{-1}\epsilon_n)$, and $\mathrm{E}(Q_n'\frac{\partial \epsilon_n(\theta_0)}{\partial \theta'}) = -Q_n'[R_nT_n(\bar{Z}_n\gamma_0 + X_n\beta_0), 0_{n\times 1}, R_n\bar{Z}_n, R_nX_n]$. Thus, G_n has the expression in (11). We next prove that $\lim_{n\to\infty} G_n$ has full column rank under Assumption 10. Let α_1 and α_2 be scalars, α_3 be a $k_z \times 1$ vector, and α_4 be a $k_x \times 1$ vector. In the case of Assumption 6(i), the last k_q elements of $\lim_{n\to\infty} G_n[\alpha_1, \alpha_2, \alpha_3', \alpha_4']' = 0$ are $-\lim_{n\to\infty} \frac{1}{n}Q_n'R_n[T_n(\bar{Z}_n\gamma_0 + X_n\beta_0), \bar{Z}_n, X_n][\alpha_1, \alpha_3', \alpha_4']' = 0$, which implies that $(\alpha_1, \alpha_3, \alpha_4) = (0, 0, 0)$, under the assumption that $\lim_{n\to\infty} \frac{1}{n}Q_n'R_n[T_n(\bar{Z}_n\gamma_0 + X_n\beta_0), \bar{Z}_n, X_n]$ has full column rank in Assumption 6(i). Then the first k_p elements of $\lim_{n\to\infty} G_n[\alpha_1, \alpha_2, \alpha_3', \alpha_4']' = 0$ become

$$-\lim_{n\to\infty} \frac{1}{n}[\mathrm{E}(\epsilon_n'P_{1n}^sH_n\epsilon_n), \dots, \mathrm{E}(\epsilon_n'P_{k_pn}^sH_n\epsilon_n)]'\alpha_2 = 0,$$

which implies that $\alpha_2 = 0$ if, for some j, $\lim_{n\to\infty} \frac{1}{n}\mathrm{E}(\epsilon_n'P_{jn}^sH_n\epsilon_n) \neq 0$. In the case of Assumption 6(ii), the first k_p elements of $\lim_{n\to\infty} G_n[\alpha_1, \alpha_2, \alpha_3', \alpha_4']' = 0$ are $\lim_{n\to\infty} G_{1n}[\alpha_1, \alpha_2, \alpha_3']' = 0$, where G_{1n} is in (12). As $\lim_{n\to\infty} G_{1n}$ has full column rank by Assumption 10, $(\alpha_1, \alpha_2, \alpha_3) = (0, 0, 0)$. Then the last k_q elements of $\lim_{n\to\infty} G_n[\alpha_1, \alpha_2, \alpha_3', \alpha_4']' = 0$ become $-\lim_{n\to\infty} \frac{1}{n}Q_n'R_nX_n\alpha_4 = 0$, which implies that $\alpha_4 = 0$ as $\lim_{n\to\infty} \frac{1}{n}Q_n'R_nX_n$ has full column rank under Assumption 6(ii). Hence, $\lim_{n\to\infty} G_n$ has full column rank under Assumption 10. As $\lim_{n\to\infty} a_nG_n$ has full column rank, $\lim_{n\to\infty} G_n'a_n'a_nG_n$ is invertible. It follows by (C.8) that $\sqrt{n}(\hat{\theta}_{\text{GMM}} - \theta_0) = -(G_n'a_n'a_nG_n)^{-1}G_n'a_n'a_n\sqrt{n}g_n(\theta_0) + o_p(1)$. By the central limit theorem for linear and quadratic forms in Kelejian and Prucha (2001), $\sqrt{n}g_n(\theta_0) \xrightarrow{d} N(0, \Omega_n)$, where $\Omega_n = n\,\mathrm{E}[g_n(\theta_0)g_n'(\theta_0)]$. Therefore, the asymptotic distribution in the proposition follows. □

Proof of Proposition 3. With $\hat{\Omega}_n = \Omega_n + o_p(1)$, the proof is similar to that for the OGMM estimator of SAR models with no endogenous regressors in Proposition 2 of Lee (2007), thus we omit the proof. □

Proof of Proposition 4. This proof follows the analytical approach in Jin et al. (2020) for the derivation of best linear and quadratic moments for spatial econometric models. The Ω_n in (13) can be written as

$$\Omega_n = \frac{1}{n}\Delta_n'\Delta_n, \tag{C.10}$$

where

$$\Delta_n = \begin{pmatrix} \frac{\sigma_0^2}{\sqrt{2}}\omega_{n\xi} & 0 \\ \frac{\mu_{30}}{\sigma_0}\omega_{nd} & \sigma_0Q_n \end{pmatrix}. \tag{C.11}$$

From (C.9),

$$
\begin{aligned}
&\mathrm{E}\Big(\frac{\partial \epsilon_n'(\theta_0) P_{jn} \epsilon_n(\theta_0)}{\partial \theta'}\Big) \\
&= -[\mathrm{tr}(P_{jn}^s R_n \mathrm{E}(\zeta_n \epsilon_n')), \sigma_0^2 \mathrm{tr}(P_{jn}^s H_n), \mathrm{tr}(P_{jn}^s R_n \mathrm{E}(\check{Z}_{n,\cdot 1}\epsilon_n')), \ldots, \mathrm{tr}(P_{jn}^s R_n \mathrm{E}(\check{Z}_{n,\cdot k_z}\epsilon_n')), 0_{1\times k_x}] \\
&= -\frac{1}{2}[\mathrm{tr}(P_{jn}^s C_{1n}^s), \cdots, \mathrm{tr}(P_{jn}^s C_{k_z+2,n}^s), 0_{1\times k_x}],
\end{aligned}
\tag{C.12}
$$

where $C_{1n} = R_n \mathrm{E}(\zeta_n \epsilon_n') - I_n \mathrm{tr}[R_n \mathrm{E}(\zeta_n \epsilon_n')]/n$, $C_{2n} = \sigma_0^2[H_n - I_n \mathrm{tr}(H_n)/n]$,[12] $C_{j+2,n} = R_n \mathrm{E}(\check{Z}_{n,\cdot j}\epsilon_n') - I_n \mathrm{tr}[R_n \mathrm{E}(\check{Z}_{n,\cdot j}\epsilon_n')]/n$ for $j = 1, \ldots, k_z$, and the second equality in (C.12) holds because P_{jn}^s has zero trace so that $\mathrm{tr}(P_{jn}^s C_n) = \frac{1}{2}\mathrm{tr}(P_{jn}^s C_n^s) = \frac{1}{2}\mathrm{tr}[P_{jn}^s(C_n^s - I_n \mathrm{tr}(C_n^s)/n)]$ for any $n\times n$ matrix C_n. For any two $n \times n$ matrices A_n and B_n, and constants a and b,

$$
\begin{aligned}
&\mathrm{tr}\big\{\big[a\,\mathrm{diag}(A_n) + \big(A_n - \mathrm{diag}(A_n)\big)\big]\big[b\,\mathrm{diag}(B_n) + \big(B_n - \mathrm{diag}(B_n)\big)\big]\big\} \\
&= \mathrm{tr}\big\{ab\,\mathrm{diag}(A_n)\,\mathrm{diag}(B_n) + \big(A_n - \mathrm{diag}(A_n)\big)\big(B_n - \mathrm{diag}(B_n)\big)\big\}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\mathrm{tr}(P_{jn}^s C_{kn}^s) &= \mathrm{tr}\big\{\big[\mathrm{diag}(P_{jn}^s) + \big(P_{jn}^s - \mathrm{diag}(P_{jn}^s)\big)\big]\big[\mathrm{diag}(C_{kn}^s) + \big(B_{kn}^s - \mathrm{diag}(C_{kn}^s)\big)\big]\big\} \\
&= \mathrm{tr}\{\mathrm{diag}(P_{jn}^s)\,\mathrm{diag}(C_{kn}^s) + [P_{jn}^s - \mathrm{diag}(P_{jn}^s)][C_{kn}^s - \mathrm{diag}(C_{kn}^s)]\} \\
&= \mathrm{tr}(P_{jn,\xi}^s C_{kn,1/\xi}^s) \\
&= \mathrm{vec}'(P_{jn,\xi}^s)\,\mathrm{vec}(C_{kn,1/\xi}^s),
\end{aligned}
\tag{C.13}
$$

where $P_{jn,\xi}^s = \xi\,\mathrm{diag}(P_{jn}^s) + [P_{jn}^s - \mathrm{diag}(P_{jn}^s)]$. By (11), (C.12) and (C.13),

$$
G_n = -\frac{1}{n}\begin{pmatrix} \frac{1}{2}\omega_{n\xi}'[\mathrm{vec}(C_{1n,1/\xi}^s), \cdots, \mathrm{vec}(C_{k_z+2,n,1/\xi}^s), 0_{n^2\times k_x}] \\ Q_n' R_n[T_n(\bar{Z}_n\gamma_0 + X_n\beta_0), 0_{n\times 1}, \bar{Z}_n, X_n] \end{pmatrix}. \tag{C.14}
$$

For simplicity, let the jth column of $R_n[T_n(\bar{Z}_n\gamma_0 + X_n\beta_0), 0_{n\times 1}, \bar{Z}_n, X_n]$ be K_{jn}. Denote $\tilde{K}_{jn} = K_{jn} - \frac{1}{n} l_n l_n' K_{jn}$, which is a vector with the sum of its elements equal to zero. Note that, as P_{jn}'s have zero traces so that $d_{P_{jn}^s}' l_n = 0$, we have $\mathrm{vec}'(P_{jn,\xi}^s)\,\mathrm{vec}(\mathrm{diag}(\tilde{K}_{jn})) = \mathrm{tr}(P_{jn,\xi}^s\,\mathrm{diag}(\tilde{K}_{jn})) = \xi\,\mathrm{tr}(\mathrm{diag}(P_{jn}^s)\,\mathrm{diag}(\tilde{K}_{jn})) = \xi d_{P_{jn}^s}' \tilde{K}_{jn} = \xi d_{P_{jn}^s}' K_{jn} = 2\xi d_{P_{jn}}' K_{jn}$. Thus, by (C.11) and (C.14),

$$
G_n = -\frac{1}{n}\Delta_n'\Gamma_n, \tag{C.15}
$$

where

$$
\Gamma_n = \begin{pmatrix} \Gamma_{n,11} & -\frac{\mu_{30}}{\sqrt{2}\xi\sigma_0^4}[\mathrm{vec}(\mathrm{diag}(\tilde{K}_{k_z+3,n})), \cdots, \mathrm{vec}(\mathrm{diag}(\tilde{K}_{k_z+k_x+2,n}))] \\ \frac{1}{\sigma_0} R_n[T_n(\bar{Z}_n\gamma_0 + X_n\beta_0), 0_{n\times 1}, \bar{Z}_n] & \frac{1}{\sigma_0} R_n X_n \end{pmatrix} \tag{C.16}
$$

[12] When defining the best moments later, the constant σ_0^2 can be removed. The C_{2n} in this proof has an extra σ_0^2 compared with that in Proposition 4.

with $\Gamma_{n,11} = [\mathrm{vec}(\frac{1}{\sqrt{2}\sigma_0^2} C^s_{1n,1/\xi} - \frac{\mu_{30}}{\sqrt{2}\xi\sigma_0^4} \mathrm{diag}(\tilde{K}_{1n})), \cdots, \mathrm{vec}(\frac{1}{\sqrt{2}\sigma_0^2} C^s_{k_z+2,n,1/\xi} - \frac{\mu_{30}}{\sqrt{2}\xi\sigma_0^4} \mathrm{diag}(\tilde{K}_{k_z+2,n}))]$. Hence, by (C.10) and (C.15), $G_n'\Omega_n^{-1}G_n = \frac{1}{n}\Gamma_n'\Delta_n(\Delta_n'\Delta_n)^{-1}\Delta_n'\Gamma_n \le \frac{1}{n}\Gamma_n'\Gamma_n$ by the Cauchy-Schwarz inequality, and $G_n'\Omega_n^{-1}G_n = \frac{1}{n}\Gamma_n'\Gamma_n$ if each column of Γ_n lies in the column space of Δ_n. As Γ_n does not depend on Q_n and P_{jn}'s, $\lim_{n\to\infty}(\frac{1}{n}\Gamma_n'\Gamma_n)^{-1}$ is the lower bound for the asymptotic variances of OGMM estimators in Proposition 3.

We next investigate the selection of Q_n and P_{jn}'s so that the lower bound $\lim_{n\to\infty}(\frac{1}{n}\Gamma_n'\Gamma_n)^{-1}$ can be attained. Let α_1, ..., α_{k_p} be constants and α be a $k_q \times 1$ vector. By (C.11),

$$\Delta_n[\alpha_1, \cdots, \alpha_{k_p}, \alpha']' = \begin{pmatrix} \frac{\sigma_0^2}{\sqrt{2}} \mathrm{vec}(P^s_{n,\xi}) \\ \frac{\mu_{30}}{\sigma_0} d_{P_n} + \sigma_0 Q_n \alpha \end{pmatrix}, \tag{C.17}$$

where $P_n = \sum_{j=1}^{k_p} \alpha_j P_{jn}$. For $1 \le j \le k_z + 2$, letting (C.17) be equal to the jth column of Γ_n in (C.16) yields

$$\begin{pmatrix} \frac{\sigma_0^2}{\sqrt{2}} \mathrm{vec}(P^s_{n,\xi}) \\ \frac{\mu_{30}}{\sigma_0} d_{P_n} + \sigma_0 Q_n \alpha \end{pmatrix} = \begin{pmatrix} \mathrm{vec}(\frac{1}{\sqrt{2}\sigma_0^2} C^s_{jn,1/\xi} - \frac{\mu_{30}}{\sqrt{2}\xi\sigma_0^4} \mathrm{diag}(\tilde{K}_{jn})) \\ \frac{1}{\sigma_0} K_{jn} \end{pmatrix}. \tag{C.18}$$

This is possible when

$$\frac{\xi\sigma_0^2}{\sqrt{2}} \mathrm{diag}(P_n^s) = \frac{1}{\sqrt{2}\xi\sigma_0^2} \mathrm{diag}(C^s_{jn}) - \frac{\mu_{30}}{\sqrt{2}\xi\sigma_0^4} \mathrm{diag}(\tilde{K}_{jn}),$$
$$\frac{\sigma_0^2}{\sqrt{2}}[P_n^s - \mathrm{diag}(P_n^s)] = \frac{1}{\sqrt{2}\sigma_0^2}[C^s_{jn} - \mathrm{diag}(C^s_{jn})],$$
$$\frac{\mu_{30}}{\sigma_0} d_{P_n} + \sigma_0 Q_n \alpha = \frac{1}{\sigma_0} K_{jn}.$$

We may let $\alpha_1 = \cdots = \alpha_{j-1} = 0$, $\alpha_j = \frac{1}{\sigma_0^4}, \alpha_{j+1} = \cdots = \alpha_{k_p} = 0$, and take P^s_{jn} to be

$$P^{*s}_{jn} = [C^s_{jn} - \mathrm{diag}(C^s_{jn})] + \frac{1}{\xi^2} \mathrm{diag}(C^s_{jn}) - \frac{\mu_{30}}{\xi^2\sigma_0^2} \mathrm{diag}(\tilde{K}_{jn}).$$

The jth column of Q_n can be taken as $Q^*_{jn} = K_{jn} - \frac{\mu_{30}}{\xi^2\sigma_0^4} d_{C_{jn}} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\tilde{K}_{jn}$. Alternatively, we can use the square matrices in P^{*s}_{jn} separately, so that we have the square matrices $C^s_{jn} - \mathrm{diag}(C^s_{jn})$, $\mathrm{diag}(C^s_{jn})$ and $\mathrm{diag}(\tilde{K}_{jn})$, because P^{*s}_{jn} is a linear combination of these matrices. If we use the IVs in Q^*_{jn} separately, then we have the IVs K_{jn}, $d_{C_{jn}}$ and $\tilde{K}_{jn}$. As $\tilde{K}_{jn} = K_{jn} - \frac{1}{n} l_n l_n' K_{jn}$, we can use the IVs K_{jn}, $d_{C_{jn}}$ and l_n equivalently.

For $k_z + 3 \le j \le k_z + k_x + 2$, letting (C.17) be equal to the jth column of Γ_n in (C.16) yields

$$\begin{pmatrix} \frac{\sigma_0^2}{\sqrt{2}} \mathrm{vec}(P^s_{n,\xi}) \\ \frac{\mu_{30}}{\sigma_0} d_{P_n} + \sigma_0 Q_n \alpha \end{pmatrix} = \begin{pmatrix} -\frac{\mu_{30}}{\sqrt{2}\xi\sigma_0^4} \mathrm{vec}(\mathrm{diag}(\tilde{K}_{jn})) \\ \frac{1}{\sigma_0} K_{jn} \end{pmatrix}. \tag{C.19}$$

This is possible when

$$\frac{\xi\sigma_0^2}{\sqrt{2}} \mathrm{diag}(P_n^s) = -\frac{\mu_{30}}{\sqrt{2}\xi\sigma_0^4} \mathrm{diag}(\tilde{K}_{jn}),$$

$$\frac{\mu_{30}}{\sigma_0}d_{P_n} + \sigma_0 Q_n \alpha = \frac{1}{\sigma_0}K_{jn}.$$

We may let $\alpha_1 = \cdots = \alpha_{j-1} = 0$, $\alpha_j = -\frac{\mu_{30}}{2\xi^2\sigma_0^6}, \alpha_{j+1} = \cdots = \alpha_{k_p} = 0$, and take P_{jn}^s to be $P_{jn}^{*s} = 2\,\text{diag}(\tilde{K}_{jn})$. The jth column of Q_n can be taken as $Q_{jn}^* = K_{jn} + \frac{\mu_{30}^2}{2\xi^2\sigma_0^6}\tilde{K}_{jn}$. Alternatively, if we use the IVs in Q_{jn}^* separately, then we have the IVs K_{jn} and l_n.

As K_{jn} is the jth column of $R_n[T_n(\bar{Z}_n\gamma_0 + X_n\beta_0), 0_{n\times 1}, \bar{Z}_n, X_n]$, we have, in particular, $Q_{2n}^* = -\frac{\mu_{30}}{\xi^2\sigma_0^4}d_{C_{2n}}$, which can be taken as $d_{C_{2n}}$ equivalently, and is redundant when $\mu_{30} = 0$. In addition, when $\mu_{30} = 0$, the square matrices $\text{diag}(\tilde{K}_{jn})$ for quadratic moments, where $k_z+3 \le j \le k_z+k_x+2$, are redundant, since they are from (C.19). Therefore, the results in the proposition follow. □

Appendix D Proof of Proposition 1 under weaker assumptions

Let $\|A\|$ be the spectral norm of a square matrix A, i.e., the square root of the largest eigenvalue of $A'A$. We replace Assumptions 2 and 5 with the following two weaker assumptions respectively, and show that Proposition 1 still holds under the weaker assumptions.

Assumption D.1. *The W_n and M_n have zero diagonals, and $\{W_n\}$, $\{M_n\}$, $\{S_n^{-1}\}$ and $\{R_n^{-1}\}$ are bounded in the spectral norm.*

Assumption D.2. *Elements of Q_n are uniformly bounded constants, and $\{P_{jn}\}$ for $j = 1, \ldots, k_p$ are bounded in the spectral norm.*

We first prove the following lemma.

Lemma D.1. *Let $\{A_n\}$ be a sequence of $n\times n$ nonstochastic matrices such that $\sup_n \|A_n\| < \infty$, b_n be an $n\times 1$ vector of constants that are uniformly bounded, and $\epsilon_n = [\epsilon_{n1}, \cdots, \epsilon_{nn}]'$, where ϵ_{ni}'s are independent with mean zero. Then (i) $\frac{1}{n}b_n'A_n'\epsilon_n = o_p(1)$ and $\frac{1}{n}b_n'A_n\epsilon_n = o_p(1)$, if $\sup_{i,n}\text{E}(\epsilon_{ni}^2) < \infty$; (ii) $\frac{1}{n}\text{E}(\epsilon_n'A_n\epsilon_n) = O(1)$ if $\sup_{i,n}\text{E}(\epsilon_{ni}^2) < \infty$; and (iii) $\frac{1}{n}\epsilon_n'A_n\epsilon_n - \frac{1}{n}\text{E}(\epsilon_n'A_n\epsilon_n) = o_p(1)$ if $\sup_{i,n}\text{E}(\epsilon_{ni}^4) < \infty$.*

Proof. (i) Denote $\Sigma_n = \text{diag}(\sigma_{n1}^2, \cdots, \sigma_{nn}^2)$, where $\sigma_{ni}^2 = \text{E}(\epsilon_{ni}^2)$. Let $A_n'A_n = \Gamma_n\Lambda_n\Gamma_n'$ be the spectral decomposition of $A_n'A_n$, where $\Gamma_n\Gamma_n' = I_n$ and Λ_n is an $n\times n$ diagonal matrix of the eigenvalues of $A_n'A_n$. The variance of $\frac{1}{n}b_n'A_n'\epsilon_n$ satisfies $\text{var}(\frac{1}{n}b_n'A_n'\epsilon_n) = \frac{1}{n^2}b_n'A_n'\Sigma_nA_nb_n \le \frac{c}{n^2}b_n'A_n'A_nb_n = \frac{c}{n^2}b_n'\Gamma_n\Lambda_n\Gamma_n'b_n \le \frac{c}{n^2}b_n'b_n\lambda_{\max}(A_n'A_n) = O(\frac{1}{n})$, where c is a constant and $\lambda_{\max}(A_n'A_n)$ denotes the largest eigenvalue of $A_n'A_n$. Thus, $\frac{1}{n}b_n'A_n'\epsilon_n = o_p(1)$. Similarly, the variance of $\frac{1}{n}b_n'A_n\epsilon_n$ satisfies $\text{var}(\frac{1}{n}b_n'A_n\epsilon_n) \le \frac{c}{n^2}b_n'b_n\lambda_{\max}(A_nA_n') = \frac{c}{n^2}b_n'b_n\lambda_{\max}(A_n'A_n) = O(\frac{1}{n})$. Thus, $\frac{1}{n}b_n'A_n\epsilon_n = o_p(1)$.

(ii) Let $A_n = [a_{n,ij}]$. We have $|\frac{1}{n}\text{E}(\epsilon_n'A_n\epsilon_n)| = |\frac{1}{n}\text{tr}(A_n\Sigma_n)| = |\frac{1}{n}\sum_{i=1}^n a_{n,ii}\sigma_{ni}^2| \le \frac{c}{n}\sum_{i=1}^n |a_{n,ii}| \le c\sqrt{\frac{1}{n}\sum_{i=1}^n a_{n,ii}^2}$, where c is a constant and the last inequality follows by the Cauchy-Schwarz inequality. Then, $|\frac{1}{n}\,\mathrm{E}(\epsilon_n' A_n \epsilon_n)| \le c\sqrt{\frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n a_{n,ij}^2} = c\sqrt{\frac{1}{n}\,\mathrm{tr}(A_n' A_n)} \le c\|A_n\|$. Thus, $\frac{1}{n}\,\mathrm{E}(\epsilon_n' A_n \epsilon_n) = O(1)$.

(*iii*) By Lemma 2(3) in Jin and Lee (2012), the variance of $\epsilon_n' A_n \epsilon_n$ is $\mathrm{var}(\epsilon_n' A_n \epsilon_n) = \sum_{i=1}^n a_{n,ii}^2[\mathrm{E}(\epsilon_{ni}^4) - 3\sigma_{ni}^4] + \mathrm{tr}[\Sigma_n A_n \Sigma_n (A_n + A_n')]$. Thus,

$$\begin{aligned}
\mathrm{var}(\epsilon_n' A_n \epsilon_n) &\le c_1 \sum_{i=1}^n a_{n,ii}^2 + \frac{1}{2}\,\mathrm{tr}[\Sigma_n^{1/2}(A_n + A_n')\Sigma_n(A_n + A_n')\Sigma_n^{1/2}] \\
&\le nc_2 + c_3\,\mathrm{tr}[\Sigma_n^{1/2}(A_n + A_n')(A_n + A_n')\Sigma_n^{1/2}] \\
&= nc_2 + c_3\,\mathrm{tr}[(A_n + A_n')\Sigma_n(A_n + A_n')] \\
&\le nc_2 + c_4\,\mathrm{tr}[(A_n + A_n')(A_n + A_n')] \\
&\le nc_2 + nc_4\|A_n + A_n'\|^2 \\
&\le nc_2 + nc_4(\|A_n\| + \|A_n'\|)^2 \le nc_5,
\end{aligned}$$

where c_j's are constants, and the last inequality uses $\|A_n\| = \|A_n'\|$. Hence, $\mathrm{var}(\epsilon_n' A_n \epsilon_n) = O(n)$ and $\frac{1}{n}\epsilon_n' A_n \epsilon_n - \frac{1}{n}\,\mathrm{E}(\epsilon_n' A_n \epsilon_n) = o_p(1)$. □

As can be seen from the proof of Proposition 1 in Appendix C, replacing Assumptions 2 and 5 by Assumptions D.1–D.2 only affects the terms with the forms in the above lemma. Since the above lemma shows that the orders of those terms do not change, Proposition 1 still holds.

References

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

Debarsy, N., Jin, F., Lee, L.F., 2015. Large sample properties of the matrix exponential spatial specification with an application to FDI. Journal of Econometrics 188, 1–21.

Delgado, M.A., Robinson, P.M., 2015. Non-nested testing of spatial correlation. Journal of Econometrics 187, 385–401.

Drukker, D.M., Egger, P., Prucha, I.R., 2013. On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. Econometric Reviews 32, 686–733.

Fingleton, B., Le Gallo, J., 2008. Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties. Papers in Regional Science 87, 319–339.

Gupta, A., 2019. Estimation of spatial autoregressions with stochastic weight matrices. Econometric Theory 35, 417–463.

Gupta, A., Robinson, P.M., 2015. Inference on higher-order spatial autoregressive models with increasingly many parameters. Journal of Econometrics 186, 19–31.

Gupta, A., Robinson, P.M., 2018. Pseudo maximum likelihood estimation of spatial autoregressive models with increasing dimension. Journal of Econometrics 202, 92–107.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Jenish, N., Prucha, I.R., 2012. On spatial processes and asymptotic inference under near-epoch dependence. Journal of Econometrics 170, 178–190.

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., 2018. Irregular N2SLS and LASSO estimation of the matrix exponential spatial specification model. Journal of Econometrics 206, 336–358.

Jin, F., Lee, L.F., Yang, K., 2020. Best linear and quadratic moments of spatial econometric models. Working paper. https://feijin.weebly.com/uploads/2/2/8/9/22895144/spatial-best-moments-jin-lee-yang.pdf.

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., 2003. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econometric Reviews 22, 307–335.

Lee, L.F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1925.

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

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

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

Liu, S.F., Yang, Z., 2015. Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and nonnormality. Regional Science and Urban Economics 52, 50–70.

Liu, X., 2012. On the consistency of the LIML estimator of a spatial autoregressive model with many instruments. Economics Letters 116, 472–475.

Liu, X., Lee, L.F., 2013. Two stage least squares estimation of spatial autoregressive models with endogenous regressors and many instruments. Econometric Reviews 32, 734–753.

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

Liu, X., Saraiva, P., 2015. GMM estimation of SAR models with endogenous regressors. Regional Science and Urban Economics 55, 68–79.

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

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.

White, H., 1994. Estimation, Inference and Specification Analysis. New York: Cambridge University Press.