Best linear and quadratic moments for spatial econometric models and an application to spatial interdependence patterns of employment growth in US counties

Fei Jin[1], Lung-fei Lee[2], and Kai Yang*[3]

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

[2]School of Economics, Shanghai University of Finance and Economics, Shanghai 200433, China

[3]School of Economics, Shanghai University of Finance and Economics, and Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai 200433, China

October 11, 2023

Abstract

We provide a novel analytic procedure to construct best linear and quadratic moments of the generalized method of moments (GMM) estimation for a large class of cross-sectional network and spatial econometric models, which generate a best GMM estimator based on linear and quadratic moments that is asymptotically more efficient than the quasi maximum likelihood estimator when the disturbances are non-normal. We apply this procedure to a high order spatial autoregressive model with spatial errors, where the disturbances are heteroskedastic with an unknown distribution. Two tests of normality of disturbances are developed. We apply the model, estimators and normality tests to employment data in US counties, which demonstrates spatial interdependence patterns and channels of regional economic growth.

Keywords: Efficiency, best generalized method of moments, spatial autoregression, spatial simultaneous equations, regional employment growth

JEL classification: C13, C21, C3, R11

*Email addresses: jinfei@fudan.edu.cn (F. Jin), lee.1777@osu.edu (L.F. Lee), yang.kai@mail.shufe.edu.cn (K. Yang). Correspondence: Lung-fei Lee, 616 School of Economics, Shanghai University of Finance and Economics, Shanghai 200433, China. Fei Jin's work is supported by the National Natural Science Foundation of China under Grants No. 71973030 and No. 71833004. Kai Yang's work is supported by the National Natural Science Foundation of China under Grant No. 72273082.

1 Introduction

It has been more than forty years since Paelinck and Klaassen (1979) published the book entitled *Spatial Econometrics*, which outlined comprehensively the field of spatial econometrics and its distinct methodology. Through forty years' development, spatial econometrics has gained popularity in applied econometrics and social science. Many different model specifications and estimation methods are investigated. See, e.g., the survey papers of Anselin and Bera (1998), Anselin (2010) and Arbia (2016).

With the increasing richness and accessibility of network and spatial data, researchers pay more attention to heterogeneity among units and complex relationships between variables, which leads to more complicated models in recent papers. Among them, two types of models attract much attention theoretically and empirically. First, high order spatial autoregressive models capture heterogeneous network and spatial interdependence (Lee and Liu, 2010; Badinger and Egger, 2011, 2015; Gupta and Robinson, 2015, to mention a few). Second, various estimation methods for simultaneous equations spatial autoregressive (SESAR) models which intend to incorporate interdependence of agents' choices or outcomes across different activities on networks have been developed (Kelejian and Prucha, 2004; Yang and Lee, 2017; Liu and Saraiva, 2019, to mention a few).

Maximum likelihood (ML) or quasi-maximum likelihood (QML), generalized methods of moments (GMM), and two (or three) stage least square (2/3SLS) estimators have been developed for the above models. ML and QML estimators are considered to have advantages on estimation efficiency. However, researchers face two issues that complicate estimation procedures in empirical applications when using ML or QML. First, the parameter space of multiple network and spatial effect coefficients need to be clearly specified as the Jacobian involved in the likelihood function can be negative at some parameter values. This drawback becomes clear and more severe in estimating models such as the high order SAR model and SESAR model, as their parameter spaces for spatial effect coefficients are complicated. Second, the computational cost of ML and QML is usually large. Compared with ML and QML, GMM (and 2/3SLS) estimators do not need to specify parameter spaces when evaluating its objective function empirically and they are less burdened in computation. A potential disadvantage of the GMM estimator (GMME) is on efficiency as the selection of moments affects its asymptotic variance.

We propose the best GMME for a large class of cross-sectional network and spatial econometric models, where the moments for estimation can be linear and quadratic functions, since it has advantages on computation and can be more efficient than the QML estimator when the distribution of disturbances is not normal.[1] Therefore, there is an issue on the search for possible best linear and quadratic moments if they exist, so that the derived GMME can be the best within the class of GMMEs derived from linear and quadratic moments. In the literature, researchers derive the best moments for some models using trial-and-error methods: Liu et al. (2010) for the first order SAR model with SAR and homoskedastic disturbances (SARAR model); Lee and Liu (2010) for high order SARAR models with homoskedastic disturbances; Doğan and Taşpınar (2013) for the SAR model with SMA disturbances; Debarsy et al. (2015) for the matrix exponential spatial specification model. However, derivations of those best moments were subject to trial and error by utilizing a characterization of redundant moments in Breusch et al. (1999). There is no general approach so far existed in the econometrics literature so researchers need to search for best moments for each model individually. The trial-and-error approach might not be easy to be applied to other network and spatial econometric models as best moments might not be easily figured out, especially for models with combined features such as high order, simultaneous equations, spatial error and moving average errors, heteroskedasticity, etc. As an evidence, Lee (2007) and Liu and Saraiva (2015) only consider best linear and quadratic moments for the special case with normal disturbances, and for the SAR model without and with endogenous regressors, respectively.

In this paper, we provide a novel analytical procedure to derive best linear and quadratic

[1]When the disturbances are normally distributed, the best GMME is asymptotically as efficient as the ML estimator. The best GMME is defined as the optimal GMME with selected linear and quadratic moments, which can achieve the smallest asymptotic variance.

moments for the GMM estimation of a large class of cross-sectional spatial econometric models, which can have a spatial autoregressive (SAR) process in the dependent variable (Whittle, 1954; Cliff and Ord, 1973), SAR and/or spatial moving average (SMA) processes in the disturbances (Haining, 1978; Cliff and Ord, 1973), higher order and multivariate versions of the above processes, and possibly other processes such as spatial error components (Kelejian and Robinson, 1993). The models we consider also feature heteroskedasticity in generality, which provides a useful modeling approach for empirical studies.

Our analytic method of deriving best linear and quadratic moments is illustrated in detail for a high order SARAR model with heteroskedasticity. There are a few merits of analyzing this class of models. First, the high order model allows researchers to investigate a number of channels of interdependence by estimating parameters for each channel represented by different weight matrices (Baltagi et al., 2022), and may be considered as alternatives of a poorly specified spatial weight matrix (Anselin and Bera, 1998). Second, the SARAR model allows spatial errors, which captures important spatial interdependence due to unobserved explanatory variables (Cliff and Ord, 1973; Kelejian and Prucha, 1998, 2010, to name a few). Third, while ignoring heteroskedasticity usually leads to inconsistent ML, QML and GMM estimators, robust methods (e.g., Kelejian and Prucha, 2010; Lin and Lee, 2010; Liu and Yang, 2015; Taşpınar et al., 2019) can lead to substantial efficiency loss, as shown in our Monte Carlo study. We show that, with a proper decomposition of the variance matrix for a corresponding set of linear-quadratic moments and reformulation of the gradient matrix of a GMM objective function, best linear and quadratic moments can be derived analytically by applying the Cauchy-Schwarz inequality to derive a variance lower bound that can be attained.

Since non-normal model disturbances indicate the existence of a GMM estimator that is asymptotically more efficient than the QML estimator, we derive tests for the normality of disturbances in the SARAR model. These tests might also be of interest in other situations, e.g., normality of disturbances implies that the information matrix equality holds for the QML estimator, so the asymptotic variance can be simplified.

To facilitate estimation for various models in empirical research, we also employ the proposed analytic procedure to derive best linear and quadratic moments for a general new multivariate model: a high order SESAR model with multivariate spatial autoregressive and moving average (MSARMA) heteroskedastic disturbances.[2] This model nests many single equation and simultaneous equations models in the literature as special cases, such as those in Kelejian and Prucha (2004), Yang and Lee (2017), etc. In particular, while SAR disturbances capture global spatial dependence, SMA disturbances capture local spatial dependence (Fingleton, 2008; Doğan and Taşpınar, 2013, e.g.,). The SESAR model has been employed to applications on, e.g., regional science (de Graaff et al., 2012; Gebremariam et al., 2010), housing economics (Baltagi and Bresson, 2011; Jeanty et al., 2010), macroeconomics (Elhorst and Emili, 2022), and fiscal policy analysis (Hauptmeier et al., 2012; Allers and Elhorst, 2011). Therefore, the potential applications of the proposed method are wide in applied econometrics.

The spatial interdependence or propagation of local economic growth is documented in the literature (Wheeler, 2001; Gebremariam et al., 2010; Feyrer et al., 2017, etc). However, the patterns and channels of interdependence have not been addressed. Postulating a coefficient for spatial correlation and a decay function in geographical distance to form a SAR model cannot fully capture heterogeneity through different channels, and at the same time heteroskedasticity and spatial correlation in errors may occur in this issue. Thus the high order SARAR model with heteroskedasticity is a more suitable model. We use county-level census data in the contiguous US to investigate the degree to which employment growth disseminates through different channels between neighbors with geographic proximity, industrial proximity, and political tendency proximity. We decompose the interdependence into three channels and measure their heterogeneous magnitudes: the interdependence decays fast geographically; and spatial interdependence among neighbors with industrial proximity is the strongest among the three channels. The proposed best GMM generates efficient and precise estimates which unveil the above insights simultaneously.

[2]The analysis is in the supplementary file.

This paper is organized as follows. In Section 2, we present the high order SARAR model with possibly heteroskedastic disturbances, derive analytically best linear and quadratic moments, develop normality tests of the disturbances, and present some Monte Carlo results on the GMM estimator with best linear and quadratic moments. Section 3 applies the high order SARAR model and best GMM to county-level census data for the investigation of spatial interdependence of local economic growth. Section 4 concludes. Analysis on the general high order SESAR model, more Monte Carlo results, proofs, and detailed derivations are provided in the supplementary file.

2 The high order SARAR model

Consider the following high order SARAR model with heteroskedasticity:

$$Y_n = \sum_{j=1}^{l_w} \lambda_{j0} W_{jn} Y_n + X_n \beta_0 + u_n, \quad u_n = \sum_{k=1}^{l_m} \rho_{k0} M_{kn} u_n + \Sigma_{n0}^{1/2} V_n, \tag{1}$$

where n is the sample size, Y_n is an $n \times 1$ vector of observations on the dependent variable, W_{jn}'s and M_{kn}'s are $n \times n$ spatial weight matrices of known constants with zero diagonals such that $W_{j_1 n} \neq W_{j_2 n}$ and $M_{k_1 n} \neq M_{k_2 n}$ for $j_1, j_2 = 1, 2, \cdots, l_w$ and $k_1, k_2 = 1, 2, \cdots, l_m$. $X_n = [x_{n,ij}]$ is an $n \times l_x$ matrix of exogenous variables, $\Sigma_{n0}^{1/2} = \mathrm{diag}(\sigma_{n1,0}, \ldots, \sigma_{nn,0})$ is a diagonal standard deviation matrix of heteroskedasticity, formed by $\sigma_{ni,0} = f(x_{n,i1}, \cdots, x_{n,il_x}, \gamma_0)$ for some known function $f(\cdot)$,[3] v_i's in $V_n = [v_1, \cdots, v_n]'$ are i.i.d. with mean zero and unit variance, $\lambda_0 = [\lambda_{10}, \cdots, \lambda_{l_w 0}]'$ and $\rho_0 = [\rho_{10}, \cdots, \rho_{l_m 0}]'$ are spatial dependence parameters for the dependent variable and disturbances respectively, and β_0 is an $l_x \times 1$ vector of parameters. Model (1) embodies multiple spatial interdependence, which is especially useful when researchers want to investigate multiple possible channels of interdependence or allow for flexible rates of decay of interdependence (Lee and Liu, 2010; Gupta and Robinson, 2015; Baltagi et al., 2022, etc.). Furthermore, Model (1) considers spatial errors in a high order

[3] With an unknown form of heteroskedasticity, we have investigated the possibility of estimating best linear and quadratic moments nonparametrically, but found it unable to achieve the same efficiency as the infeasible best moments. The literature on best moments for spatial econometric models typically assumes homoskedasticity, while our analytical procedure allows us to handle heteroskedastic errors with quite flexible known functional forms.

form as additional sources of spatial dependence.

Denote $R_n(\rho) = I_n - \sum_{k=1}^{l_m} \rho_k M_{kn}$ and $S_n(\lambda) = I_n - \sum_{j=1}^{l_w} \lambda_j W_{jn}$, where ρ and λ are parameter values in their parameter spaces, and I_n is the n-dimensional identity matrix. Under the regularity condition that $R_n \equiv R_n(\rho_0)$ and $S_n \equiv S_n(\lambda_0)$ are invertible, Y_n has the reduced form $Y_n = S_n^{-1}(X_n\beta_0 + R_n^{-1}\Sigma_{n0}^{1/2}V_n)$.

2.1 GMM estimation and best moments

We consider the GMM estimation of model (1) with the following vector of linear and quadratic moments:[4]

$$g_n(\theta) = [V_n'(\theta)P_{1n}V_n(\theta) - \mathrm{tr}(P_{1n}), \cdots, V_n'(\theta)P_{l_pn}V_n(\theta) - \mathrm{tr}(P_{l_pn}), V_n'(\theta)Q_n]', \tag{2}$$

where $\theta = [\lambda', \rho', \beta', \gamma']'$, which is a finite dimensional parameter vector of the model, $V_n(\theta) = \Sigma_n^{-1/2}(\gamma)R_n(\rho)[S_n(\lambda)Y_n - X_n\beta]$ with $\Sigma_n^{1/2}(\gamma) = \mathrm{diag}(\sigma_{n1}(\gamma), \ldots, \sigma_{nn}(\gamma))$ for $\sigma_{ni}(\gamma) = f(x_{n,i1}, \cdots, x_{n,il_x}, \gamma)$, P_{rn}'s are $n \times n$ matrices constructed from W_{jn}'s, M_{kn}'s and X_n, and Q_n is an $n \times l_q$ instrumental variable (IV) matrix as a function of W_{jn}'s, M_{kn}'s and X_n. For example, P_{rn}'s can be I_n, W_{jn}, M_{kn}, W_{jn}^2, M_{kn}^2 and so on, and Q_n can be the matrix formed by the independent columns of X_n, $W_{jn}X_n$, $M_{kn}X_n$, $W_{jn}^2X_n$, $M_{kn}^2X_n$ and so on.[5] The moments linear and quadratic in $V_n(\theta)$ are motivated from the quasi maximum likelihood estimation of the SARAR model, as in Lee (2007). Our moments quadratic in $V_n(\theta)$ differ from those of Lee (2007) for homoskedastic SAR models in that P_{rn}'s are not required to have zero traces. Our approach simplifies the search for best moments since possible best P_{rn}'s do not need to have zero races either. Efficiency implications due to this difference are discussed in Appendix A. We consider the GMM estimation based on the objective function $g_n'(\theta)a_n'a_ng_n(\theta)$ for some weighting matrix $a_n'a_n$, where a_n is an $l_a \times (l_p + l_q)$ full row rank matrix with l_a greater than or equal to the number of parameters l_θ in θ and the limit

[4]The moment $V_n'(\theta)P_{jn}V_n(\theta) - \mathrm{tr}(P_{jn})$ is called a quadratic moment as it is quadratic in $V_n(\theta)$. Similarly, $Q_n'V_n(\theta)$ is called a linear moment.

[5]It is possible to use higher powers of spatial weight matrices to construct instruments. But there can be a many instruments problem (Liu and Lee, 2013) and potentially a weak instruments problem. One method of solving these problems is to use synthetic instruments as in Fingleton (2023). Our GMM estimator with best linear and quadratic moments not only solve these problems, but also is efficient among a class of GMM estimators with linear and quadratic moments.

$a = \lim_{n\to\infty} a_n$ exists by design.

The following assumption contains basic regularity conditions for model (1) and its GMM estimation. We abbreviate "bounded in both row and column sum norms" as UB.

Assumption 1. *(i) v_i's are i.i.d. with mean zero and unit variance, and $\mathrm{E}(|v_i|^{4+\eta}) < \infty$ for some $\eta > 0$; (ii) The elements of X_n are uniformly bounded constants, and $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ exists and is nonsingular. (iii) $\{W_{jn}\}$ for $j = 1, \ldots, l_w$ and $\{M_{kn}\}$ for $k = 1, \ldots, l_m$, which are UB, have known constant entries and zero diagonals; $\{R_n\}$ and $\{S_n\}$ are invertible and $\{R_n^{-1}\}$ and $\{S_n^{-1}\}$ are UB. (iv) $\{P_{rn}\}$ for $r = 1, \ldots, l_p$ are UB, and the elements of Q_n are uniformly bounded constants. (v) The parameter space Θ of θ is a compact subset of $\mathbb{R}^{l_\theta}$. (vi) $\sup_{i,n} |f(x_{n,i1}, \cdots, x_{n,il_x}, \gamma_0)| < \infty$, $\inf_{\gamma\in\Gamma,i,n} |f(x_{n,i1}, \cdots, x_{n,il_x}, \gamma)| > 0$, and $f(x_{n,i1}, \cdots, x_{n,il_x}, \gamma)$ is differentiable such that $\sup_{\gamma\in\Gamma,i,n} \|\frac{\partial}{\partial\gamma} f(x_{n,i1}, \cdots, x_{n,il_x}, \gamma)\| < \infty$, where Γ is the compact parameter space of γ.*

Assumptions 1(i)–(iv) are typical regularity conditions in the spatial econometric literature (see, e.g., Kelejian and Prucha, 1998, 1999; Lee, 2004, 2007). In Assumption 1(i), the finite moment condition with an order higher than four on the disturbances is needed for the applicability of the central limit theorem for linear and quadratic forms in Kelejian and Prucha (2001). The nonstochasticity of X_n in Assumption 1(ii) is maintained for simplicity.[6] The UB properties of spatial weight matrices in Assumption 1(iii) originate in Kelejian and Prucha (1998, 1999). As the analysis will use the reduced form of Y_n, $\{R_n^{-1}\}$ and $\{S_n^{-1}\}$ are also assumed to be UB. These UB properties are also required on the matrices P_{rn}'s, as they are functions of spatial weight matrices. Assumption 1(v) is a usual condition on the parameter space for an extremum estimation. In Assumption 1(vi), the function $f(\cdot)$ for heteroskedasticity is required to be uniformly bounded at the true γ_0, uniformly bounded away from zero and satisfy some smoothness conditions. As γ is of finite dimension, the norm $\|\cdot\|$ for the derivative of $f(\cdot)$ can be taken as the familiar Euclidean norm. This general model

[6] The asymptotic distribution of the GMM estimator with best linear-quadratic moments will use the central limit theorem for linear-quadratic forms in Kelejian and Prucha (2001), which require square matrices in linear-quadratic forms to be nonstochastic. The square matrices in our derived best quadratic moments will involve X_n. We leave the extension to a random X_n to future research.

nests the SARAR model with homoskedastic disturbances when $f(\cdot)$ is a constant function.

We have $\mathrm{E}[Q_n' V_n(\theta)] = Q_n' \bar{V}_n(\theta)$, and

$$\mathrm{E}[V_n'(\theta) P_{rn} V_n(\theta)] = \bar{V}_n'(\theta) P_{rn} \bar{V}_n(\theta) + \mathrm{tr}[\mathcal{A}_n'(\lambda,\rho,\gamma) P_{rn} \mathcal{A}_n(\lambda,\rho,\gamma)] \quad (3)$$

for $r = 1, \ldots, l_p$, where

$$\bar{V}_n(\theta) = \mathrm{E}[V_n(\theta)] = \Sigma_n^{-1/2}(\gamma) R_n(\rho) \Big[\sum_{j=1}^{l_w} (\lambda_{j0} - \lambda_j) W_{jn} S_n^{-1} X_n \beta_0 + X_n(\beta_0 - \beta) \Big], \quad (4)$$

and $\mathcal{A}_n(\lambda,\rho,\gamma) = \Sigma_n^{-1/2}(\gamma) R_n(\rho) S_n(\lambda) S_n^{-1} R_n^{-1} \Sigma_{n0}^{1/2}$. The following assumption provides sufficient parameter identification conditions for $\lim_{n\to\infty} \frac{1}{n}\mathrm{E}[g_n(\theta)]$ to be uniquely zero at $\theta = \theta_0$, which are similar to those in Lee and Liu (2010).

Assumption 2. *Either the following (i) or (ii) is satisfied:*

(i) $\lim_{n\to\infty} \frac{1}{n} Q_n' \Sigma_n^{-1/2}(\gamma) R_n(\rho) [W_{1n} S_n^{-1} X_n \beta_0, \cdots, W_{l_w n} S_n^{-1} X_n \beta_0, X_n]$ *has full column rank for any* (ρ,γ)*, and the equation system* $\lim_{n\to\infty}\{\frac{1}{n}\mathrm{tr}[\mathcal{A}_n'(\lambda_0,\rho,\gamma) P_{rn} \mathcal{A}_n(\lambda_0,\rho,\gamma)] - \frac{1}{n}\mathrm{tr}(P_{rn})\} = 0$ *for* $r = 1, \ldots, l_p$ *holds uniquely at* $(\rho,\gamma) = (\rho_0,\gamma_0)$*;*

(ii) *For any* (ρ,γ), $\lim_{n\to\infty} \frac{1}{n} Q_n' \Sigma_n^{-1/2}(\gamma) R_n(\rho) X_n$ *has full column rank,* $\lim_{n\to\infty} \frac{1}{n} Q_n' \Sigma_n^{-1/2}(\gamma) R_n(\rho) [W_{1n} S_n^{-1} X_n \beta_0, \cdots, W_{l_w n} S_n^{-1} X_n \beta_0, X_n]$ *has column rank* $l_x + l_w - l_{w0}$ *for some* $1 \leq l_{w0} \leq l_w$*, and the equation system* $\lim_{n\to\infty}\{\frac{1}{n}\mathrm{tr}[\mathcal{A}_n'(\lambda,\rho,\gamma) P_{rn} \mathcal{A}_n(\lambda,\rho,\gamma)] - \frac{1}{n}\mathrm{tr}(P_{rn})\} = 0$ *for* $r = 1, \ldots, l_p$ *has a unique solution at* $(\lambda,\rho,\gamma) = (\lambda_0,\rho_0,\gamma_0)$*.*

For the asymptotic distribution of the GMM estimator, the gradient matrix $\frac{1}{n}\frac{\partial g_n(\theta_0)}{\partial \theta'}$ in the limit needs to be considered. For any square matrix C_n, let $C_n^s = C_n + C_n'$ and $\mathrm{vec}(C_n)$ be the vectorization of C_n. Denote $\bar{G}_n = -\frac{1}{n}\mathrm{E}(\frac{\partial g_n(\theta_0)}{\partial \theta'})$, the dimension of the vector γ by l_γ, and the jth element of γ by γ_j. Then,

$$\bar{G}_n = \frac{1}{n}\begin{pmatrix} \frac{1}{2}\omega_n' \Psi_{1n} & \frac{1}{2}\omega_n' \Psi_{2n} & 0 & \frac{1}{2}\omega_n' \Psi_{3n} \\ Q_n' \Upsilon_n & 0 & Q_n' \Sigma_{n0}^{-1/2} R_n X_n & 0 \end{pmatrix},$$

where $\omega_n = [\mathrm{vec}(P_{1n}^s), \cdots, \mathrm{vec}(P_{l_p n}^s)]$, $\Psi_{1n} = [\mathrm{vec}(J_{1n}^s), \cdots, \mathrm{vec}(J_{l_w n}^s)]$ with $J_{rn} = \Sigma_{n0}^{-1/2} R_n W_{rn} \times S_n^{-1} R_n^{-1} \Sigma_{n0}^{1/2}$, $\Psi_{2n} = [\mathrm{vec}(K_{1n}^s), \cdots, \mathrm{vec}(K_{l_m n}^s)]$ with $K_{rn} = \Sigma_{n0}^{-1/2} M_{rn} R_n^{-1} \Sigma_{n0}^{1/2}$, $\Psi_{3n} =$

$[\mathrm{vec}(L_{1n}^s), \cdots, \mathrm{vec}(L_{l_\gamma n}^s)]$ with $L_{rn} = \Sigma_{n0}^{-1/2} \frac{\partial \Sigma_{n0}^{1/2}}{\partial \gamma_r}$, and $\Upsilon_n = [\Upsilon_{1n}, \cdots, \Upsilon_{l_w n}]$ with $\Upsilon_{jn} = \Sigma_{n0}^{-1/2} R_n W_{jn} S_n^{-1} X_n \beta_0$. Thus, for $\lim_{n\to\infty} \bar{G}_n$ to have full column rank, under Assumption 2(i), a sufficient condition is that $\lim_{n\to\infty} \frac{1}{n}\omega_n'[\Psi_{2n}, \Psi_{3n}]$ has full column rank; on the other hand, for the situation of Assumption 2(ii), a sufficient condition is that $\lim_{n\to\infty} \frac{1}{n}\omega_n'[\Psi_{1n}, \Psi_{2n}, \Psi_{3n}]$ has full column rank. In addition, for the asymptotic distribution of the GMM estimator, we also need the usual condition that γ_0 is in the interior of its parameter space.

Assumption 3. (i) *In the case of Assumption 2(*i*), we assume that* $\lim_{n\to\infty} \frac{1}{n}\omega_n'[\Psi_{2n}, \Psi_{3n}]$ *has full column rank; in the case of Assumption 2(*ii*), we assume that* $\lim_{n\to\infty} \frac{1}{n}\omega_n'[\Psi_{1n}, \Psi_{2n}, \Psi_{3n}]$ *has full column rank.* (ii) θ_0 *is in the interior of* Θ.

Proposition 1. *Under Assumptions 1–2 and the condition that* $\lim_{n\to\infty} \frac{1}{n} a_n \,\mathrm{E}[g_n(\theta)]$ *is uniquely zero at* $\theta = \theta_0$*, the GMM estimator* $\hat{\theta}_{\mathrm{GMM}} = \arg\min_{\theta\in\Theta} g_n'(\theta) a_n' a_n g_n(\theta)$ *is consistent. If Assumption 3 also holds and* $\lim_{n\to\infty} \frac{1}{n} a_n \bar{G}_n$ *has full column rank, then* $\hat{\theta}_{\mathrm{GMM}}$ *has the asymptotic distribution*

$$\sqrt{n}(\hat{\theta}_{\mathrm{GMM}} - \theta_0) \xrightarrow{d} N\Big(0, \lim_{n\to\infty} (\bar{G}_n' a_n' a_n \bar{G}_n)^{-1} \bar{G}_n' a_n' a_n \bar{\Omega}_n a_n' a_n \bar{G}_n (\bar{G}_n' a_n' a_n \bar{G}_n)^{-1}\Big),$$

where $\bar{\Omega}_n = \mathrm{var}\big(\frac{1}{\sqrt{n}} g_n(\theta_0)\big)$.

As in Hansen (1982), the optimal choice of the weighting matrix $a_n' a_n$ is the inverse of the variance matrix $\bar{\Omega}_n$. Let $\mu_{30} = \mathrm{E}(v_i^3)$, $\mu_{40} = \mathrm{E}(v_i^4)$, and $\omega_{nd} = [d_{P_{1n}}, \cdots, d_{P_{l_p n}}]$, where d_C for a square matrix C is a column vector formed by the diagonal elements of C. Then, as in Lee (2007),

$$\bar{\Omega}_n = \frac{1}{n} \begin{pmatrix} (\mu_{40} - 3)\omega_{nd}'\omega_{nd} + \frac{1}{2}\omega_n'\omega_n & \mu_{30}\omega_{nd}' Q_n \\ \mu_{30} Q_n' \omega_{nd} & Q_n' Q_n \end{pmatrix}. \tag{5}$$

As $\bar{\Omega}_n$ is a variance matrix of moments, it would be a positive (or at least a nonnegative) definite matrix. It is well known that a nonnegative definite matrix can be written as a product of a matrix and its transpose. From the above expression of $\bar{\Omega}_n$, it is not written as a product of a square matrix and its transpose without a careful manipulation, because

$(\mu_{40}-3)$ might be negative. With a manipulation, it can be rewritten as

$$\bar{\Omega}_n = \frac{1}{n}\begin{pmatrix} \mu_{30}^2\omega_{nd}'\omega_{nd} + (\mu_{40}-1-\mu_{30}^2)\omega_{nd}'\omega_{nd} + (\frac{1}{2}\omega_n'\omega_n - 2\omega_{nd}'\omega_{nd}) & \mu_{30}\omega_{nd}'Q_n \\ \mu_{30}Q_n'\omega_{nd} & Q_n'Q_n \end{pmatrix}.$$

As $\mathrm{E}(v_i^2)=1$ and $\mathrm{E}[(v_i^2-1)^2]\cdot\mathrm{E}(v_i^2) \geq \mu_{30}^2$ by the Cauchy-Schwarz inequality, $\mu_{40}-1-\mu_{30}^2 \geq 0$. Note that the (j,k)th element of $\omega_n'\omega_n$ is $\mathrm{tr}(P_{jn}^s P_{kn}^s)$. For any $n \times n$ matrices C_{1n} and C_{2n}, $\mathrm{tr}\{\mathrm{diag}(C_{1n})[C_{2n} - \mathrm{diag}(C_{2n})]\} = 0$, where $\mathrm{diag}(C_{jn})$ is the diagonal matrix formed by the diagonal elements of C_{jn}. By writing each matrix P_{rn}^s as $\mathrm{diag}(P_{rn}^s) + [P_{rn}^s - \mathrm{diag}(P_{rn}^s)]$, we have $\frac{1}{2}\omega_n'\omega_n - 2\omega_{nd}'\omega_{nd} = \frac{1}{2}\Xi_n'\Xi_n$, where $\Xi_n = [\mathrm{vec}(P_{1n}^s - \mathrm{diag}(P_{1n}^s)), \ldots, \mathrm{vec}(P_{l_pn}^s - \mathrm{diag}(P_{l_pn}^s))]$. Let $\zeta_0 = \sqrt{(\mu_{40}-1-\mu_{30}^2)/2}$, $P_{rn,\zeta_0}^s = \zeta_0\,\mathrm{diag}(P_{rn}^s) + [P_{rn}^s - \mathrm{diag}(P_{rn}^s)]$, and $\omega_{n\zeta_0} = [\mathrm{vec}(P_{1n,\zeta_0}^s), \cdots, \mathrm{vec}(P_{l_pn,\zeta_0}^s)]$. Then $\bar{\Omega}_n$ can be rewritten as

$$\bar{\Omega}_n = \frac{1}{n}B_n'B_n, \text{ where } B_n = \begin{pmatrix} \frac{\sqrt{2}}{2}\omega_{n\zeta_0} & 0 \\ \mu_{30}\omega_{nd} & Q_n \end{pmatrix}. \tag{6}$$

According to this expression, it is apparent that $\bar{\Omega}_n$ is non-negative definite. The following assumption guarantees the positive definiteness of $\bar{\Omega}_n$ in the limit, and hence, it is so for a large enough n.

Assumption 4. $\lim_{n\to\infty}\frac{1}{n}Q_n'Q_n$ *and* $\lim_{n\to\infty}\frac{1}{n}\{\omega_{n\zeta_0}'\omega_{n\zeta_0} + 2\mu_{30}^2\omega_{nd}'[I_n - Q_n(Q_n'Q_n)^{-1}Q_n']\omega_{nd}\}$ *exist and are nonsingular.*

Let $\hat{\Omega}_n$ be an estimate of $\bar{\Omega}_n$ such that $\hat{\Omega}_n = \bar{\Omega}_n + o_p(1)$, e.g., $\hat{\Omega}_n$ can be derived by replacing the unknown true parameters in $\bar{\Omega}_n$ with their corresponding consistent estimators. A feasible optimal GMM (OGMM) estimator is $\hat{\theta}_{\mathrm{OGMM}} = \arg\min_{\theta\in\Theta} g_n'(\theta)\hat{\Omega}_n^{-1}g_n(\theta)$.

Proposition 2. *Under Assumptions 1–4,* $\hat{\theta}_{\mathrm{OGMM}}$ *is consistent and has the asymptotic distribution* $\sqrt{n}(\hat{\theta}_{\mathrm{OGMM}} - \theta_0) \xrightarrow{d} N(0, \lim_{n\to\infty}(\bar{G}_n'\bar{\Omega}_n^{-1}\bar{G}_n)^{-1})$.

The remaining issue of interest is to consider possible selections of P_{rn}'s and Q_n to minimize the asymptotic variance of an OGMM estimator based on linear and quadratic moments. For that propose, we are going to derive an upper bound of the precision matrix

$\bar{G}_n'\bar{\Omega}_n^{-1}\bar{G}_n$, and then consider the specification of P_{rn}'s and Q_n for a best GMM estimator which can attain that upper bound.

(1) Derive the upper bound. In order to derive an upper bound for precision matrices of GMM estimates, we investigate the use of the Cauchy-Schwarz inequality. Corresponding to the decomposition of $\bar{\Omega}_n$, we would like to have a product decomposition of $\bar{G}_n$, which involves B_n. Theoretically, this can always be done in principle because $\bar{G}_n = B_n' \cdot B_n(B_n'B_n)^{-1}\bar{G}_n = B_n'C_n$, where $C_n = B_n(B_n'B_n)^{-1}\bar{G}_n$, but C_n should depend only on structures of the model but not on those arbitrary P_{rn}'s and Q_n. Our subsequent analysis provides a more direct decomposition. Let $X_{n,\cdot j}$ be the jth column of X_n. The $\bar{G}_n$ can be written as $\bar{G}_n = \frac{1}{n}B_n'F_n$, where

$$F_n = \begin{pmatrix} F_{n,11} & F_{n,12} & F_{n,13} & F_{n,14} \\ \Upsilon_n & 0 & \Sigma_{n0}^{-1/2}R_nX_n & 0 \end{pmatrix},$$

with $F_{n,11} = \frac{1}{\sqrt{2}}\big[\mathrm{vec}\big(J^s_{1n,1/\zeta_0} - \frac{\mu_{30}}{\zeta_0}\mathrm{diag}(\Upsilon_{1n})\big), \cdots, \mathrm{vec}\big(J^s_{l_wn,1/\zeta_0} - \frac{\mu_{30}}{\zeta_0}\mathrm{diag}(\Upsilon_{l_wn})\big)\big]$, $F_{n,12} = \frac{1}{\sqrt{2}}[\mathrm{vec}(K^s_{1n,1/\zeta_0}), \cdots, \mathrm{vec}(K^s_{l_mn,1/\zeta_0})]$, $F_{n,13} = -\frac{\mu_{30}}{\sqrt{2}\zeta_0}\big[\mathrm{vec}\big(\mathrm{diag}(\Sigma_{n0}^{-1/2}R_nX_{n,\cdot 1})\big), \cdots, \mathrm{vec}\big(\mathrm{diag}(\Sigma_{n0}^{-1/2} \times R_nX_{n,\cdot l_x})\big)\big]$, and $F_{n,14} = \frac{1}{\sqrt{2}\zeta_0}[\mathrm{vec}(L^s_{1n}), \cdots, \mathrm{vec}(L^s_{l_\gamma n})]$, where $J^s_{rn,1/\zeta_0} = \frac{1}{\zeta_0}\mathrm{diag}(J^s_{rn}) + [J^s_{rn} - \mathrm{diag}(J^s_{rn})]$ and $K^s_{rn,1/\zeta_0}$ is similarly defined with K^s_{rn}. By the Cauchy-Schwarz inequality, $\bar{G}_n'\bar{\Omega}_n^{-1}\bar{G}_n = \frac{1}{n}F_n'B_n(B_n'B_n)^{-1}B_n'F_n \leq \frac{1}{n}F_n'F_n$. This upper bound is valid for the selection of any finite number of quadratic matrices P_{rn}'s and IV matrix Q_n because F_n depends on neither P_{rn}'s nor Q_n. The bound can be attained if B_n includes F_n in its column space. For that purpose, we shall find proper P_{rn}, $r = 1, \ldots, l_p$ and Q_n which can generate B_n.

(2) Derive the specification of P_{rn}'s and Q_n which can attain that upper bound. For a column vector $[\alpha_1, \ldots, \alpha_{l_p}, \alpha']'$ conformable with the rows of B_n, $B_n[\alpha_1, \ldots, \alpha_{l_p}, \alpha']' = \big[\frac{1}{\sqrt{2}}\mathrm{vec}\big(\zeta_0\,\mathrm{diag}(P_n^s) + [P_n^s - \mathrm{diag}(P_n^s)]\big), \frac{1}{2}\mu_{30}d_{P_n^s} + Q_n\alpha\big]'$, where $P_n^s = \sum_{j=1}^{l_p} P^s_{jn}\alpha_j$. The above vector can equal the rth column of F_n for $1 \leq r \leq l_w$ if $\zeta_0\,\mathrm{diag}(P_n^s) = \frac{1}{\zeta_0}\mathrm{diag}(J^s_{rn}) - \frac{\mu_{30}}{\zeta_0}\mathrm{diag}(\Upsilon_{rn})$, $P_n^s - \mathrm{diag}(P_n^s) = J^s_{rn} - \mathrm{diag}(J^s_{rn})$, and $\frac{1}{2}\mu_{30}d_{P_n^s} + Q_n\alpha = \Upsilon_{rn}$. By taking $\alpha_1 = \cdots = \alpha_{r-1} = 0$, $\alpha_r = 1$ and $\alpha_{r+1} = \cdots = \alpha_{l_p} = 0$, P^s_{rn} can be taken as $P^{*s}_{rn} = [J^s_{rn} - \mathrm{diag}(J^s_{rn})] + \frac{1}{\zeta_0^2}\mathrm{diag}(J^s_{rn}) - \frac{\mu_{30}}{\zeta_0^2}\mathrm{diag}(\Upsilon_{rn})$, and a column of Q_n can be taken as $Q^*_{rn} =$

$(1+\frac{\mu_{30}^2}{2\zeta_0^2})\Upsilon_{rn} - \frac{\mu_{30}}{2\zeta_0^2} d_{J_{rn}^s}$. Alternatively, we can have three matrices for quadratic moments: $J_{rn}^s - \operatorname{diag}(J_{rn}^s)$, $\operatorname{diag}(J_{rn}^s)$ and $\operatorname{diag}(\Upsilon_{rn})$, since P_{rn}^{*s} is a linear combination of these three matrices; and we can let an IV matrix Q_n contain Υ_{rn} and $d_{J_{rn}^s}$ as IVs. Note that when $\mu_{30} = 0$, the quadratic matrix $\operatorname{diag}(\Upsilon_{rn})$ and the IV $d_{J_{rn}^s}$ are redundant. These combined moments can avoid the use of more separate moments. But, on the other hand, using the moments separately avoids the estimation of the third and fourth moments of disturbances.

The (l_w+j)th column of F_n for $j = 1, \dots, l_m$ has a similar form as the first column of F_n, but with J_{1n} and Υ_{1n} replaced by, respectively, K_{jn} and $0_{n\times 1}$. Then we have the quadratic matrix $P_{l_w+j,n}^{*s} = [K_{jn}^s - \operatorname{diag}(K_{jn}^s)] + \frac{1}{\zeta_0^2}\operatorname{diag}(K_{jn}^s)$, and the IV $Q_{2n}^* = d_{K_{jn}^s}$. The (l_w+l_m+j)th column of F_n for $j = 1, \dots, l_x$ has a similar form as the first column of F_n, but with J_{1n} and Υ_{1n} replaced by, respectively, $0_{n\times n}$ and $\Sigma_{n0}^{-1/2} R_n X_{n,\cdot j}$. Then we have $P_{l_w+l_m+j,n}^{*s} = \operatorname{diag}(\Sigma_{n0}^{-1/2} R_n X_{n,\cdot j})$, and a column of Q_n can be taken as $Q_{l_w+l_m+j,n}^* = \Sigma_{n0}^{-1/2} R_n X_{n,\cdot j}$, where $j = 1, \dots, l_x$. The last l_γ columns of F_n have similar forms as the (l_w+1)th column of F_n, but L_{jn}'s are diagonal matrices, thus we have $P_{l_w+l_m+l_x+j,n}^{*s} = L_{jn}^s$, and a column of Q_n can be taken as $Q_{l_w+l_m+l_x+j,n}^* = d_{L_{jn}^s}$, where $j = 1, \dots, l_\gamma$.

The best P_{rn}'s and Q_n are summarized in the following proposition and the use of separate moments is presented in its corollary.

Proposition 3. *Suppose that Assumptions 1–4 hold. Let* $P_{jn}^* = [J_{jn} - \operatorname{diag}(J_{jn})] + \frac{1}{\zeta_0^2}\operatorname{diag}(J_{jn}) - \frac{\mu_{30}}{2\zeta_0^2}\operatorname{diag}(\Upsilon_{jn})$ *for* $j = 1, \dots, l_w$, $P_{l_w+j,n}^* = [K_{jn} - \operatorname{diag}(K_{jn})] + \frac{1}{\zeta_0^2}\operatorname{diag}(K_{jn})$ *for* $j = 1, \dots, l_m$, $P_{l_w+l_m+j,n}^* = \operatorname{diag}(\Sigma_{n0}^{-1/2} R_n X_{n,\cdot j})$ *for* $j = 1, \dots, l_x$, $P_{l_w+l_m+l_x+j,n}^* = L_{jn}$ *for* $j = 1, \dots, l_\gamma$, *and*

$$Q_n^* = \Big[\Big(1+\frac{\mu_{30}^2}{2\zeta_0^2}\Big)\Upsilon_{1n} - \frac{\mu_{30}}{2\zeta_0^2} d_{J_{1n}^s}, \cdots, \Big(1+\frac{\mu_{30}^2}{2\zeta_0^2}\Big)\Upsilon_{l_w n} - \frac{\mu_{30}}{2\zeta_0^2} d_{J_{l_w n}^s},$$
$$d_{K_{1n}^s}, \cdots, d_{K_{l_m n}^s}, \Sigma_{n0}^{-1/2} R_n X_n, d_{L_{1n}^s}, \cdots, d_{L_{l_\gamma n}^s}\Big].$$

The OGMM estimator with these quadratic matrices and Q_n^*, *denoted by* $\hat\theta_{\mathrm{BGMM}}$, *has the asymptotic distribution* $\sqrt{n}(\hat\theta_{\mathrm{BGMM}} - \theta_0) \xrightarrow{d} N(0, \lim_{n\to\infty}(\frac{1}{n}F_n'F_n)^{-1})$, *where* $(\frac{1}{n}F_n'F_n)^{-1} \le (\bar G_n'\bar\Omega_n^{-1}\bar G_n)^{-1}$ *for the asymptotic variance* $(\bar G_n'\bar\Omega_n^{-1}\bar G_n)^{-1}$ *of* $\hat\theta_{\mathrm{OGMM}}$ *given in Proposition 2. When* $\mu_{30} = 0$, *the quadratic matrices* $\operatorname{diag}(\Sigma_{n0}^{-1/2} R_n X_{n,\cdot 1})$, ..., $\operatorname{diag}(\Sigma_{n0}^{-1/2} R_n X_{n,\cdot l_x})$ *and the IVs* $d_{K_{1n}^s}$, ..., $d_{K_{l_m n}^s}$, $d_{L_{1n}^s}$, ..., $d_{L_{l_\gamma n}^s}$ *are redundant.*

Corollary 1. *Suppose that Assumptions 1–4 hold. Let P_{jn}^{**}'s be the matrices: $J_{rn}-\mathrm{diag}(J_{rn})$ for $r=1,\ldots,l_w$, $K_{rn}-\mathrm{diag}(K_{rn})$ for $r=1,\ldots,l_m$, $\mathrm{diag}(J_{rn})$ for $r=1,\ldots,l_w$, $\mathrm{diag}(K_{rn})$ for $r=1,\ldots,l_m$, L_{rn} for $r=1,\ldots,l_\gamma$, $\mathrm{diag}(\Upsilon_{rn})$ for $r=1,\ldots,l_w$, $\mathrm{diag}(\Sigma_{n0}^{-1/2}R_nX_{n,\cdot r})$ for $r=1,\ldots,l_x$; and denote $Q_n^{**}=[\Sigma_{n0}^{-1/2}R_nX_n,\Upsilon_n,d_{J_{1n}^s},\cdots,d_{J_{l_wn}^s},d_{K_{1n}^s},\cdots,d_{K_{l_mn}^s},d_{L_{1n}^s},\cdots,d_{L_{l_\gamma n}^s}]$. The OGMM estimator $\hat\theta_{\mathrm{BGMM2}}$ with these P_{jn}^{**}'s and Q_n^{**} satisfies $\sqrt{n}(\hat\theta_{\mathrm{BGMM2}}-\theta_0)\xrightarrow{d}N(0,\lim_{n\to\infty}(\frac{1}{n}F_n'F_n)^{-1})$, where $(\frac{1}{n}F_n'F_n)^{-1}\le(\bar G_n'\bar\Omega_n^{-1}\bar G_n)^{-1}$. When $\mu_{30}=0$, the quadratic matrices $\mathrm{diag}(\Upsilon_{1n})$, ..., $\mathrm{diag}(\Upsilon_{l_wn})$, $\mathrm{diag}(\Sigma_{n0}^{-1/2}R_nX_{n,\cdot 1})$, ..., $\mathrm{diag}(\Sigma_{n0}^{-1/2}R_nX_{n,\cdot l_x})$ and the IVs $d_{J_{1n}^s}$, ..., $d_{J_{l_wn}^s}$, $d_{K_{1n}^s}$, ..., $d_{K_{l_mn}^s}$, $d_{L_{1n}^s}$, ..., $d_{L_{l_\gamma n}^s}$ are redundant.*

The P_{rn}^*'s and Q_n^* in Proposition 3 involve some unknown parameters. In practice, they can be replaced with their respective consistent estimates, which results in a feasible best GMM (FBGMM) estimator. The FBGMM estimator with the estimated P_{rn}^*'s and Q_n^* will have the same asymptotic distribution as that in Proposition 3. The same comment applies to Corollary 1.

When $\mu_{30}=0$, the best quadratic matrices for quadratic moments reduce to $[J_{rn}-\mathrm{diag}(J_{rn})]+\frac{1}{\zeta_0^2}\mathrm{diag}(J_{rn})$ for $r=1$, ..., l_w, $[K_{rn}-\mathrm{diag}(K_{rn})]+\frac{1}{\zeta_0^2}\mathrm{diag}(K_{rn})$ for $r=1$, ..., l_m, L_{rn} for $r=1$, ..., l_γ, and the best IV matrix is $[\Upsilon_n,\Sigma_{n0}^{-1/2}R_nX_n]$. Furthermore, if $\mu_{40}=3$ also holds, e.g., when v_i's are normal, then $\zeta_0=1$ and the best quadratic matrices become J_{1n}, ..., J_{l_wn}, K_{1n}, ..., K_{l_mn}, L_{1n}, ..., $L_{l_\gamma n}$. Thus, when $\mu_{30}=0$ and $\mu_{40}=3$, the best GMM has the moment vector

$$\begin{aligned}g_n^{\#}(\theta)=\big[&V_n'(\theta)J_{1n}V_n(\theta)-\mathrm{tr}(J_{1n}),\cdots,V_n'(\theta)J_{l_wn}V_n(\theta)-\mathrm{tr}(J_{l_wn}),\\&V_n'(\theta)K_{1n}V_n(\theta)-\mathrm{tr}(K_{1n}),\cdots,V_n'(\theta)K_{l_mn}V_n(\theta)-\mathrm{tr}(K_{l_mn}),\\&V_n'(\theta)L_{1n}V_n(\theta)-\mathrm{tr}(L_{1n}),\cdots,V_n'(\theta)L_{l_\gamma n}V_n(\theta)-\mathrm{tr}(L_{l_\gamma n}),V_n'(\theta)[\Upsilon_n,\Sigma_{n0}^{-1/2}R_nX_n]\big]'.\end{aligned}\tag{7}$$

This best moment vector when $\mu_{30}=0$ and $\mu_{40}=3$ is similar to that implied by the QML score of model (1). These can be seen as follows. The quasi log likelihood function of model (1), as if the disturbances were normal, is

$$\ln L_n(\theta)=-\frac{n}{2}\ln(2\pi)+\ln|R_n(\rho)S_n(\lambda)|-\ln|\Sigma_n^{1/2}(\gamma)|-\frac{1}{2}V_n'(\theta)V_n(\theta).\tag{8}$$

With $Y_n = S_n^{-1}(X_n\beta_0 + R_n^{-1}\Sigma_{n0}^{1/2}V_n)$, the first order derivatives of $\ln L_n(\theta)$ at $\theta = \theta_0$ are

$$\frac{\partial \ln L_n(\theta_0)}{\partial \lambda_r} = V_n' J_{rn} V_n - \mathrm{tr}(J_{rn}) + \Upsilon_{rn}' V_n, \text{ for } r = 1, \ldots, l_w,$$
$$\frac{\partial \ln L_n(\theta_0)}{\partial \rho_r} = V_n' K_{rn} V_n - \mathrm{tr}(K_{rn}), \text{ for } r = 1, \ldots, l_m,$$
$$\frac{\partial \ln L_n(\theta_0)}{\partial \beta} = X_n' R_n' \Sigma_{n0}^{-1/2} V_n,$$
$$\frac{\partial \ln L_n(\theta_0)}{\partial \gamma_r} = V_n' L_{rn} V_n - \mathrm{tr}(L_{rn}), \text{ for } r = 1, \ldots, l_\gamma.$$

Thus, for the case with $\mu_{30} = 0$ and $\mu_{40} = 3$, the best moment vector $g_n^{\#}(\theta)$ in (7) is similar to that implied by the QML score. In the derivative of $\ln L_n(\theta)$ with respect to λ, a quadratic moment and a linear one are combined with equal weights, while $g_n^{\#}(\theta)$ uses them separately. However, in the case with $\mu_{30} = 0$ and $\mu_{40} = 3$, which includes the case with normal disturbances, the QML estimator can be shown to be asymptotically as efficient as our best GMM estimator. Debarsy et al. (2015) provide a proof for the case of the matrix exponential spatial specification model. If $\mu_{30} \neq 0$ or $\mu_{40} \neq 3$, the OGMM estimator with $g_n^{\#}(\theta)$ can be asymptotically more efficient than the QML estimator due to its use of optimal weighting. Nevertheless, in general, this OGMM estimator is asymptotically less efficient relative to the best GMM estimator in Proposition 3 or its corollary when disturbances are not normally distributed.

2.2 Normality tests

In this section, we derive two normality tests for the SARAR model in Section 2: one employs the Lagrangian multiplier principle and the Pearson distribution as in Jarque and Bera (1980), and the other directly uses the vector of estimated skewness and excess kurtosis coefficients, as in Bera, Doğan and Taşpınar (2021, 2022).

The Pearson distribution has the probability density function:

$$h(x,a) = \frac{\exp(-\int \frac{x+a_1}{a_2x^2+a_1x+a_0}\,\mathrm{d}x)}{\int_{-\infty}^{\infty} \exp(-\int \frac{x+a_1}{a_2x^2+a_1x+a_0}\,\mathrm{d}x)\,\mathrm{d}x} = \frac{\exp(-h_1(x,a))}{\int_{-\infty}^{\infty} \exp(-h_1(x,a))\,\mathrm{d}x},$$

where a_0, a_1 and a_2 are constants, $a = [a_0, a_1, a_2]'$, and $h_1(x,a) = \int \frac{x+a_1}{a_2x^2+a_1x+a_0}\,\mathrm{d}x$. Suppose that v_i in model (1) has the density function $h(x,a)$. As v_i has unit variance, a satisfies the restriction $\sigma^2(a) = 1$, where $\sigma^2(a) = \int_{-\infty}^{\infty} x^2 h(x,a)\,\mathrm{d}x - [\int_{-\infty}^{\infty} x h(x,a)\,\mathrm{d}x]^2$. The null hypothesis for the normality test is H_0: $a_1 = a_2 = 0$. Under the null, $a_0 = 1$. The log likelihood function of the SARAR model (1) is

$$\ln L_n(\delta) = \sum_{i=1}^{n} \ln h(v_i(\theta), a) + \ln |S_n(\lambda)| + \ln |R_n(\rho)| - \ln |\Sigma_n^{1/2}(\gamma)|,$$

where $\delta = [\theta', a']'$, and $v_i(\theta) = e'_{ni}\Sigma_n^{-1/2}(\gamma)R_n(\rho)[S_n(\lambda)Y_n - X_n\beta]$ with e_{ni} being the ith column of I_n. Denote $a_{12} = [a_1, a_2]'$. The Lagrangian function for the restricted estimation with $a_{12} = 0$ imposed is $\ln L_n(\delta) - b_1[\sigma^2(a) - 1] - b_2' a_{12}$, where b_1 is a scalar, and b_2 is a 2×1 vector. Let $\check{a} = [1, 0, 0]'$, and $\check{\delta} = [\check{\theta}', \check{a}']'$ be the restricted MLE with $a_{12} = 0$ imposed. Then we have the following first order conditions $\frac{\partial \ln L_n(\check{\delta})}{\partial \theta} = 0$, $\frac{\partial \ln L_n(\check{\delta})}{\partial a_0} - \check{b}_1 \frac{\partial \sigma^2(\check{a})}{\partial a_0} = 0$, and $\frac{\partial \ln L_n(\check{\delta})}{\partial a_{12}} - \check{b}_1 \frac{\partial \sigma^2(\check{a})}{\partial a_{12}} - \check{b}_2 = 0$. Solving for the Lagrangian multiplier $\check{b}_2$ yields $\check{b}_2 = \frac{\partial \ln L_n(\check{\delta})}{\partial a_{12}} - \frac{\partial \sigma^2(\check{a})}{\partial a_{12}} (\frac{\partial \sigma^2(\check{a})}{\partial a_0})^{-1} \frac{\partial \ln L_n(\check{\delta})}{\partial a_0}$. Our test is on the basis of the asymptotic distribution of $\check{b}_2$. Since $\frac{\partial \ln L_n(\check{\delta})}{\partial a} = -\sum_{i=1}^{n} \frac{\partial h_1(v_i(\check{\theta}), \check{a})}{\partial a} + \frac{n}{\int_{-\infty}^{\infty} \exp(-h_1(x,\check{a}))\,\mathrm{d}x} \int_{-\infty}^{\infty} \exp(-h_1(x,\check{a})) \frac{\partial h_1(x,\check{a})}{\partial a}\,\mathrm{d}x$, where $\frac{\partial h_1(x,\check{a})}{\partial a} = [-\frac{1}{2}x^2, x - \frac{1}{3}x^3, -\frac{1}{4}x^4]'$, and $\frac{\partial \sigma^2(\check{a})}{\partial a} = [1, 0, 3]'$, we have

$$\check{b}_2 = \left[\frac{1}{3}\sum_{i=1}^{n} v_i^3(\check{\theta}) - \sum_{i=1}^{n} v_i(\check{\theta}),\ \frac{1}{4}\sum_{i=1}^{n}[v_i^4(\check{\theta}) - 3] - \frac{3}{2}\sum_{i=1}^{n}[v_i^2(\check{\theta}) - 1]\right]'. \tag{9}$$

Since $\check{\theta}$ is the QMLE of model (1) as if the disturbances were normally distributed, as in Lee (2004), we can show that $\sqrt{n}(\check{\theta} - \theta_0) = O_p(1)$ under regularity conditions. Then by the mean value theorem, we have

$$\frac{1}{\sqrt{n}}\check{b}_2 = \frac{1}{\sqrt{n}}\left[\frac{1}{3}\sum_{i=1}^{n} v_i^3 - \sum_{i=1}^{n} v_i,\ \frac{1}{4}\sum_{i=1}^{n}(v_i^4 - 3) - \frac{3}{2}\sum_{i=1}^{n}(v_i^2 - 1)\right]' + o_p(1)$$

under the null hypothesis, i.e., using the efficient estimator $\check{\theta}$ of θ does not affect the asymptotic distribution of $\check{b}$. By the Lindeberg-Lévy central limit theorem, $\frac{1}{\sqrt{n}}\check{b}_2$ follows the asymptotic distribution $N\Big(0, \left(\begin{smallmatrix} \frac{2}{3} & 0 \\ 0 & \frac{3}{2} \end{smallmatrix}\right)\Big)$, where the two components of $\frac{1}{\sqrt{n}}\check{b}_2$ are asymptotically independent. Therefore, we have the test statistic

$$\boldsymbol{t}_{1n} = \frac{1}{6n}\left[\sum_{i=1}^{n} v_i^3(\check{\theta}) - 3\sum_{i=1}^{n} v_i(\check{\theta})\right]^2 + \frac{1}{24n}\left[\sum_{i=1}^{n}[v_i^4(\check{\theta}) - 3] - 6\sum_{i=1}^{n}[v_i^2(\check{\theta}) - 1]\right]^2, \tag{10}$$

which is asymptotically chi-squared distributed with two degrees of freedom.

We also construct a normality test based on estimated skewness and excess kurtosis coefficients with the QMLE $\check{\theta}$, following Bera et al. (2021, 2022). Denote $\check{v}_i = v_i(\check{\theta})$ for simplicity. As v_i has unit variance, the skewness and excess kurtosis coefficient estimates of v_i with $\check{\theta}$ are, respectively, $\frac{1}{n}\sum_{i=1}^n \check{v}_i^3$ and $\frac{1}{n}\sum_{i=1}^n(\check{v}_i^4 - 3)$. A normality test can be based the asymptotic distribution of $\hat{\mathcal{S}}_n \equiv [\frac{1}{\sqrt{n}}\sum_{i=1}^n \check{v}_i^3, \frac{1}{\sqrt{n}}\sum_{i=1}^n(\check{v}_i^4 - 3)]'$. By the mean value theorem, $\hat{\mathcal{S}}_n = \mathcal{S}_n + [\frac{3}{n}\sum_{i=1}^n v_i^2(\tilde{\theta})\frac{\partial v_i(\tilde{\theta})}{\partial \theta}, \frac{4}{n}\sum_{i=1}^n v_i^3(\tilde{\theta})\frac{\partial v_i(\tilde{\theta})}{\partial \theta}]'\sqrt{n}(\check{\theta} - \theta_0)$, where $\mathcal{S}_n = [\frac{1}{\sqrt{n}}\sum_{i=1}^n v_i^3, \frac{1}{\sqrt{n}}\sum_{i=1}^n(v_i^4 - 3)]'$, and $\tilde{\theta}$ lies between $\check{\theta}$ and θ_0 elementwise. As in Lee (2004), we can show that $\sqrt{n}(\check{\theta} - \theta_0) = \Delta_{1n}^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial \theta} + o_p(1)$, where $\Delta_{1n} \equiv -\frac{1}{n}\mathrm{E}(\frac{\partial^2 \ln L_n(\theta_0)}{\partial \theta \partial \theta'}) = \frac{1}{n}\Lambda_n'\Lambda_n$ with

$$\Lambda_n = \begin{pmatrix} \frac{1}{\sqrt{2}}\Psi_{1n} & \frac{1}{\sqrt{2}}\Psi_{2n} & 0 & \frac{1}{\sqrt{2}}\Psi_{3n} \\ \Upsilon_n & 0 & \Sigma_{n0}^{-1/2}R_n X_n & 0 \end{pmatrix}.$$

Under the null hypothesis of normal disturbances, $[\frac{3}{n}\sum_{i=1}^n v_i^2(\tilde{\theta})\frac{\partial v_i(\tilde{\theta})}{\partial \theta}, \frac{4}{n}\sum_{i=1}^n v_i^3(\tilde{\theta})\frac{\partial v_i(\tilde{\theta})}{\partial \theta}]' = \Delta_{2n} + o_p(1)$, where $\Delta_{2n} = -\frac{3}{n}\begin{pmatrix} 0 & 1_n' \\ 2\sqrt{2}\,\mathrm{vec}'(I_n) & 0 \end{pmatrix}\Lambda_n$. Then $\hat{\mathcal{S}}_n = \mathcal{S}_n + \Delta_{2n}\Delta_{1n}^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial \theta} + o_p(1)$. Under the null hypothesis, the variance of $\mathcal{S}_n$ is $\Delta_3 = (\begin{smallmatrix} 15 & 0 \\ 0 & 96 \end{smallmatrix})$, and the variance of $\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial \theta}$ is equal to Δ_{1n} by the information matrix equality. As $\mathrm{E}(\mathcal{S}_n \cdot \frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial \theta'}) = -\Delta_{2n}$, the variance of $\mathcal{S}_n + \Delta_{2n}\Delta_{1n}^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial \theta}$ is $\Delta_n = \Delta_3 - \Delta_{2n}\Delta_{1n}^{-1}\Delta_{2n}'$. By the central limit theorem for general linear quadratic forms in Lemma 6 of Yang and Lee (2017), $\hat{\mathcal{S}}_n \xrightarrow{d} N(0, \lim_{n\to\infty}\Delta_n)$. Let $\hat{\Delta}_{jn}$ be the estimate of Δ_{jn} for $j = 1, 2$. The $\hat{\Delta}_{jn}$ can be derived by replacing the unknown θ_0 in Δ_{jn} with $\check{\theta}$. Denote $\hat{\Delta}_n = \Delta_3 - \hat{\Delta}_{2n}\hat{\Delta}_{1n}^{-1}\hat{\Delta}_{2n}'$. Then we have the test statistic

$$\boldsymbol{t}_{2n} = \hat{\mathcal{S}}_n'\hat{\Delta}_n^{-1}\hat{\mathcal{S}}_n \tag{11}$$

for the normality of v_i, which is asymptotically chi-squared distributed with two degrees of freedom under the null. The test statistic $\boldsymbol{t}_{1n}$ has a simple form, and it is locally most powerful if the true disturbance distribution is a Pearson distribution, which might not be the case for other disturbance distributions, while $\boldsymbol{t}_{2n}$ follows a simple principle to jointly test the skewness and excess kurtosis.

2.3 Monte Carlo experiments

In this section, we report some Monte Carlo results on the finite sample performance of GMM estimators with best linear and quadratic moments for the SARAR model (1).[7]

We consider a SARAR model with two spatial lags of the dependent variable and one spatial lag of disturbances. We specify the spatial weight matrix W_{1n} for the first spatial lag as a block diagonal matrix with each diagonal block matrix being the matrix for the study of crimes across 49 districts in Columbus, Ohio, in Anselin (1988). The spatial weight matrix W_{2n} for the second spatial lag is based on the queen criterion. The spatial weight matrix M_n for disturbances is set to be the same as W_{2n}. The W_{1n}, W_{2n} and M_n are normalized to have row sums equal to 1. The exogenous variable matrix X_n contains three variables, where the first one is an intercept term and the other two are randomly drawn from $N(0,1)$. The heteroskedasticity function $f(x_{n,i1}, x_{n,i2}, x_{n,i3}, \gamma)$ is set to $\exp(x_{n,i1}\gamma_1 + x_{n,i2}\gamma_2 + x_{n,i3}\gamma_3)$. The true parameter values of λ_1, λ_2, ρ, $\beta = [\beta_1, \beta_2, \beta_3]'$ and $\gamma = [\gamma_1, \gamma_2, \gamma_3]'$ are set to, respectively, 0.3, 0.3, 0.3, $[1,1,1]'$, and $[\frac{\ln(2)}{4}, \sqrt{\frac{\ln(2)}{8}}, \sqrt{\frac{\ln(2)}{8}}]'$. With these parameter values, the unconditional variance of $f(x_{i1}, x_{i2}, \gamma_0)v_i$ is equal to 2. The sample size is either 196 or 392, and the number of Monte Carlo repetitions is $2,000$.

We consider the following estimators: the QML estimator, an OGMM estimator with some simple linear and quadratic moments (denoted as SGMM), an OGMM estimator with theoretically best linear and quadratic moments (BGMM),[8] an OGMM estimator with estimated best linear and quadratic moments (feasible BGMM, i.e., FBGMM), an OGMM estimator robust to heteroskedasticity in Lin and Lee (2010) (denoted as RGMM), and the generalized spatial two-stage least squares estimator robust to heteroskedasticity in Kelejian and Prucha (2010) (denoted as GS2SLS). While SGMM, BGMM and FBGMM are based on moments of the form in (2), which takes into account heteroskedastic variances in estimation, RGMM is based on the moment vector $[\epsilon_n'(\delta)\mathring{P}_{1n}\epsilon_n(\delta), \cdots, \epsilon_n'(\delta)\mathring{P}_{l_p n}\epsilon_n(\delta), \epsilon_n'(\delta)Q_n]'$, where $\epsilon_n(\delta) = R_n(\rho)[S_n(\lambda)Y_n - X_n\beta]$ with $\delta = [\lambda', \rho, \beta']'$ and $\mathring{P}_{jn}$'s have zero diagonals. For SGMM,

[7]The empirical sizes and powers of the proposed normality tests of disturbances are reported in the supplementary file.

[8]The optimal weighting matrix is also computed using the true parameters.

the IVs for linear moments are $[X_n, W_{1n}X_{1n}, W_{2n}X_{1n}]$, where X_{1n} excludes the intercept term in X_n to avoid multicollinearity, and the square matrices for quadratic moments are I_n, W_{1n}, W_{2n}, W_{1n}^2, W_{2n}^2, $\mathrm{diag}(x_{n,11}, \cdots, x_{n,n1})$ and $\mathrm{diag}(x_{n,12}, \cdots, x_{n,n2})$, where the last two matrices are included since the heteroskedasticity function involves X_n. For RGMM, the IVs for linear moments are $[X_n, W_{1n}X_{1n}, W_{2n}X_{1n}]$, and the square matrices for quadratic moments are W_{1n}, W_{2n}, $W_{1n}^2 - \mathrm{diag}(W_{1n}^2)$ and $W_{2n}^2 - \mathrm{diag}(W_{2n}^2)$. The SGMM estimator is used as the initial estimate to estimate theoretically best linear and quadratic moments for FBGMM. Recall that BGMM and FBGMM are asymptotically equivalent, and they are asymptotically more efficient than QML when disturbances are non-normal. As there are a few outliers for SGMM, RGMM, BGMM and FBGMM, we report the following robust measures of bias and dispersion for various estimators: the median bias (MB), median absolute deviation (MAD), and interdecile range (IDR), which is the difference between the 0.9 and 0.1 quantiles of data points.

Table 1 reports the estimation results for the case with gamma$(1, 1)$-distributed disturbances, which have been adjusted to have mean zero.[9] All estimators have a relatively small MB. FBGMM generally performs similarly to BGMM, and neither FBGMM nor BGMM dominates each other. FBGMM and BGMM have smaller MADs and IDRs than other estimators, except those for β_2 in the case of the smaller sample size $n = 196$. For example, for λ_2, the MAD and IDR of FBGMM are about 20% smaller than those of QML. The MAD and IDR of SGMM are significantly larger than those of FBGMM and QML. RGMM and GS2SLS have larger MADs and IDRs than SGMM. The MAD and IDR of RGMM are smaller than those of G2SLS for all parameters except ρ. As n increases from 196 to 392, the MB, MAD and IDR of various estimators decrease.

[9]The disturbances have unit variance, skewness 2 and excess kurtosis 6. Results for other disturbance distributions including the normal distribution are in the supplementary file.

Table 1: Performances of various estimators for the SARAR model

			QML	BGMM	FBGMM	SGMM	RGMM	GS2SLS
$n = 196$	λ_1	MB	-0.007	0.001	0.000	-0.001	-0.006	0.006
		MAD	0.039	0.034	0.033	0.044	0.049	0.085
		IDR	0.148	0.135	0.130	0.171	0.194	0.338
	λ_2	MB	-0.010	0.003	0.002	0.006	-0.009	0.026
		MAD	0.075	0.060	0.060	0.091	0.118	0.118
		IDR	0.300	0.243	0.235	0.353	0.459	0.454
	ρ	MB	-0.006	0.011	0.009	-0.005	0.003	-0.057
		MAD	0.105	0.098	0.098	0.122	0.158	0.153
		IDR	0.410	0.367	0.386	0.470	0.597	0.559
	β_1	MB	0.003	0.072	0.075	0.041	-0.074	-0.070
		MAD	0.199	0.174	0.177	0.232	0.281	0.300
		IDR	0.790	0.719	0.706	0.975	1.092	1.172
	β_2	MB	-0.009	0.004	-0.001	0.004	-0.045	-0.014
		MAD	0.049	0.051	0.049	0.061	0.075	0.078
		IDR	0.189	0.196	0.190	0.227	0.281	0.288
$n = 392$	λ_1	MB	-0.004	-0.001	-0.001	-0.001	-0.004	0.006
		MAD	0.027	0.023	0.022	0.029	0.035	0.062
		IDR	0.100	0.086	0.084	0.112	0.135	0.232
	λ_2	MB	-0.007	-0.001	-0.003	0.001	-0.007	0.011
		MAD	0.053	0.042	0.042	0.063	0.083	0.084
		IDR	0.214	0.157	0.154	0.238	0.314	0.317
	ρ	MB	0.001	0.004	0.005	-0.001	-0.001	-0.027
		MAD	0.076	0.064	0.063	0.084	0.108	0.105
		IDR	0.288	0.254	0.255	0.313	0.414	0.402
	β_1	MB	0.005	0.041	0.048	0.021	-0.022	-0.052
		MAD	0.132	0.117	0.115	0.157	0.194	0.212
		IDR	0.544	0.446	0.443	0.614	0.769	0.795
	β_2	MB	-0.004	0.002	0.000	0.001	-0.024	-0.008
		MAD	0.035	0.033	0.033	0.040	0.055	0.055
		IDR	0.135	0.127	0.126	0.157	0.210	0.215

Notes: The true value of $[\lambda_1, \lambda_2, \rho, \beta_1, \beta_2, \beta_3]$ is $[0.3, 0.3, 0.3, 1, 1, 1]$. The statistics for β_3 are similar to those for β_2, thus they are omitted to save space. The number of Monte Carlo repetitions is $2,000$. MB, MAD and IDR stand for, respectively, median bias, median absolute deviation, and interdecile range.

3 Spatial interdependence patterns of employment growth in US counties

Regional urban expansion and employment growth in the United States and worldwide has attracted much attention in urban economics and regional science. Numerous studies in general interest and economic journals (for instance, Combes, 2000; Wheeler, 2001; Burchfield et al., 2006; Duranton and Turner, 2012; Bettencourt, 2013; Reia et al., 2022; Peters, 2022, among others) focus on social, economic, industrial, and infrastructural causes and consequences of regional growth. As these social and economic factors could be spatially interdependent, certain questions on regional growth remain unanswered: does a county's growth resemble its spatially or economically neighboring counties? What are the spatial interdependence patterns of regional employment growth?

We study the degree to which employment growth disseminates through different channels by investigating the spatial interdependence patterns of growth on employments and other economic variables in US counties using a high order SARAR model with spatial errors and heteroskedasticity. It is documented in previous studies that local economic activities could propagate spatially. For example, Wheeler (2001) estimates spatial correlograms of population, employment, income and earnings of US counties and shows the spatial correlation structure of county-level growth. Gebremariam et al. (2010) find strong agglomerative effects using employment growth and median household income growth in the Appalachia region in the 1990s. Feyrer et al. (2017) and James and Smith (2020) examine how economic shocks propagate geographically using the recent boom in oil and gas production in US. Economic linkage, technological spillover, and suburbanization due to transportation can be explanations of spatial interdependence in economic growth (Ertur and Koch, 2007; Baum-Snow, 2007).

We characterize different neighbors of a county by geographic adjacency and distance, industrial proximity, and political tendency. The spatial autoregressive effects through these channels are estimated together by putting them in the high order SARAR model, which allows us to compare the spatial correlations of a county to its neighbors defined differently. Extending the literature, we could characterize the spatial interdependence patterns of regional economic growth in various angels. Such an empirical model usually have larger standard deviations as more spatial lags are simultaneously estimated. Thus an efficient estimator is beneficial by reducing the standard deviations and narrowing the confidence interval as shown below.

Moreover, the results provide evidence on conditional convergence of local economic growth in spatial viewpoints. Employment growth in a county is not only conditional on initial conditions but also is conditional upon growth in neighboring counties. A county's employment growth has a larger spatial correlation with its neighbours which are geographically adjacent or with a higher degree of industrial proximity.

3.1 Empirical model specification

The key focus is on the spatial dependence of employment growth in county-level, where the dependence materializes in various ways which are captured by four different spatial or network weight matrices. The first matrix, W_{co}, which is named as an adjacent matrix in the literature, represents spatial interdependence between counties which share borders. Specifically, its (i, j)th element equals to one when the ith and jth counties are adjacent to each other. Numerous studies on economics and regional science adopt this spatial weight matrix. The second matrix, W_{di}, is a distance-based matrix. Wheeler (2001) shows that a county's growth rate is correlated to its neighbors within 200 miles and the correlations decline at a substantially higher rate when distances are beyond 40 miles. Feyrer et al. (2017) and James and Smith (2020) examine how economic shocks propagate geographically and find the outward spillover is limited to within 100 to 200 miles depending on measures for economic variables. Therefore, we construct the entries of W_{di} being inversely proportional to the square of distances between two counties i and j, i.e., $1/d_{ij}^2$ with d_{ij} being the distance, which is similar to specifications in the literature (Ertur and Koch, 2007). The third spatial weight matrix, W_{ec}, is constructed by the "difference" of industrial structures among counties ("ec" stands for economic structure). As economic and industrial structures influence local employment growth (Combes, 2000), a county's growth pattern might resemble its neighbors which have similar industrial structures. In the benchmark model, we use the squared differences of employments in 7 sectors between two counties within D_L miles to construct the weights in W_{ec}, i.e., $\mathbf{1}(d_{ij} \leq D_L) \times 1/[(Sec_{1i} - Sec_{1j})^2 + \cdots + (Sec_{7i} - Sec_{7j})^2]$, where the 7 sectors are natural resources and mining, construction, manufacturing, transportation, educational and health services, other services, and governments, and the classification is similar to that in Feyrer et al. (2017). Sec_{ki} is the number of workers in sector k of county i on the initial year 2000. The fourth matrix, W_{po}, captures the neighbors that share similar political tendency. Its (i, j)th entry is calculated as the inverse of the squared difference of democratic and republican shares of voting in the 2000 presidential elections between two counties within D_L miles, i.e., $\mathbf{1}(d_{ij} \leq D_L) \times 1/[(Dem_i - Dem_j)^2 + (Rep_i - Rep_j)^2]$, where Dem_i (or Rep_i) represents county i's average share of voting for a democratic (or republican) candidate in these three presidential elections. For matrices W_{ec} and W_{po}, the neighbors are limited to counties up to 200 miles away, which is consistent with the findings in Wheeler (2001) and Feyrer et al. (2017). Other distance limits are also considered for robustness analysis.

We employ a county's employment growth rate, G_{Emp} that is the difference between logged employment in 2000 and 2005, as the dependent variable in the benchmark model. We also study the county-level economic growth that can be measured in a variety of ways in order to keep the empirical investigation broad. Other measures are the growth rate of establishments, the growth rate of income per capita, and the growth rates of employment, establishments, and income per capita in good-producing sectors: natural resources and mining, construction, and manufacturing.

The explanatory variables include variables at the initial period which can reflect the initial economic conditions: employment in 2000; and variables at the initial period which have effects on economic growth: population, establishment spatial density (defined as the ratio between establishment and area), female householders density (ratio between female householders and population), owner-occupied house units density, percentages of employed in manufacturing and service sectors, percentage of adults (larger than 25 years old) with high school/college degrees and above in 2000. Variables reflecting the changes of economic conditions during the investigated period such as growth rate of population are also included. Denote them as X, and the empirical model is

$$\begin{aligned} G_{emp,i} =& \lambda_{co} \sum_{j=1}^{n} W_{ij,co} G_{emp,j} + \lambda_{di} \sum_{j=1}^{n} W_{ij,di} G_{emp,j} + \lambda_{ec} \sum_{j=1}^{n} W_{ij,ec} G_{emp,j} \\ &+ \lambda_{po} \sum_{j=1}^{n} W_{ij,po} G_{emp,j} + X_i \beta + u_i, \text{ and} \\ u_i =& \rho \sum_{j=1}^{n} M_{ij} u_i + \Sigma_i^{1/2}(pop_{2000,i}) v_i. \end{aligned} \tag{12}$$

The proposed GMM estimator with best moments is more efficient than other methods such as QML, which allows multiple spatial lags, spatial errors and heteroskedastic disturbances to be estimated simultaneously. We also consider heteroskedasticity and spatial errors in the model. The disturbance u_i could be heteroskedastic for counties with different population sizes in the initial year and we consider a linear form such that $\sigma_{ni} = \gamma_0 + \gamma_1 \ln(pop_{2000,i})$ where σ_{ni} is the ith diagonal entry of $\Sigma_i^{1/2}(pop_{2000,i})$. We use a distance-based matrix M_n as the spatial weight matrix for spatial errors, where M_n captures the spatial interdependence from neighbors which are within 200 to 500 miles of a county in order to avoid any missing spillover from others.

3.2 Data

We employ census-type data on counties' economic performance in 2000 and 2005 in the United States. The data are from the US Census Bureau and the US Bureau of Labor Statistics, and cover all the counties in the contiguous US. The variable list is in Table 2, with description, mean, standard deviations, minimum, and maximum being provided. We use data for 3074 counties in estimation. The growth rates of employments during 2000-2005 vary from -32.6% to 37.8% with a mean growth rate of 1.30%. The counties with faster growth rates are located mainly in the east coast states and western states while counties in great lakes states mostly have the slowest growth rates. All the spatial weight matrices

Table 2: Variable descriptions and summary statistics

Variable	Description	Mean	Std	Min	Max
G_{emp}	Growth rate of employment	1.30%	9.80%	-32.60%	37.80%
G_{pop}	Growth rate of population	1.50%	7.40%	-20.50%	49.50%
pop_{2000}	Population in 2000 (in thousands)	90.507	294.986	0.356	9519.338
emp_{2000}	Employment in 2000 (in thousands)	40.977	147.784	0.06	4110.915
est_{2000}	Establishment in 2000 (in thousands)	2.408	8.593	0.015	297.191
$pop_{2000}^{(d)}$	Spatial density of population in 2000	0.241	1.632	0	64.157
$pop_{2000}^{(25-54)}$	Percentage of population in ages 25-54	41.067	3.534	23.057	61.207
ump_{2000}	Unemployment rate in 2000	4.328	1.644	1.3	17.4
fhh_{2000}	female householder density in 2000	0.04	0.014	0.006	0.107
hmw_{2000}	Owner-occupied house units density in 2000	0.284	0.038	0.053	0.394
mnf_{2000}	Percentage of employed in manufacturing in 2000	15.7	12.8	0	70.4
sev_{2000}	Percentage of employed in professional, financial, information and other services (excluding health and education)	20.8	9.8	0	89.9
$est_{2000}^{(d)}$	Spatial density of establishment in 2000	0.007	0.085	0	4.498
col_{2000}	Percentage of adults (>25) with college degrees and above in 2000	16.483	7.746	4.9	63.7
hig_{2000}	Percentage of adults (>25) with high school diplomas and above	77.31	8.71	34.7	96.3

Notes: the descriptive statistics summarize employment, establishment, income and other variables of economic performance for 3074 counties in contiguous US with around 1% being excluded to avoid outliers as the estimation results are sensitive to their inclusion. The spatial density of population (establishments) are calculated as the ratio between population (number of establishments) and the area of any county. The density of female householders (owner-occupied house units) is calculated as the ratio between the number of female householders (owner-occupied house units) and the population of any county.

are row-normalized. W_{co} is constructed by the information of adjacency among counties as described in the last subsection. Counties have 5.56 adjacent neighbors on average, with a maximum of 13 adjacent neighbors and a minimum of 1 adjacent neighbor. We include counties which are within 200-miles around each county in W_{di} and the adjacent neighbors are excluded.[10] The entries in W_{di} are constructed by the inverse of squared distances. Thus, W_{co} and W_{di} together capture the geographic neighbors, while W_{co} mainly captures the spatial interdependence among closest counties which share borders and W_{di} captures spatial interdependence between relatively farther counties. W_{ec} captures industrial structure proximate neighbors and W_{po} captures political tendency proximate neighbors. We also set the distance limit $D_L = 200$ miles for them in the benchmark model. Therefore, three different kinds of channels of spatial interdependence in geographical, industrial structure and political tendency proximity, are considered simultaneously. As all the four matrices capture spatial effects within 200 miles, the spatial weight matrix M_n in spatial errors captures all neighbors' other potential spillovers within 500 miles.

Table 2 also presents descriptive statistics for independent variables, which are consistent with those employed in local growth studies (for example, Schmitt and Henry, 2000; Wheeler, 2003; Gebremariam et al., 2010). Total population, population density and the percentage of population between the ages of 25-54 control for the agglomeration effects. Unemployment rate and female householders density (defined as the ratio between the number of female householders and population) could measure counties' economic distress as a high unemployment rate reflects poor business environments and a high female householders density is associated with low median household incomes. The two indicators are also relative to weak demand in local markets. Establishment spatial density, the ratio of the number of establishment and area, is associated with severity of competition in a county. Owner-occupied house units density, which is the ratio between owner-occupied house units and population, is related to availability of financing resources and at the same time mortgage as the cost of

[10]The 200-miles setup is consistent with findings in the literature that growth of a county is correlated with its neighbors in 200 miles or that shocks in a county can propagate up to 200 miles (e.g., Wheeler, 2001; Feyrer et al., 2017; James and Smith, 2020).

owning housing units might result in deterioration of economic environment. Percentages of adults with high school diploma and college degree or above measures human capital in a county.

3.3 Estimation results

We present the estimation results for the high order SAR models with spatial error and heteroskedasticity in Table 3, employing the proposed FBGMM and traditional QML estimation methods. The simple GMM estimates with X, W_jX, W_j^2X for each $j = co, di, se, po$ as IVs and W_j, W_j^2, M, M^2 as quadratic matrices are used as the initial estimates for FBGMM. Following LeSage and Pace (2009), we calculate the average direct impact (ADI) and average indirect impact (AII) for each explanatory variable. The model is stable as $\| \sum_{j=co,di,ec,po} \hat{\lambda}_j W_{jn} \|_\infty$ is 0.5279 (FBGMM), 0.5287 (QML), or 0.5449 (GS2SLS). The FBGMM estimates has smaller standard deviation than QML and GS2SLS estimates for all λ's and most impact coefficients, which is consistent with the theoretical superiority of FBGMM in the Section 2. Specifically, in the estimation results for model (12), the FBGMM estimates' standard deviations of λ's are 14%-42% smaller than those of QML's and 57%-66% smaller than those of GS2SLS's. The reduction of standard deviations for impact estimates are up to 80%. The advantage of FBGMM can be further elucidated when the complexity of a model increases or the sample size becomes smaller as FBGMM demonstrates a more efficient exploitation of information. For instance, upon examining heterogeneous in-state and out-of-state spatial interdependence in Table S7 in the supplementary file, certain QML estimates lose their significance or experience a decrease in their level of significance.

The spatial interdependence among neighbors with industrial proximity, measured by the estimate of λ_{ec} in Table 3, is the strongest, followed by the interdependence among the neighbors that share borders. For relatively farther neighbors, their estimated spatial interdependence is only -2.47% and insignificant, i.e., one percentage point of employment growth in neighbors only results in -2.47 percentage point of reduction for a county. The spatial interdependence among neighbors with political tendency proximity is relatively small with an estimate 8.40%.

Table 3: Empirical results of model (12) for growth rates of employment

		FBGMM		QML		GS2SLS	
		$\hat{\theta}_{fbgmm}$	std_θ	$\hat{\theta}_{qml}$	std_θ	$\hat{\theta}_{gs2sls}$	std_θ
λ_{co}		0.0937***	(0.0220)	0.0936***	(0.0255)	0.0594	(0.0535)
λ_{di}		-0.0247	(0.0224)	-0.0249	(0.0281)	-0.0508	(0.0523)
λ_{se}		0.3292***	(0.0296)	0.3293***	(0.0511)	0.3586***	(0.0883)
λ_{po}		0.084***	(0.0188)	0.0846***	(0.0240)	0.0813*	(0.0442)
Constant		0.0239	(0.0350)	0.0236	(0.0358)	0.0190	(0.0401)
G_{pop}	ADI	0.6015***	(0.0239)	0.6018***	(0.0271)	0.6196***	(0.0269)
	AII	0.5502***	(0.0233)	0.5512***	(0.0861)	0.4957***	(0.1141)
$\ln(pop_{2000})$	ADI	0.0438***	(0.0062)	0.0448***	(0.0063)	0.0442***	(0.0072)
	AII	0.04***	(0.0057)	0.041***	(0.0087)	0.0354***	(0.0102)
$pop^{(d)}_{2000}$	ADI	-0.0016	(0.0011)	-0.0028	(0.0019)	-0.0028	(0.0017)
	AII	-0.0015	(0.0010)	-0.0026	(0.0018)	-0.0023	(0.0015)
$pop^{(25-54)}_{2000}$	ADI	-0.0015	(0.0012)	-0.0004	(0.0011)	-0.0005	(0.0011)
	AII	-0.0013	(0.0011)	-0.0004	(0.0011)	-0.0004	(0.0009)
ump_{2000}	ADI	-0.0011**	(0.0005)	-0.0011**	(0.0005)	-0.001**	(0.0005)
	AII	-0.001**	(0.0005)	-0.001**	(0.0005)	-0.0008**	(0.0004)
fhh_{2000}	ADI	-0.5081***	(0.1431)	-0.5081***	(0.1453)	-0.5413***	(0.1584)
	AII	-0.4648***	(0.1289)	-0.4654***	(0.1465)	-0.4331***	(0.1517)
hmw_{2000}	ADI	-0.0859*	(0.0509)	-0.0858*	(0.0516)	-0.0876	(0.0534)
	AII	-0.0786*	(0.0462)	-0.0786	(0.0479)	-0.0701	(0.0449)
mnf_{2000}	ADI	-0.1076***	(0.0159)	-0.1078***	(0.0163)	-0.1017***	(0.0206)
	AII	-0.0984***	(0.0143)	-0.0987***	(0.0184)	-0.0814***	(0.0188)
sev_{2000}	ADI	0.0093	(0.0261)	0.0096	(0.0255)	0.0140	(0.0312)
	AII	0.0085	(0.0238)	0.0088	(0.0232)	0.0112	(0.0246)
$est^{(d)}_{2000}$	ADI	0.0286	(0.0204)	0.0285	(0.0353)	0.0287	(0.0257)
	AII	0.0262	(0.0186)	0.0261	(0.0326)	0.0230	(0.0213)
col_{2000}	ADI	-0.0001	(0.0003)	0.0005*	(0.0003)	0.0005*	(0.0003)
	AII	-0.0001	(0.0003)	0.0005*	(0.0003)	0.0004	(0.0003)
hig_{2000}	ADI	0.0008***	(0.0003)	0.0005*	(0.0003)	0.0005*	(0.0003)
	AII	0.0007**	(0.0003)	0.0005*	(0.0003)	0.0004	(0.0003)
emp_{2000}	ADI	-0.0467***	(0.0060)	-0.0469***	(0.0060)	-0.0462***	(0.0067)
	AII	-0.0427***	(0.0055)	-0.0429***	(0.0088)	-0.037***	(0.0104)
ρ		0.0051	(0.0568)	0.0051	(0.0596)	0.0302	(0.0627)
Constant		0.1762***	(0.0000)	0.0874***	(0.0088)		
$\ln(pop_{2000})$		-0.0092***	(0.0001)	-0.0005	(0.0009)		
$\boldsymbol{t}_{1n}$		125.348***		300.664***			
$\boldsymbol{t}_{2n}$		9.2885***		41.2012***			

Notes: The dependent variable is the growth rate of employment in 3074 counties ($n = 3074$). We use FBGMM, QML, and GS2SLS in estimation with spatial errors and heteroskedasticity being considered. All λ's in the first part are corresponding spatial effects in model (12). The second part presents estimates for average direct effect (ADI) and average indirect effect (AII) for each X using LeSage and Pace (2009). The third part presents results of the second equation in (12) for errors: ρ is the coefficient for spatial error while constant and $\ln(pop_{2000})$ are coefficients in the heteroskedastic function for σ_{ni}. The last part presents $\boldsymbol{t}_{1n}$ and $\boldsymbol{t}_{2n}$ which are test statistics for the normality of disturbances, where $\boldsymbol{t}_{1n}$ is in (10) and $\boldsymbol{t}_{2n}$ is in (11). Both test statistics are asymptotically chi-squared distributed with 2 degrees of freedom. The numbers in parentheses are standard errors. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

The significant estimates for explanatory variables are mostly consistent with predictions in theory. The positive estimates for logged population ($\ln(pop_{2000})$) and population growth (G_{pop}) provide evidence that agglomeration economies are significant determinants for a county's employment growth. Unemployment rate (ump_{2000}) and female householders density (fhh_{2000}) in the initial year have negative effects on local employment growth. Poor economic environment provides inadequate incentives for new business establishment or business expansion and insufficient demand due to unemployment and relatively low household income could shrink local markets. Owner-occupied house units density (hmw_{2000}) has negative coefficients, which is not in favor of the hypothesis of loosening financial constraint when business owners can use houses as collateral. On the other hand, high density of owner-occupied house units might also increase economic pressure during recessions due to mortgage and deteriorated aggregate demand in local markets. The employment in the initial year and percentage of employees in the manufacturing sector are negatively related to growth, which suggests that rapid growth occurs more likely in less developed areas.

We apply our tests for the normality of disturbances and the null hypothesis that disturbances are normal are rejected. Take the FBGMM estimation as an example, the test statistic that employs the Lagrangian multiplier principle as in Jarque and Bera (1980) is 125.3 and the test statistic that uses the vector of estimated skewness and excess kurtosis coefficients is 9.3, while the critical value at 1% significance level for both tests is 9.2. Our data does not exhibits significant spatial error dependence, indicated by the last part of Table 3. We also conduct FBGMM without spatial errors in column (4) of Table 4, where the estimates for λ's do not depart much from those in Table 3.

Table 4 provides additional estimates with other estimation methods and model specifications. They show that (i) the empirical findings in the benchmark model are robust to estimation methods and restrictions on the spatial error parameter; (ii) the proposed FBGMM estimation outperforms other estimation methods in efficiency by substantially reducing standard errors;(iii) the estimation results are robust when we estimate a high order SAR model; (iv) the estimation results are robust when we take the endogeneity of spatial weight matrices W_{ec} and W_{po} in column (5) of Table 4. We adopt Qu and Lee (2015)'s control function method to handle possible endogeneity problem. We use the each counties' historical industrial structure and political tendency data - the employments in 7 sectors on 1990 and share of voting for democratic or republican candidates on 1988 and 1992 - as the instrumental variables for entries of W_{ec} and W_{po}. The estimates are almost similar in magnitudes with these in Table 3.

Table 4: Empirical results for growth rates of employment: additional specifications

	2SLS	RGMM	GMM	FBGMM with alternative M	FBGMM with endogenous W
	(1)	(2)	(3)	(4)	(5)
λ_{co}	0.0842*	0.113***	0.0936***	0.1159***	0.1183***
	(0.0507)	(0.0263)	(0.0237)	(0.0092)	(0.0250)
λ_{di}	-0.0372	0.0007	-0.0249	0.0102	-0.0219
	(0.0521)	(0.0278)	(0.0255)	(0.0071)	(0.0270)
λ_{se}	0.3268***	0.3411***	0.3293***	0.3334***	0.3609***
	(0.0837)	(0.0553)	(0.0347)	(0.0116)	(0.0488)
λ_{po}	0.0978**	0.0349	0.0846***	0.0438***	0.0847***
ρ		0.1226*	0.0051		0.0001
		(0.0642)	(0.0599)		(0.0570)
Full set of covariants	Yes	Yes	Yes	Yes	Yes
$\boldsymbol{t}_{1n}$			300.664***	111.6399***	111.6399***
$\boldsymbol{t}_{2n}$			41.2012***	48.5543***	48.5543***

Notes: The dependent variable is the growth rate of employment in 3074 counties ($n = 3074$). Column (1) presents 2SLS estimates; column (2) presents RGMM estimates proposed by Lee and Liu (2010) with X, WX, and W^2X as IVs and W, M, $W^2-\text{diag}(W^2)$ and $M^2-\text{diag}(M^2)$ as the quadratic matrices in the quadratic moments; column (3) presents GMM estimates with X, WX, and W^2X as IVs and W, M, W^2 and M^2 as the quadratic matrices in the quadratic moments; column (4) presents FBGMM estimates for a modified model of (12) which does not consider spatial errors ;column (5) presents FBGMM estimates for a modified model of (12) which allows endogenous spatial weight matrices. $\boldsymbol{t}_{1n}$ and $\boldsymbol{t}_{2n}$ which are test statistics for the normality of disturbances, where $\boldsymbol{t}_{1n}$ is in (10) and $\boldsymbol{t}_{2n}$ is in (11). Both test statistics are asymptotically chi-squared distributed with 2 degrees of freedom. The numbers in parentheses are standard errors. *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively.

4 Conclusion

This paper proposes a novel analytic procedure to derive best linear and quadratic moments for a large class of spatial econometric models with heteroskedasticity. The resulting best GMM estimator is asymptotically more efficient than the QML estimator when disturbances are non-normal. The consideration of heteroskedasticity in our models also provides a useful modeling approach for empirical studies, which can yield estimates more robust to heteroskedasticity and improve estimation efficiency relative to existing robust estimation methods. We derive the best linear and quadratic moments in detail for a high order SARAR model with heteroskedasticity. In the appendix, we derive two normality tests on the normality of disturbances for the SARAR model and derive best moments for a general high order SESAR model with MSARMA disturbances which nest many models in the literature, with details in the supplementary file.

We apply the high order SARAR model with heteroskedastic errors and the best GMM estimator to county-level census data on local employment growth in the contiguous United States. We investigate the spatial interdependence patterns and channels of regional economics growth. Estimation results show heterogeneous magnitudes of spatial interdependence among neighboring counties with geographic proximity, industrial proximity, and political tendency proximity, with the interdependence between counties with similar industrial structure being the strongest. Classic SAR models with single parameter for spatial interdependence tends to overestimate the spatial effect through the specified channel as other channels are omitted. The results are robust to model specifications and estimation methods.

The best GMM estimator for a larger class of network and spatial econometric models might be studied in future research. Furthermore, it might be difficult to search for best moments for nonlinear spatial models while they have potentials in empirical studies. In addition, the search for best moments is confined to linear and quadratic moments in this paper as it is motivated by the score of QML estimators, which can be extended.

References

Allers, M.A., Elhorst, J.P., 2011. A simultaneous equations model of fiscal policy interactions. Journal of Regional Science 51, 271–291.

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

Anselin, L., 2010. Thirty years of spatial econometrics. Papers in Regional Science 89, 3–25.

Anselin, L., Bera, A., 1998. Spatial dependence in linear regression models with an introduction to spatial econometrics, in: Ullah, A., Giles, D.E. (Eds.), Handbook of Applied Economic Statistics. Marcel Dekker, New York, pp. 237–289.

Arbia, G., 2016. Spatial econometrics: A broad view, in: Foundations and Trends® in Econometrics. volume 8, pp. 145–265.

Badinger, H., Egger, P., 2011. GM estimation of higher-order spatial autoregressive processes in cross-section models with heteroskedastic disturbances. Papers in Regional Science 90, 213–235.

Badinger, H., Egger, P., 2015. Fixed effects and random effects estimation of higher-order spatial autoregressive models with spatial autoregressive and heteroscedastic disturbances. Spatial Economic Analysis 10, 11–35.

Baltagi, B.H., Bresson, G., 2011. Maximum likelihood estimation and lagrange multiplier tests for panel seemingly unrelated regressions with spatial lag and spatial errors: An application to hedonic housing prices in Paris. Journal of Urban Economics 69, 24–42.

Baltagi, B.H., Egger, P.H., Kesina, M., 2022. Bayesian estimation of multivariate panel probits with higher-order network interdependence and an application to firms' global market participation in guangdong. Journal of Applied Econometrics 37, 1356–1378.

Baum-Snow, N., 2007. Did highways cause suburbanization? Quarterly Journal of Economics 122, 775–805.

Bera, A.K., Doğan, O., Taşpınar, S., 2021. Asymptotic variance of test statistics in the ML and QML frameworks. Journal of Statistical Theory and Practice 15, 1–26.

Bera, A.K., Doğan, O., Taşpınar, S., 2022. Distribution of test statistics under parameter uncertainty for time series data: an application to testing skewness, kurtosis and normality. Hacettepe Journal of Mathematics and Statistics 51, 253–272.

Bettencourt, L.M.A., 2013. The origins of scaling in cities. Science 340, 1438–1441.

Breusch, T., Qian, H., Schmidt, P., Wyhowski, D., 1999. Redundancy of moment conditions. Journal of Econometrics 91, 89–111.

Burchfield, M., Overman, H.G., Puga, D., Turner, M.A., 2006. Causes of sprawl: A portrait from space. Quarterly Journal of Economics 121, 587–633.

Cliff, A., Ord, J.K., 1973. Spatial Autocorrelation. Pion, London.

Combes, P.P., 2000. Economic structure and local growth: France, 1984–1993. Journal of Urban Economics 47, 329–355.

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.

Doğan, O., Taşpınar, S., 2013. GMM estimation of spatial autoregressive models with moving average disturbances. Regional Science and Urban Economics 43, 903–926.

Duranton, G., Turner, M.A., 2012. Urban growth and transportation. Review of Economic Studies 79, 1407–1440.

Elhorst, J.P., Emili, S., 2022. A spatial econometric multivariate model of Okun's law. Regional Science and Urban Economics 93, 103756.

Ertur, C., Koch, W., 2007. Growth, technological interdependence and spatial externalities: Theory and evidence. Journal of Applied Econometrics 22, 1033–1062.

Feyrer, J., Mansur, E.T., Sacerdote, B., 2017. Geographic dispersion of economic shocks: Evidence from the fracking revolution. American Economic Review 107, 1313–1334.

Fingleton, B., 2008. A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors. Spatial Economic Analysis 3, 27–44.

Fingleton, B., 2023. Estimating dynamic spatial panel data models with endogenous regressors using synthetic instruments. Journal of Geographical Systems 25, 121–152.

Gebremariam, G.H., Gebremedhin, T.G., Schaeffer, P.V., 2010. Analysis of county employment and income growth in Appalachia: A spatial simultaneous-equations approach. Empirical Economics 38, 23–45.

de Graaff, T., van Oort, F.G., Raymond J.G.M. Florax, Frank G. van Oort, R.J.F., 2012. Regional population-employment dynamics across different sectors of the economy. Journal of Regional Science 52, 60–84.

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

Haining, R.P., 1978. The moving average model for spatial interaction. Transactions of the Institute of British Geographers 3, 202–225.

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

Hauptmeier, S., Mittermaier, F., Rincke, J., 2012. Fiscal competition over taxes and public inputs. Regional Science and Urban Economics 42, 407–419.

James, A.G., Smith, B., 2020. Geographic dispersion of economic shocks: Evidence from the fracking revolution: Comment. American Economic Review 110, 1905–1913.

Jarque, C.M., Bera, A.K., 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics letters 6, 255–259.

Jeanty, P.W., Partridge, M., Irwin, E., 2010. Estimation of a spatial simultaneous equation model of population migration and housing price dynamics. Regional Science and Urban Economics 40, 343–352.

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., 2004. Estimation of simultaneous systems of spatially interrelated cross sectional equations. Journal of Econometrics 118, 27–50.

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.

Kelejian, H.H., Robinson, D.P., 1993. A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. Papers in Regional Science 72, 297–312.

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.

LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics. Chapman & Hall/CRC.

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., 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.

Liu, X., Saraiva, P., 2019. GMM estimation of spatial autoregressive models in a system of simultaneous equations with heteroskedasticity. Econometric Reviews 38, 359–385.

Paelinck, J.H.P., Klaassen, L.H., 1979. Spatial Econometrics. Saxon House, Farnborough.

Peters, M., 2022. Market size and spatial growth-evidence from germany's post-war population expulsions. Econometrica 90, 2357–2396.

Qu, X., Lee, L.F., 2015. Estimating a spatial autoregressive model with an endogenous spatial weight matrix. Journal of Econometrics 184, 209–232.

Reia, S.M., Rao, P.S.C., Barthelemy, M., Ukkusuri, S.V., 2022. Spatial structure of city population growth. Nature Communications 13.

Schmitt, B., Henry, M.S., 2000. Size and growth of urban centers in french labor market areas: consequences for rural population and employment. Regional Science and Urban Economics 30, 1–21.

Taşpınar, S., Doğan, O., Bera, A.K., 2019. Heteroskedasticity-consistent covariance matrix estimators for spatial autoregressive models. Spatial Economic Analysis 14, 241–268.

Wheeler, C.H., 2001. A note on the spatial correlation structure of county-level growth in the U.S. Journal of Regional Science 41, 433–449.

Wheeler, C.H., 2003. Evidence on agglomeration economies, diseconomies, and growth. Journal of Applied Econometrics 18, 79–104.

Whittle, P., 1954. On stationary processes in the plane. Biometrika 41, 434–449.

Yang, K., Lee, L.F., 2017. Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models. Journal of Econometrics 196, 196–214.

Appendix A Efficiency implications of using quadratic matrices with zero traces

For our SARAR model with heteroskedasticity, we estimate the unknown parameter γ_0 in the assumed function $f(x_{n,i1}, \cdots, x_{n,il_x}, \gamma_0)$ for heteroskedasticity together with $[\lambda', \rho', \beta']'$, while Lee (2007) considers the homoskedastic case and does not estimate the variance parameter together with other parameters. In the homoskedastic case of Lee (2007) where the variance parameter is not estimated together with other parameters, P_{rn}'s can be required to have zero traces without losing any efficiency (Liu et al., 2010), but doing so in our case can lead to an efficiency loss. We investigate this problem in this section. We shall see that a quadratic moment $[V_n'(\theta)V_n(\theta) - n]$ corresponding to the variance of v_i can help improve the estimation efficiency in the heteroskedastic case, while it only contributes to the estimation of a joint variance parameter in the homoskedastic case. Even for the homoskedastic case, a special case with $f(\cdot)$ being a constant, we see from the main text that not requiring P_{rn}'s to have zero traces simplifies the analysis on searching for the best moments, since there is no need to make sure that the quadratic matrices for the best moments have zero traces.

Note that the derivative $\frac{\partial \ln L_n(\theta_0)}{\partial \rho_j}$ of the quasi log likelihood function $\ln L_n(\theta)$ in (12) of the main text for the SARAR can be written as

$$V_n'\Big[K_{jn} - \frac{1}{n}I_n \operatorname{tr}(K_{jn})\Big]V_n + \frac{1}{n}(V_n'V_n - n)\operatorname{tr}(K_{jn}),$$

of which the first term has a quadratic matrix with a zero trace, but the second term has a quadratic matrix equal to the identity matrix with a nonzero trace, which corresponds to the variance of v_i. The derivative $\frac{\partial \ln L_n(\theta_0)}{\partial \lambda_j}$ can be similarly decomposed. Each $\frac{\partial \ln L_n(\theta_0)}{\partial \gamma_j}$ is quadratic in V_n, where the quadratic matrix L_{jn} is diagonal but generally not proportional to an identity matrix. Thus restricting P_{rn}'s to have zero traces would lead to an efficiency loss. In particular, the case with $\Sigma_n^{1/2}(\gamma) = \gamma I_n$ for a scalar γ corresponds to the homoskedastic case and $\frac{\partial \ln L_n(\theta)}{\partial \gamma}$ reduces to $\frac{1}{\gamma}[V_n'(\theta)V_n(\theta) - n]$. In this case, the moment $V_n'(\theta)V_n(\theta) - n$ only contributes to the estimation of γ and P_{rn}'s can be restricted to have zero traces without losing efficiency for the estimation of other parameters.