GEL estimation and tests of spatial autoregressive models

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

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

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

July 16, 2018

Abstract

This paper considers the generalized empirical likelihood (GEL) estimation and tests of high order spatial autoregressive (SAR) models by exploring an inherent martingale structure. The GEL estimator has the same asymptotic distribution as the generalized method of moments estimator explored with same moment conditions for estimation, but circumvents a first step estimation of the optimal weighting matrix with a preliminary estimator, and thus can be robust to unknown heteroskedasticity and non-normality. While the GEL removes the asymptotic bias from the preliminary estimator and partially removes the bias due to the correlation between the moment conditions and their Jacobian, the empirical likelihood as a special member of GELs further partially removes the bias from estimating the second moment matrix. We also formulate the GEL overidentification test, Moran's I test, and GEL ratio tests for parameter restrictions and non-nested hypotheses. While some of the conventional tests might not be robust to non-normality and/or unknown heteroskedasticity, the corresponding GEL tests can be robust.

Keywords: Spatial autoregressive model, empirical likelihood, higher order asymptotic bias, unknown heteroskedasticity, non-normality, EL ratio test

JEL classification: C12, C13, C14, C21, C52

1 Introduction

In this paper, we consider empirical likelihood (EL) and generalized EL (GEL) estimation and tests of popular spatial autoregressive (SAR) models with spatially dependent data. The EL approach is introduced in Owen (1991) for independent sample observations. It can be interpreted as a nonparametric maximum likelihood and a generalized minimum contrast estimation method (Kitamura, 2007).[1] The class of GEL estimators includes the EL,

*Corresponding author. Tel.: +1 614 292 5508; fax: +1 614 292 4192. E-mail addresses: jin.fei@sufe.edu.cn (F. Jin), lee.1777@osu.edu (L.-F. Lee).

[1]Helpful reviews include, among others, Hall and La Scala (1990), Owen (2001), Kitamura (2007) and Chen and Keilegom (2009).

the exponential tilting (ET) of Kitamura and Stutzer (1997) and Imbens et al. (1998), and the continuous updating generalized method of moments (GMM) of Hansen et al. (1996). With independent sample observations, the EL and GEL can have various advantages over other methods as shown in the literature. They can be robust against distributional assumptions but may still have good properties analogous to the parametric likelihood approach in estimation and testing. As alternatives to the two-step optimal GMM estimation which usually requires a first step estimation of an optimal weighting matrix with a preliminary estimator, the EL and GEL estimators are one-step estimators. They are consistent and have the same asymptotic distribution as the two-step optimal GMM estimator by using same moment conditions, but invariant to parameter-dependent linear transformations of moment conditions, and have improved high order properties (Imbens et al., 1998; Owen, 2001; Newey and Smith, 2004). In particular, Newey and Smith (2004) show that, for i.i.d. data, the GEL estimator has no asymptotic bias from estimation of the Jacobian or the preliminary estimator, and the EL further removes a bias component from estimation of the second moment matrix. In finite samples, while the two-step optimal GMM can have large bias (e.g., Altonji and Segal, 1996), the GEL estimators are observed to perform better than the GMM estimator (e.g., Hansen et al., 1996; Imbens, 1997; Ramalho, 2002; Mittelhammer et al., 2005; Newey et al., 2005). The EL and GEL can also be applied to testing problems. As a nonparametric analog of the parametric likelihood ratio statistic, a GEL ratio statistic follows an asymptotic chi-squared distribution under the null. An EL ratio test and confidence region are often Bartlett correctable (Corcoran, 1998; DiCiccio et al., 1991; Lazar and Mykland, 1999), and EL tests are Bahadur efficient (Otsu, 2010) and have optimality properties in terms of large deviations (Kitamura, 2001).

The EL and GEL have originally been considered for independent data. Later on, there are attempts to generalize them for time series data (e.g., Kitamura, 1997). For time series, some authors have studied the EL for models with martingale structures. Mykland (1995) generalizes the EL definition for i.i.d. data to models with martingale structures and introduces the concept of dual likelihood, and Chuang and Chan (2002) develop the EL for autoregressive models with innovations that form a martingale difference sequence. The EL has also been considered for regularly (e.g., Nordman, 2008a,b) and irregularly (e.g., Bandyopadhyay et al., 2015; Van Hala et al., 2015) spaced spatial data. In the spatial econometric literature, Marsh and Mittelhammer (2004) and Perevodchikov et al. (2012) investigate some GEL estimation methods for the SAR model, Fernández-Vázquez et al. (2009) propose to use the ET to estimate spatial weights and other parameters in the SAR model, and Kostov (2013) considers the EL estimation of a spatial quantile regression.

However, existing research on the GEL approach for spatial econometric models is limited in the following ways: First, only linear moments are used in estimation. As a result, the GEL approach is similar to that for i.i.d. data in its setting, even though dependent variables are spatially dependent. However, quadratic moments that capture spatial dependence might lead to significant efficiency gain (Lee, 2007). Both linear and quadratic moments are fundamental statistics in likelihood estimation of an SAR model. In the extreme case of pure SAR models with no exogenous variables, which includes spatially dependent disturbances in a regression equation, no instrumental variables (IV) are available and existing GEL procedures are not directly applicable. Second, even though sample observations are spatially dependent, the GEL estimation procedures are described but no theoretical justification has been provided in the existing econometric literature. Third, only first order SAR models are considered. These motivate our investigation of the GEL approach for SAR models. We focus our study on the GEL estimation and tests of a general high order SAR model with SAR disturbances (SARAR model) using both linear and quadratic moments. Such a model includes the SAR model with a single spatial lag, and the regression model with spatially dependent disturbances.[2] Quadratic moments for SARAR models, motivated from the quasi-maximum likelihood (QML) estimation and Moran's I test, are quadratic forms of the whole disturbance vector and thus involve all observations. For the GEL estimation, a key step is to write them as sums of martingale differences.[3] Treating each martingale difference as if it were a data observation, we can set up the GEL objective function to derive corresponding estimates and relevant test statistics. We shall investigate the asymptotic properties of the GEL estimator, its higher order asymptotic bias, and various tests in the GEL framework.[4]

We show that, for spatial data, the GEL estimation with moment conditions can remove the asymptotic bias from the preliminary estimator and partially remove the asymptotic bias due to the correlation between moment conditions and their Jacobian. The EL further partially removes the bias from estimation of the second moment matrix. This conclusion is consistent with that in Anatolyev (2005) for stationary time series models with non-i.i.d. but serially uncorrelated data. In the event that only linear moments are used, the EL has the ability to completely remove the asymptotic bias from estimation of the second moment matrix.

We also consider test statistics in the GEL framework. The GEL objective function (with proper normalization) evaluated at the GEL estimator is an overidentification test statistic that can be used to test for validity of moment conditions. Tests of parameter restrictions can be conveniently implemented with GEL ratio statistics. The popular Moran's I test for spatial dependence formulated with a GEL ratio is robust to unknown heteroskedasticity. In addition, we employ the GEL ratio statistic to construct a spatial J test for competing SARAR models (Kelejian, 2008; Kelejian and Piras, 2011). Unlike original spatial J tests based on the two-stage least squares (2SLS) or generalized spatial 2SLS (GS2SLS) estimation (Kelejian and Prucha, 1998), the spatial J test with a GEL ratio conveniently employs quadratic moments in addition to linear ones to obtain more efficient estimators for testing. These tests do not involve estimation of variances and are robust to unknown heteroskedasticity. For testing with quadratic moments, GEL tests are also robust to non-normality in the sense that (higher order) moment parameters do not need to be evaluated.

Various estimation methods for the SARAR model, which includes the SAR and spatial error (SE) models

[2]We appreciate suggestions from an Associate Editor to devote our attention to the high order SARAR model for its generality rather than the first order one.

[3]In the time series literature, quadratic statistics have long been written as martingales (See, e.g., Hall and Heyde, 1980). The importance of martingale processes for spatial random variables has been recognized by Kelejian and Prucha (2001). They develop a central limit theorem for a general linear-quadratic form of independent disturbances by exploring its martingale structure.

[4]Although we only focus on the SARAR model, by exploring martingale structures, estimation of other spatial econometric models including nonlinear ones may be possibly studied by an extension of kernel smoothing of moment conditions motivated by Kitamura and Stutzer (1997) and Smith (1997). But that will be investigated in a future study.

as special cases, have been proposed in the literature, e.g., the GS2SLS estimation (Kelejian and Prucha, 1998), the QML estimation (Lee, 2004), and the GMM estimation (Lee, 2007; Liu et al., 2010; Lee and Liu, 2010).[5] The GS2SLS estimates the equation by the 2SLS, thus it is computationally simple, but can be asymptotically inefficient compared to the QML. Although being relatively efficient, the QML may be computationally intensive for large sample sizes, especially for high order SARAR models. The GMM, which may employ not only linear moments in disturbances but also quadratic ones, can be computationally simpler than the QML and asymptotically as efficient as or efficient relative to the QML.[6] In the presence of unknown heteroskedasticity, by selecting quadratic matrices with zero diagonals, the GMM can yield robust estimates (Kelejian and Prucha, 2010; Lin and Lee, 2010). Liu and Yang (2015) propose to modify the QML scores to obtain estimators robust to unknown heteroskedasticity.

This paper is organized as follows. Section 2 introduces a general high order SARAR model, and the GEL and GMM estimation in both homoskedastic and heteroskedastic cases based on its martingale structure. Section 3 shows the consistency and asymptotic normality of the GEL estimator and compares its asymptotic bias with that of the GMM estimator. Section 4 investigates test statistics in the GEL framework. Section 5 reports some Monte Carlo results, which demonstrate that GEL estimators and test statistics have desirable finite sample performance. Section 6 concludes. Lemmas and proofs of theorems are collected in appendices.[7]

2 The SARAR model and GEL estimation

Consider the SARAR model with p-order spatial lags and q-order spatial errors (for short, SARAR(p,q)):

$$Y_n = \sum_{j=1}^{p} \kappa_j W_{jn} Y_n + X_n \beta + U_n, \quad U_n = \sum_{k=1}^{q} \tau_k M_{kn} U_n + V_n, \tag{1}$$

where n is the sample size, Y_n is an $n \times 1$ vector of observations on the dependent variable, X_n is an $n \times k_x$ matrix of exogenous variables with coefficient vector β, W_{jn} and M_{kn} are $n \times n$ nonstochastic spatial weights matrices with zero diagonals, κ_j and τ_k are scalar spatial dependence parameters, and $V_n = [v_{ni}]$ is an $n \times 1$ vector of independent disturbances with mean zero and finite variances. In this paper, we consider both the homoskedastic case that v_{ni}'s have the same variance and the heteroskedastic case that v_{ni}'s have different variances with unknown form. Let I_n be the $n \times n$ identity matrix, $\kappa = (\kappa_1, \ldots, \kappa_p)'$, $\tau = (\tau_1, \ldots, \tau_q)'$, $S_n(\kappa) = I_n - \sum_{j=1}^{p} \kappa_j W_{jn}$, $R_n(\tau) = I_n - \sum_{k=1}^{q} \tau_k M_{kn}$, and $(\kappa_0', \tau_0', \beta_0')$ be the true value of (κ', τ', β'). As an equilibrium model, Y_n has the reduced form $Y_n = S_n^{-1}(X_n\beta_0 + R_n^{-1}V_n)$, where $S_n = S_n(\kappa_0)$ and $R_n = R_n(\tau_0)$ are assumed to be invertible. The X_n is assumed to be nonstochastic for convenience, as in Kelejian and Prucha (1998) and Lee (2004).[8]

If the disturbances v_{ni}'s in model (1) are i.i.d. with mean 0 and variance σ_0^2, the moment vector for a GMM

[5] Due to endogeneity of the spatial lag in an SAR model, the least squares estimator is only consistent in certain cases (Lee, 2002).

[6] The GMM estimator with properly chosen moments can be as efficient as the QML estimator for the SARAR model with normal disturbances, but it can be more efficient than the latter in the case of non-normal disturbances (Liu et al., 2010; Lee and Liu, 2010).

[7] Proofs of lemmas and some other material are provided in an online supplementary file available at: `http://econ.shufe.edu.cn/kindeditor-4.1.10/attached/file/20180716/20180716231318_52942.pdf`.

[8] Alternatively, X_n can be stochastic with finite moments of certain order.

estimation can be

$$g_n(\theta) = \frac{1}{n}[V_n'(\theta)P_{1n}V_n(\theta) - \sigma^2 \operatorname{tr}(P_{1n}), \ldots, V_n'(\theta)P_{k_pn}V_n(\theta) - \sigma^2 \operatorname{tr}(P_{k_pn}), V_n'(\theta)Q_n]', \tag{2}$$

where $V_n(\theta) = R_n(\tau)[S_n(\kappa)Y_n - X_n\beta]$, with $\theta = (\kappa', \tau', \beta', \sigma^2)'$ being a k_θ-dimensional vector for $k_\theta = k_x + p + q + 1$, P_{ln} for $l = 1, \ldots, k_p$ are $n \times n$ nonstochastic matrices, and Q_n is an $n \times k_q$ IV matrix with full column rank k_q. Without loss of generality, assume that P_{ln}, for $l = 1, \ldots, k_p$, are symmetric and linearly independent.[9] The quadratic moments are valid since $\mathrm{E}(V_n'P_{ln}V_n) = \sigma_0^2 \operatorname{tr}(P_{ln})$. The IV matrix Q_n may consist of independent columns of $X_n, W_{jn}X_n$ and so on, and P_{ln}'s can be functions of W_{jn} and M_{kn} such as $W_{jn}^{(s)}$, $M_{kn}^{(s)}$, $(W_{jn}^2)^{(s)}$, $(M_{kn}^2)^{(s)}$, $(W_{jn}W_{kn}')^{(s)}$ and $(M_{jn}M_{kn}')^{(s)}$, where $A^{(s)} = A + A'$ for any square matrix A. The total number of moments in $g_n(\theta)$ is $k_g = k_p + k_q$, which is greater than or equal to k_θ.

A quadratic form has in general a double summation. However, it can also be written in a single summation with the remaining summation written into partial sums: $V_n'(\theta)P_{ln}V_n(\theta) - \sigma^2 \operatorname{tr}(P_{ln}) = \sum_{i=1}^n \omega_{ln,i}(\theta)$ for $l = 1, \ldots, k_p$, where

$$\omega_{ln,i}(\theta) = p_{ln,ii}[v_{ni}^2(\theta) - \sigma^2] + 2v_{ni}(\theta)\sum_{j=1}^{i-1} p_{ln,ij}v_{nj}(\theta) \tag{3}$$

with $v_{nj}(\theta)$ being the jth element of $V_n(\theta)$, and

$$g_{ni}(\theta) = [\omega_{1n,i}(\theta), \ldots, \omega_{k_pn,i}(\theta), Q_{ni}'v_{ni}(\theta)]', \tag{4}$$

where Q_{ni} contains k_q IVs of the ith unit, i.e., $Q_n = [Q_{n1}, \ldots, Q_{nn}]'$. Then $g_n(\theta) = \frac{1}{n}\sum_{i=1}^n g_{ni}(\theta)$. The quadratic moments involve the variance parameter σ^2 due to (3) in order that $g_n(\theta)$ can be decomposed into a sum of $g_{ni}(\theta)$'s in (4), where $g_{ni}(\theta_0)$ for $i = 1, \ldots, n$, are martingale differences (Kelejian and Prucha, 2001).[10] Thus the variance of $g_n(\theta_0)$ is $\frac{1}{n^2}\sum_{i=1}^n \mathrm{E}[g_{ni}(\theta_0)g_{ni}'(\theta_0)]$. Our quadratic moments involving the estimation of σ^2 are in line with those in Kelejian and Prucha (1998, 1999).[11]

In the case that there is unknown heteroskedasticity, we may select all P_{ln}'s to have zero diagonals in order to derive valid moment conditions, as in Kelejian and Prucha (2010) and Lin and Lee (2010). Such P_{ln}'s can be $W_{jn}^{(s)}$, $M_{kn}^{(s)}$, $(W_{jn}^2)^{(s)} - 2\operatorname{diag}(W_{jn}^2)$, $(M_{kn}^2)^{(s)} - 2\operatorname{diag}(M_{kn}^2)$, $(W_{jn}W_{kn}')^{(s)} - 2\operatorname{diag}(W_{jn}W_{kn}')$, $(M_{jn}M_{kn}')^{(s)} - 2\operatorname{diag}(M_{jn}M_{kn}')$ and so on, where $\operatorname{diag}(A)$ for a square matrix A denotes a diagonal matrix formed by the diagonal elements of A. Let the moment vector be

$$g_n(\theta) = \frac{1}{n}[V_n'(\theta)P_{1n}V_n(\theta), \ldots, V_n'(\theta)P_{k_pn}V_n(\theta), V_n'(\theta)Q_n]', \tag{5}$$

[9] If P_{ln} is not symmetric, replacing it with $(P_{ln} + P_{ln}')/2$ does not change the value of the moment vector.

[10] This is a technical issue in the GEL estimation framework. For that reason, it is not possible to unify the GEL estimation for the homoskedastic and heteroskedastic cases into a single estimation framework, unless we restrict P_{jn}'s to have zero diagonals in the homoskedastic case. Furthermore, the variance σ^2 can be estimated in the above GEL framework, while it is not meaningful to be estimated in a heteroskedastic case.

[11] Lee (2001, 2007) use quadratic moments of the form $\mathrm{E}(V_n'P_nV_n) = 0$ with $\operatorname{tr}(P_n) = 0$ to formulate the GMM estimation, which do not involve σ^2. The quadratic moments involving no σ^2 in Lee (2001, 2007) would not be technically appropriate to be used here due to the required martingale difference property. As shown in Lee (2001, 2007), in the case that v_{ni}'s are normal, there are moment vectors with which the resulting GMM estimators are as efficient as the maximum likelihood estimator.

where P_{ln}'s have zero diagonals, and $V_n(\theta)$ is the same as above, but $\theta = (\kappa', \tau', \beta')'$ would not contain σ^2 so that $k_\theta = k_x + p + q$ is the dimension of θ.[12] Then $\omega_{ln,i}(\theta)$ and $g_{ni}(\theta)$ can still have the forms in (3) and (4), except that the first term of $\omega_{ln,i}(\theta)$ on the r.h.s. of (3) is zero.

We consider the GEL estimator:

$$\hat{\theta}_{n,\mathrm{GEL}} = \arg\min_{\theta\in\Theta}\sup_{\lambda\in\Lambda_n(\theta)} \sum_{i=1}^{n} \rho(\lambda' g_{ni}(\theta)), \tag{6}$$

where $\Lambda_n(\theta) = \{\lambda : \lambda' g_{ni}(\theta) \in \mathcal{V}, i = 1, \ldots, n\}$ for an open interval $\mathcal{V}$ containing 0, and $\rho(v)$ is a twice continuously differentiable concave function of a scalar v on $\mathcal{V}$.[13] Denote $\rho_k(v) = \frac{d^k \rho(v)}{dv^k}$ and $\rho_k = \rho_k(0)$ for $k = 1$ and 2. As long as $\rho_1 \neq 0$ and $\rho_2 < 0$, without loss of generality, we may let $\rho_1 = \rho_2 = -1$ (Newey and Smith, 2004). The EL is a special case of the GEL with $\rho(v) = \ln(1 - v)$ for $v < 1$ (Qin and Lawless, 1994; Smith, 1997); the ET is a special case with $\rho(v) = -e^v$ (Kitamura and Stutzer, 1997; Smith, 1997); and the continuous updating GMM is a special case with a quadratic $\rho(v) = -\frac{1}{2}(v + 1)^2$ (Newey and Smith, 2004).

To study large sample properties of the GEL estimator, we assume formally the following regularity conditions.

Assumption 1. *Either (i) v_{ni}'s are i.i.d. with mean zero, variance σ_0^2 and $\mathrm{E}(|v_{ni}|^{4+\iota}) < \infty$ for some $\iota > 0$; or (ii) v_{ni}'s are independent with mean zero and variances σ_{ni}^2's, and $\sup_n \sup_{1\leq i\leq n} \mathrm{E}(|v_{ni}|^{4+\iota}) < \infty$.*

Assumption 2. *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.*

Assumption 3. *(i) W_{jn} and M_{kn} have zero diagonals for $j = 1, \ldots, p$ and $k = 1, \ldots, q$;[14] (ii) S_n and R_n are nonsingular; (iii) the sequences of matrices $\{S_n^{-1}\}$, $\{R_n^{-1}\}$, $\{W_{1n}\}$, ..., $\{W_{pn}\}$, $\{M_{1n}\}$, ..., $\{M_{qn}\}$ are bounded in both row and column sum norms.*

Assumption 4. *(i) The sequences of matrices $\{P_{1n}\}$, ..., $\{P_{k_p n}\}$ are bounded in both row and column sum norms, and the elements of Q_n are uniformly bounded constants; (ii) P_{jn} for $j = 1, \ldots, k_p$ have zero diagonals if v_{ni}'s are heteroskedastic.*

Assumption 5. *θ_0 in a compact parameter space $\Theta \subset \mathbb{R}^{k_\theta}$ is the unique solution to $\lim_{n\to\infty} \frac{1}{n} \mathrm{E}[g_n(\theta)] = 0$.*

Assumption 6. *$\rho(v)$ is concave on $\mathcal{V}$, twice continuously differentiable in a neighborhood of zero, and $\rho_1 = \rho_2 = -1$.*

We shall consider both homoskedastic and heteroskedastic cases, so Assumption 1 gives general conditions to allow both cases. Assumptions 1(i) and 2–4(i) are the same as Assumptions 1–4 in Lee and Liu (2010). Assumption 1(ii) for the heteroskedastic case is the same as that in Lin and Lee (2010). The existence of moments higher

[12] This is proper because a single σ^2 would not be meaningful with heteroskedastic errors.

[13] In practice, λ can be chosen from $\mathbb{R}^{k_g}$. If there is some θ such that, for any λ, there exists some i such that $\lambda' g_{ni}(\theta)$ falls out of the domain of $\rho(\cdot)$, it is theoretically appropriate to set the GEL objective function at θ to infinity. If not, but $\lambda' g_{ni}(\theta)$ falls out of the domain of $\rho(\cdot)$ for some i and λ, then this λ is not the solution of the problem. This is because $\hat{\lambda}_n = O_p(n^{-1/2})$ by Theorem 3.1, and with probability approaching one, $\lambda' g_{ni}(\theta) \in \mathcal{V}$ for all $1 \leq i \leq n$, $\theta \in \Theta$ and $\|\lambda\| \leq n^{-\zeta}$, where ζ is a positive number, by Lemma 10.

[14] In this paper, whenever we mention a spatial weights matrix, it is implicit that it has a zero diagonal.

than the fourth order in Assumption 1 is needed for the application of the central limit theorem on linear-quadratic forms as in Kelejian and Prucha (2001). In Assumption 2, explanatory variables are assumed to be constants and uniformly bounded for convenience and multicollinearity is ruled out. Assumption 3(i) is a normalization often assumed in order to provide meaningful interpretation on interaction effects in the literature. Assumption 3(ii) guarantees the existence of an equilibrium of the model. Assumption 3(iii), originated in Kelejian and Prucha (1998, 1999), restricts the degree of spatial dependence to be manageable. As P_{ln}'s and Q_n are functions of W_{jn}'s, M_{kn}'s and X_n, it is reasonable to maintain Assumption 4(i). Assumption 4(ii) is needed in the heteroskedastic case so that the moments are valid. Compactness of parameter spaces in Assumption 5 is a standard assumption on extremum estimation. A high level identification condition is maintained in Assumption 5 for simplicity. Low level conditions can be derived as in Lee and Liu (2010), which are discussed in the supplementary file. These conditions are full rank conditions on the IV matrix, the quadratic matrices P_{ln}'s and the spatial weights matrices.[15] Assumption 6 is a smoothness condition on $\rho(\cdot)$ as in Newey and Smith (2004).

It is of interest to compare asymptotic properties of the GEL and GMM. Let $\Omega_n(\theta) = \frac{1}{n}\sum_{i=1}^n g_{ni}(\theta)g'_{ni}(\theta)$, then $\mathrm{var}[\sqrt{n}g_n(\theta_0)] = \mathrm{E}[\Omega_n(\theta_0)] \equiv \bar{\Omega}_n$. An estimate of $\bar{\Omega}_n$ is given by $\Omega_n(\tilde{\theta}_n)$ with some initial consistent estimator $\tilde{\theta}_n$. With $\Omega_n(\tilde{\theta}_n)$, we consider the following feasible optimal GMM (FOGMM) estimator:

$$\hat{\theta}_{n,\mathrm{GMM}} = \arg\min\nolimits_{\theta\in\Theta} g'_n(\theta)\Omega_n^{-1}(\tilde{\theta}_n)g_n(\theta). \tag{7}$$

We shall compare this FOGMM estimator with the GEL estimator. For these estimators, $\bar{\Omega}_n$ is required to be nonsingular in the limit. For any square matrix A, let $\mathrm{vec}(A)$ be the vectorization of A and $\mathrm{vec}_{\mathrm{D}}(A)$ be the column vector formed by the diagonal elements of A. In the homoskedastic case, let μ_j be the jth moment of v_{ni} for $j = 3, 4$, $P^*_{jn} = \sqrt{\mu_4 - \sigma_0^4 - \frac{\mu_3^2}{\sigma_0^2}}\,\mathrm{diag}(P_{jn}) + \sqrt{2}\sigma_0^2(P_{jn} - \mathrm{diag}(P_{jn}))$, $\Xi_n = [\mathrm{vec}(P^*_{1n}), \ldots, \mathrm{vec}(P^*_{k_pn})]$ and $\Xi_{nd} = [\mathrm{vec}_{\mathrm{D}}(P_{1n}), \ldots, \mathrm{vec}_{\mathrm{D}}(P_{k_pn})]$.[16] Then, as in Debarsy et al. (2015),

$$\bar{\Omega}_n = \frac{1}{n}\begin{pmatrix} \Xi'_n\Xi_n & 0 \\ 0 & 0 \end{pmatrix} + \big[\frac{\mu_3}{\sigma_0}\Xi_{nd}, \sigma_0 Q_n\big]'\big[\frac{\mu_3}{\sigma_0}\Xi_{nd}, \sigma_0 Q_n\big].$$

In the heteroskedastic case, let Π_n be the diagonal matrix formed by σ_{ni}^2's, and

$$\Xi^*_n = \sqrt{2}[\mathrm{vec}(\Pi_n^{1/2}P_{1n}\Pi_n^{1/2}), \ldots, \mathrm{vec}(\Pi_n^{1/2}P_{k_pn}\Pi_n^{1/2})].$$

Then $\bar{\Omega}_n = \frac{1}{n}\begin{pmatrix} \Xi^{*\prime}_n\Xi^*_n & 0 \\ 0 & Q'_n\Pi_n Q_n \end{pmatrix}$. The following assumption provides sufficient conditions for the nonsingularity of $\bar{\Omega}_n$ in the limit.

[15] As pointed out by an anonymous referee, the true parameters may not be uniquely identified in the just-identification case via the moment condition alone. However, they are generally uniquely identified in the over-identification case. For the first order SAR model, even though the true parameters are not uniquely identified in the just-identification case via the moment condition alone, one closed-form root of the sample moment equation can be shown to be consistent (Jin and Lee, 2012).

[16] By the Cauchy-Schwarz inequality, $\mathrm{E}[(v_{ni}^2 - \sigma_{ni}^2)^2]\,\mathrm{E}(v_{ni}^2) \geq [E(v_{ni}^3 - \sigma_{ni}^2 v_{ni})]^2$, i.e., $(\mu_4 - \sigma_0^4)\sigma_0^2 \geq \mu_3^2$. Thus, $\mu_4 - \sigma_0^4 - \frac{\mu_3^2}{\sigma_0^2} \geq 0$.

Assumption 7. *In the homoskedastic case, either (i)* $\lim_{n\to\infty}\frac{1}{n}Q_n'Q_n$ *and* $\lim_{n\to\infty}\frac{1}{n}\Xi_n'\Xi_n$ *have full rank, or (ii)* $\lim_{n\to\infty}\frac{1}{n}[\frac{\mu_3}{\sigma_0}\Xi_{nd},\sigma_0 Q_n]'[\frac{\mu_3}{\sigma_0}\Xi_{nd},\sigma_0 Q_n]$ *has full rank. Furthermore, in the heteroskedastic case,* $\lim_{n\to\infty}\frac{1}{n}Q_n'\Pi_n Q_n$ *and* $\lim_{n\to\infty}\frac{1}{n}\Xi_n^{*\prime}\Xi_n^*$ *have full rank.*

In the homoskedastic case, the condition $\lim_{n\to\infty}\frac{1}{n}\Xi_n'\Xi_n$ requires P_{jn}'s to be linearly independent for large enough n. When $\mu_3=0$, e.g., v_{ni}'s are normal, the condition in (ii) of Assumption 7 will not hold and will not be needed. In the heteroskedastic case, $\bar{\Omega}_n$ is block diagonal and does not involve third and fourth moments of v_{ni}'s, as P_{jn}'s have zero diagonals. The conditions are similar to those in the homoskedastic case and P_{jn}'s are required to be linearly independent for large enough n.

The initial estimator $\tilde{\theta}_n$ for the FOGMM can be derived from $\min_{\theta\in\Theta} g_n'(\theta)\hat{J}_n^{-1}g_n(\theta)$, where $\hat{J}_n$ is a $k_g\times k_g$ weighting matrix. Following Newey and Smith (2004), we assume that $\hat{J}_n$ satisfies the following assumption.

Assumption 8. $\hat{J}_n=\bar{J}_n+n^{-1/2}\xi_n^J+O_p(n^{-1})$*, where* $\lim_{n\to\infty}\bar{J}_n$ *is positive definite,* $\xi_n^J=O_p(1)$ *and* $\mathrm{E}(\xi_n^J)=0$.

This statistical property on the estimated weighting matrix provides a rigorous setting for subsequent analysis on higher order asymptotic properties of the FOGMM estimator.

3 Large sample properties of estimators

In this section, we investigate the consistency and asymptotic normality of the GEL estimator, and compare its asymptotic bias of some higher order with that of the FOGMM estimator.

3.1 Consistency and asymptotic distribution

For the GEL estimation, it is convenient to present results on asymptotic properties in both the homoskedastic and heteroskedastic cases together, though θ and other terms below may have different expressions in the two cases. The following theorem establishes the consistency of $\hat{\theta}_{n,\mathrm{GEL}}$ and related probability orders of the moment vector and the corresponding GEL estimate $\hat{\lambda}_{n,\mathrm{GEL}}$ of λ.

Theorem 3.1. *Under Assumptions 1–7,* $\hat{\theta}_{n,\mathrm{GEL}}\xrightarrow{p}\theta_0$, $g_n(\hat{\theta}_{n,\mathrm{GEL}})=O_p(n^{-1/2})$*; furthermore,* $\hat{\lambda}_{n,\mathrm{GEL}}=\arg\max_{\lambda\in\Lambda_n(\hat{\theta}_{n,\mathrm{GEL}})}\sum_{i=1}^n\rho(\lambda' g_{ni}(\hat{\theta}_{n,\mathrm{GEL}}))$ *exists with probability approaching one, and* $\hat{\lambda}_{n,\mathrm{GEL}}=O_p(n^{-1/2})$.

With the consistency of the GEL estimator, its asymptotic distribution can be derived by expanding the first order condition. Denote $\bar{G}_n=\mathrm{E}(\frac{\partial g_n(\theta_0)}{\partial\theta'})$.

Assumption 9. *(i)* $\theta_0\in\mathrm{int}(\Theta)$*; (ii)* $\lim_{n\to\infty}\bar{G}_n$ *has full rank.*

As usual, Assumption 9(i) guarantees the validity of the first order condition, which will be used to derive the asymptotic distribution of the estimator, and Assumption 9(ii) rules out functionally dependent moments. Let $\gamma=(\theta',\lambda')'$ and $\gamma_0=(\theta_0',0_{1\times k_g})'$. Furthermore, denote $\bar{\Sigma}_n=(\bar{G}_n'\bar{\Omega}_n^{-1}\bar{G}_n)^{-1}$ and $\bar{D}_n=\bar{\Omega}_n^{-1}-\bar{\Omega}_n^{-1}\bar{G}_n(\bar{G}_n'\bar{\Omega}_n^{-1}\bar{G}_n)^{-1}\bar{G}_n'\bar{\Omega}_n^{-1}$. The next theorem shows that $\hat{\gamma}_{n,\mathrm{GEL}}=(\hat{\theta}_{n,\mathrm{GEL}}',\hat{\lambda}_{n,\mathrm{GEL}}')'$ is asymptotically normal.

Theorem 3.2. *Under Assumptions 1–7 and 9,* $\sqrt{n}(\hat{\gamma}_{n,\text{GEL}}-\gamma_0) \xrightarrow{d} N\big(0, \lim_{n\to\infty} \text{diag}(\bar{\Sigma}_n, \bar{D}_n)\big)$*, where* $\text{diag}(\bar{\Sigma}_n, \bar{D}_n)$ *is the block diagonal matrix formed by the blocks* $\bar{\Sigma}_n$ *and* $\bar{D}_n$*.*

This theorem shows that the GEL estimator $\hat{\theta}_{n,\text{GEL}}$ of θ_0 has the same asymptotic distribution as the GMM estimator $\hat{\theta}_{n,\text{GMM}}$ in (7) (see e.g., Lee (2007) and Lee and Liu (2010)).

3.2 Stochastic expansion and high order asymptotic bias

To study high order asymptotic biases of the GMM and GEL estimators, we shall first derive Nagar-type expansions (Nagar, 1959) of a $\sqrt{n}$-consistent estimator $\hat{\gamma}_n$ of γ_0 with the form

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) = \xi_n + n^{-1/2}\psi_n + O_p(n^{-1}), \tag{8}$$

where $\xi_n = O_p(1)$, $\text{E}(\xi_n) = 0$ and $\psi_n = O_p(1)$. High order bias of $\hat{\gamma}_n$ can be computed as $\frac{1}{n}\,\text{E}(\psi_n)$.

For the FOGMM estimator $\hat{\theta}_{n,\text{GMM}}$, following Newey and Smith (2004), an auxiliary parameter vector

$$\hat{\lambda}_{n,\text{GMM}} = -\Omega_n^{-1}(\tilde{\theta}_n)g_n(\hat{\theta}_{n,\text{GMM}})$$

can be defined to make the derivation of its corresponding Nagar-type expansion easier.[17] With $\hat{\lambda}_{n,\text{GMM}}$, the first order condition for $\hat{\theta}_{n,\text{GMM}}$ can be written as

$$0 = -\begin{pmatrix} G_n'(\hat{\theta}_{n,\text{GMM}})\hat{\lambda}_{n,\text{GMM}} \\ g_n(\hat{\theta}_{n,\text{GMM}}) + \Omega_n(\tilde{\theta}_n)\hat{\lambda}_{n,\text{GMM}} \end{pmatrix}, \tag{9}$$

and we derive the expansion for the whole vector $\hat{\gamma}_{n,\text{GMM}} = (\hat{\theta}_{n,\text{GMM}}', \hat{\lambda}_{n,\text{GMM}}')'$. The expansion requires the existence of higher order moments of disturbances.

Assumption 10. $\sup_n \sup_{1\le i\le n} \text{E}\,|v_{ni}|^8 < \infty$.

Theorem 3.3. *For the FOGMM estimator* $\hat{\gamma}_{n,\text{GMM}}$*, under Assumptions 1–5 and 7–10, the expansion* (8) *holds with* $\xi_n = -\binom{\bar{H}_n}{\bar{D}_n}\sqrt{n}g_n(\theta_0)$*, where* $\bar{H}_n = (\bar{G}_n'\bar{\Omega}_n^{-1}\bar{G}_n)^{-1}\bar{G}_n'\bar{\Omega}_n^{-1}$*.*

The explicit form of ψ_n for the asymptotic expansion of $\hat{\gamma}_{n,\text{GMM}}$ is rather complex, but can be found in Appendix A in the proof of this theorem. A similar expansion for the GEL estimator $\hat{\theta}_{n,\text{GEL}}$ is also in Appendix A. The expansion requires further smoothness condition on $\rho(v)$.

Assumption 11. $\rho(v)$ *is four times continuously differentiable with Lipschitz fourth derivative in a neighborhood of zero.*

Theorem 3.4. *For the GEL estimator* $\hat{\gamma}_{n,\text{GEL}}$*, under Assumptions 1–7 and 9–11, the expansion* (8) *holds with* $\xi_n = -\binom{\bar{H}_n}{\bar{D}_n}\sqrt{n}g_n(\theta_0)$*.*

The above two theorems show that $\hat{\gamma}_{n,\text{GEL}}$ and $\hat{\gamma}_{n,\text{GMM}}$ are asymptotically equivalent. With the expansions, we can compute the asymptotic biases of the FOGMM and GEL estimators with the form $\frac{1}{n}\,\text{E}(\psi_n)$. Let $\Omega_n = \Omega_n(\theta_0)$,

[17] Alternatively, the expansion can be directly derived from the GMM first order condition as in Rilstone et al. (1996).

$G_n = G_n(\theta_0)$, $\bar{G}_n^{(l)} = \mathrm{E}(\frac{\partial G_n(\theta_0)}{\partial \theta_l})$, $g_n = g_n(\theta_0)$, $g_{ni} = g_{ni}(\theta_0)$, $g_{ni}^{(l)} = \frac{\partial g_{ni}(\theta_0)}{\partial \theta_l}$, and $e_{k_\theta,l}$ be the lth column of the $k_\theta \times k_\theta$ identity matrix, where θ_l denotes the lth element of θ.

Theorem 3.5. *Under Assumptions 1–5 and 7–10, the bias of the FOGMM estimator* $\hat{\theta}_{n,\mathrm{GMM}}$ *is* $B_n^I + B_n^G + B_n^\Omega + B_n^J$, *where* $B_n^I = \bar{H}_n \mathrm{E}(G_n \bar{H}_n g_n) - \frac{1}{2n}\sum_{l=1}^{k_\theta} \bar{H}_n \bar{G}_n^{(l)} \bar{\Sigma}_n e_{k_\theta,l}$, $B_n^G = -\bar{\Sigma}_n \mathrm{E}(G_n' \bar{D}_n g_n)$, $B_n^\Omega = \bar{H}_n \mathrm{E}(\Omega_n \bar{D}_n g_n)$ *and* $B_n^J = -\sum_{l=1}^{k_\theta} \frac{1}{n^2}\sum_{i=1}^n \bar{H}_n[\mathrm{E}(g_{ni} g_{ni}^{(l)\prime} + g_{ni}^{(l)} g_{ni}')](\bar{H}_n^J - \bar{H}_n)' e_{k_\theta,l}$, *with* $\bar{H}_n^J = (\bar{G}_n' \bar{J}_n^{-1} \bar{G}_n)^{-1} \bar{G}_n' \bar{J}_n^{-1}$.

In Theorem 3.5, B_n^I is the asymptotic bias for a GMM estimator dealing with the optimal linear combination $\bar{G}_n' \bar{\Omega}_n^{-1} g_n(\theta)$ of empirical moments $g_n(\theta)$; B_n^G arises from estimating $\bar{G}_n$; B_n^Ω arises from estimating the second moment matrix $\bar{\Omega}_n$ with the empirical variance Ω_n; and B_n^J arises from the choice of the initial GMM estimator. For the latter, if $\bar{J}_n$ is a scalar multiple of $\bar{\Omega}_n$, then $B_n^J = 0$ as $\bar{H}_n = \bar{H}_n^J$. With exact identification, $\bar{D}_n = 0$. In that case, $B_n^G = B_n^\Omega = B_n^J = 0$. Let $G_{ni} = \frac{\partial g_{ni}(\theta_0)}{\partial \theta'} = [g_{ni}^{(1)}, \ldots, g_{ni}^{(k_\theta)}]$.

Theorem 3.6. *Under Assumptions 1–7 and 9–11, the bias of the GEL estimator* $\hat{\theta}_{n,\mathrm{GEL}}$ *is* $B_n^I + B_n^G - \tilde{B}_n^G + B_n^\Omega + \frac{\rho_3}{2}\tilde{B}_n^\Omega$, *where* $\tilde{B}_n^G = -\frac{1}{n^2}\bar{\Sigma}_n \sum_{i=1}^n \mathrm{E}(G_{ni}' \bar{D}_n g_{ni})$, $\rho_3 = \frac{d^3\rho(0)}{dv^3}$, *and* $\tilde{B}_n^\Omega = \frac{1}{n^2}\sum_{i=1}^n \bar{H}_n \mathrm{E}(g_{ni} g_{ni}' \bar{D}_n g_{ni})$.

Again with exact identification, $\tilde{B}_n^G$ and $\tilde{B}_n^\Omega$ are also zero, so the bias is simply B_n^I. For overidentification cases, since $g_{ni}(\theta_0)$'s are not independent across i, in general, $B_n^G \neq \tilde{B}_n^G$ and $B_n^\Omega \neq \tilde{B}_n^\Omega$. Thus, unlike the case with i.i.d. data, the bias of the GEL estimator does not reduce to $B_n^I + B_n^\Omega + \frac{\rho_3}{2}\tilde{B}_n^\Omega$ and does not reduce further to B_n^I for the EL with $\rho_3 = -2$. The GEL only partially removes the asymptotic bias from the correlation between $G_n(\theta_0)$ and $g_n(\theta_0)$. This conclusion is similar to that in Anatolyev (2005) for stationary time series models with non-i.i.d. but serially uncorrelated data. In the presence of serial correlation, by using smoothed moment conditions, Anatolyev (2005) shows that, for certain kernel functions, the GEL can completely remove some high order bias terms from those of the GMM based on the same smoothed moment conditions. However, the main purpose of introducing smoothing is to derive efficient GEL estimates. When sample observations are uncorrelated, the GEL yields efficient estimates and the standard practice is to use unsmoothed moment conditions. For SAR models, the martingale difference are uncorrelated, so the GEL can generate efficient estimates.[18]

When $g_n(\theta)$ only contains linear moments, g_{ni} becomes $Q_{ni} v_{ni}$. Then, with only IV estimation, $B_n^\Omega = \tilde{B}_n^\Omega$.

Corollary 3.1. *When* $g_n(\theta) = \frac{1}{n}Q_n' V_n(\theta)$, *the bias of the EL estimator reduces to* $B_n^I + B_n^G - \tilde{B}_n^G$, *and the bias of the FOGMM estimator is* $B_n^I + B_n^G + B_n^\Omega + B_n^J$, *where* $B_n^\Omega = \frac{1}{n^2}\bar{H}_n \sum_{i=1}^n Q_{ni} Q_{ni}' \bar{D}_n Q_{ni} \mathrm{E}(v_{ni}^3)$.

The bias of the EL estimator reduces to $B_n^I + B_n^G - \tilde{B}_n^G$ because the EL has $\rho_3 = -2$, and hence, it does not have a bias from estimation of the second moment matrix $\bar{\Omega}_n$. If we further assume that $\mathrm{E}(v_{ni}^3) = 0$ for $i = 1, \ldots, n$, then $B_n^\Omega = \tilde{B}_n^\Omega = 0$. In that case, B_n^Ω is removed from the bias of the FOGMM estimator, and $B_n^\Omega + \frac{\rho_3}{2}\tilde{B}_n^\Omega$ is removed from the bias of any GEL estimator.

[18] It might be possible to consider smoothed moment conditions, e.g., smoothing the martingale differences. However, it is not clear whether the resulting GEL estimate can have lower high order bias than that based on the unsmooth moment conditions. Smoothing also involves additional practical complications on bandwidth selection. We thus leave this interesting question to future research.

4 Test statistics

In this section, we investigate several popular test statistics for SAR models in the GEL framework, including the parameter restriction test, overidentification test, Moran's I test and spatial J test. As shown below, an interesting aspect of those test statistics in the GEL framework is their robustness to unknown heteroskedasticity as long as their moment conditions are valid, while conventional test statistics without taking into account carefully their heteroskedastic variances for relevant evaluation might not be robust. Furthermore, while certain conventional test statistics might not be robust to non-normal distributions if higher order moments are not properly taken into account, test statistics in the GEL framework can be robust without referring to higher order moments as they are internalized.

4.1 Test for parameter restrictions

We may test for parameter restrictions in the GEL framework. Let $\theta = (\alpha', \phi')'$, where α is a subvector of θ. Suppose that we are interested in testing whether the true value of α is equal to zero or more generally a known constant vector c_α. For example, if our interest is to test spatial dependence, α might be the vector of spatial dependence parameters κ and/or τ in (1). Let $\dot{\theta}_n = (c'_\alpha, \dot{\phi}'_n)'$ be the restricted GEL estimator with the restriction $\alpha = c_\alpha$ imposed, and $\dot{\lambda}_n = \arg\max_{\lambda \in \Lambda_n(\dot{\theta}_n)} \sum_{i=1}^n \rho(\lambda' g_{ni}(\dot{\theta}_n))$. By the max-min characterization of the saddle point of the GEL objective function, $\sum_{i=1}^n \rho(\dot{\lambda}'_n g_{ni}(\dot{\theta}_n)) \geq \sum_{i=1}^n \rho(\hat{\lambda}'_n g_{ni}(\dot{\theta}_n)) \geq \sum_{i=1}^n \rho(\hat{\lambda}'_n g_{ni}(\hat{\theta}_n))$. Then we have the following GEL ratio test.

Theorem 4.1. *Suppose that Assumptions 1–7 and 9 are satisfied. Then under the null hypothesis $H_0 : \alpha_0 = c_\alpha$, $2[\sum_{i=1}^n \rho(\dot{\lambda}'_n g_{ni}(\dot{\theta}_n)) - \sum_{i=1}^n \rho(\hat{\lambda}'_n g_{ni}(\hat{\theta}_n))] \xrightarrow{d} \chi^2(k_\alpha)$, where k_α is the dimension of α_0.*

The GEL ratio test is asymptotically equivalent to the distance difference test in the GMM framework (Donald et al., 2003). This is so also for an SAR model. But it does not involve estimation of an optimal weighting of moments as in the GMM distance difference test. The GEL ratio has a similarity to a classical likelihood ratio statistic. As long as the moment vector $g_n(\theta)$ is valid, this test statistic can be formulated and is robust to unknown heteroskedasticity. These latter and distribution-free features are more attractive than those of a likelihood ratio test statistic. In a likelihood ratio test, the likelihood function needs to be properly specified to take into account heteroskedasticity and distributions of sample observations. For this GEL, one relies only on moments and does not need to have the proper formulation of heteroskedastic variances and distributions of disturbances. Since it is invariant under equivalent forms of the null hypothesis, it does not have the drawback of the Wald test that virtually any value of the Wald statistic can be obtained by writing the null hypothesis in different ways (Lafontaine and White, 1986).

To understand power properties of this test statistic, we consider a local alternative sequence. Suppose that the true value of α is subject to a Pitman drift $\alpha_n = c_\alpha + n^{-1/2} d_\alpha$, where d_α is a $k_\alpha \times 1$ vector of constants. Let $\bar{G}_{n\alpha} = \mathrm{E}(\frac{\partial g_n(\theta_0)}{\partial \alpha'})$, $\bar{G}_{n\phi} = \mathrm{E}(\frac{\partial g_n(\theta_0)}{\partial \phi'})$, $\bar{D}_{n\phi} = \bar{\Omega}_n^{-1} - \bar{\Omega}_n^{-1} \bar{G}_{n\phi} (\bar{G}'_{n\phi} \bar{\Omega}_n^{-1} \bar{G}_{n\phi})^{-1} \bar{G}'_{n\phi} \bar{\Omega}_n^{-1}$, and $\chi^2(a, b)$ be a noncentral chi-squared distribution with a degrees of freedom and a noncentrality parameter b.

Theorem 4.2. *Suppose that Assumptions 1–7 and 9 are satisfied. Then, under the Pitman drift $\alpha_n = c_\alpha + n^{-1/2} d_\alpha$,*

$$2\Big[\sum_{i=1}^{n} \rho(\dot{\lambda}_n' g_{ni}(\dot{\theta}_n)) - \sum_{i=1}^{n} \rho(\hat{\lambda}_n' g_{ni}(\hat{\theta}_n))\Big] \xrightarrow{d} \chi^2\big(k_\alpha, \lim_{n\to\infty} d_\alpha' \bar{G}_{n\alpha}' \bar{D}_{n\phi} \bar{G}_{n\alpha} d_\alpha\big).$$

The GEL ratio statistic is asymptotically distributed with a noncentral chi-squared distribution, which is the same as that for a distance difference test in the GMM framework (see also Newey and West, 1987).

4.2 Overidentification test

Like the GMM, a properly normalized GEL objective function at the GEL estimator $(\hat{\theta}_n', \hat{\lambda}_n')'$ can provide an overidentification test of moment conditions. The test statistic $2[\sum_{i=1}^{n} \rho(\hat{\lambda}_n' g_{ni}(\hat{\theta}_n)) - n\rho(0)]$ is non-negative as $\rho(0)$ is the restricted value of $\frac{1}{n}\sum_{i=1}^{n} \rho(\lambda' g_{ni}(\hat{\theta}_n))$ with the restriction $\lambda = 0$ while $\frac{1}{n}\sum_{i=1}^{n} \rho(\hat{\lambda}_n' g_{ni}(\hat{\theta}_n))$ is an unrestricted maximum for λ.

Theorem 4.3. *Suppose that Assumptions 1–7 and 9 are satisfied. Then $2[\sum_{i=1}^{n} \rho(\hat{\lambda}_n' g_{ni}(\hat{\theta}_n)) - n\rho(0)] \xrightarrow{d} \chi^2(k_g - k_\theta)$, where the number of moments k_g is strictly greater than the number of parameters k_θ.*

This GEL overidentification test is asymptotically equivalent to the GMM overidentification test. In general, misspecification of an SAR model may come from different sources which give misspecified moment conditions. The overidentification test will be able to detect those misspecifications. If one believes that misspecification might come only from a particular source, then the overidentifcation test might detect it. However, for a specific direction of departure, it is desirable to design more powerful test statistics. In a subsequent section, we consider a non-nested test, namely, a J-test, for SAR models with different specifications of spatial weights matrices. Before that, we consider a test of spatial dependence, the well-known Moran's I statistic.

4.3 Moran's I test

Moran's I test is a popular test for spatial dependence. In practice, the vector of ordinary least squares (OLS) residuals $\hat{V}_n = [I_n - X_n(X_n'X_n)^{-1}X_n']Y_n$ from the regression of Y_n on X_n in the regression model $Y_n = X_n\beta + V_n$ is often used and the test is based on the asymptotic distribution of $\frac{1}{\sqrt{n}}\hat{V}_n' M_n \hat{V}_n$ for a spatial weights matrix M_n. After normalization with a proper standard error, an asymptotically normal distribution of the normalized statistic is used for testing. Such a test has a null hypothesis that v_{ni}' in V_n are independent.[19] Here we show that such a test of spatial dependence can be conveniently implemented in the GEL framework. This GEL test can be robust against unknown heteroskedasticity in disturbances, while there is no need to estimate any variance. To allow for possible spatial dependence with different specifications on spatial weights matrices, we use the vector of q

[19]Kelejian and Prucha (2001) propose a generalized Moran's I test that covers the SARAR and limited dependent variable models. Qu and Lee (2012, 2013) have considered the use of generalized residuals for the construction of locally most powerful LM tests for the spatial Tobit model.

moments $\sqrt{n}\hat{g}_n = \frac{1}{\sqrt{n}}[\hat{V}_n' M_{1n}\hat{V}_n, \ldots, \hat{V}_n' M_{qn}\hat{V}_n]'$ to construct a joint test, where M_{1n}, …, M_{qn} are spatial weights matrices. Let $g_{ni} = [v_{ni}\sum_{j=1}^{i-1}(m_{1n,ij} + m_{1n,ji})v_{nj}, \ldots, v_{ni}\sum_{j=1}^{i-1}(m_{qn,ij} + m_{qn,ji})v_{nj}]'$ and $\hat{g}_{ni} = [\hat{v}_{ni}\sum_{j=1}^{i-1}(m_{1n,ij} + m_{1n,ji})\hat{v}_{nj}, \ldots, \hat{v}_{ni}\sum_{j=1}^{i-1}(m_{qn,ij} + m_{qn,ji})\hat{v}_{nj}]'$, where $\hat{v}_{ni}$ is the ith element of $\hat{V}_n$, for $i = 1, \ldots, n$, and $\hat{\Lambda}_n = \{\lambda : \lambda'\hat{g}_{ni} \in \mathcal{V}, i = 1, \ldots, n\}$.[20]

Theorem 4.4. *Suppose that $Y_n = X_n\beta_0 + V_n$, the $n \times n$ nonstochastic matrices M_{1n}, …, M_{qn} have zero diagonals, the sequences $\{M_{1n}\}$, …, $\{M_{qn}\}$ are bounded in row and column sum norms, and $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(g_{ni}g_{ni}')$ is positive definite. Under Assumptions 1, 2 and 6,*

$$2\Big[\max_{\lambda\in\hat{\Lambda}_n}\sum_{i=1}^{n}\rho(\lambda'\hat{g}_{ni}) - n\rho(0)\Big] = \Big(\sum_{i=1}^{n}\hat{g}_{ni}\Big)'\Big(\sum_{i=1}^{n}\hat{g}_{ni}\hat{g}_{ni}'\Big)^{-1}\Big(\sum_{i=1}^{n}\hat{g}_{ni}\Big) + o_p(1) \xrightarrow{d} \chi^2(q).$$

The GEL test statistic can use the estimated $\hat{g}_{ni}$ in place of the true g_{ni}, because $\sqrt{n}\hat{g}_n$ with the OLS estimated $\hat{V}_n$ has the same asymptotic distribution as $\sqrt{n}g_n = \frac{1}{\sqrt{n}}[V_n' M_{1n}V_n, \ldots, V_n' M_{qn}V_n]$ with true V_n due to an orthogonality property. A conventional Moran's I test would need to evaluate the asymptotic variance of $\sqrt{n}g_n$ under the null. If we use $\sum_{i=1}^{n}\hat{g}_{ni}\hat{g}_{ni}'$ to estimate the variance of $\sum_{i=1}^{n}\hat{g}_{ni}$, a Moran's I test, which is robust to unknown heteroskedasticity, can be computed as $(\sum_{i=1}^{n}\hat{g}_{ni})'(\sum_{i=1}^{n}\hat{g}_{ni}\hat{g}_{ni}')^{-1}(\sum_{i=1}^{n}\hat{g}_{ni})$, as given in the above theorem. The GEL version of Moran's I test can bypass such calculations as the GEL takes care of unknown heteroskedasticity internally.

For the local power of Moran's I test, we consider the alternative model being an SE model, $Y_n = X_n\beta + U_n$ with $U_n = \sum_{j=1}^{q}\tau_{nj}M_{jn}U_n + V_n$, where the spatial error dependence parameter $\tau_n = [\tau_{n1}, \ldots, \tau_{nq}]'$ is subject to the Pitman drift $\tau_n = n^{-1/2}d_\tau = n^{-1/2}[d_{\tau 1}, \ldots, d_{\tau q}]'$ to zero.

Theorem 4.5. *Suppose that $Y_n = X_n\beta_0 + U_n$ with $U_n = \sum_{j=1}^{q}n^{-1/2}d_{\tau j}M_{jn}U_n + V_n$, where $d_{\tau j}$'s are constants, the $n \times n$ nonstochastic matrices M_{1n}, …, M_{qn} have zero diagonals, the sequences $\{M_{1n}\}$, …, $\{M_{qn}\}$ are bounded in row and column sum norms, and $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(g_{ni}g_{ni}')$ is positive definite. Under Assumptions 1, 2 and 6,*

$$2\Big[\max_{\lambda\in\hat{\Lambda}_n}\sum_{i=1}^{n}\rho(\lambda'\hat{g}_{ni}) - n\rho(0)\Big] \xrightarrow{d} \chi^2\Big(q, \lim_{n\to\infty}L_n' T_n^{-1}L_n\Big),$$

where $L_n = \frac{1}{n}\mathrm{E}[\sum_{j=1}^{q}V_n'(M_{1n} + M_{1n}')M_{jn}V_n d_{\tau j}, \ldots, \sum_{j=1}^{q}V_n'(M_{qn} + M_{qn}')M_{jn}V_n d_{\tau j}]'$ and T_n is a $q \times q$ matrix with its (j,k)th element being $\frac{1}{n}\mathrm{E}(V_n' M_{jn}V_n V_n' M_{kn}V_n)$.

We may compare this GEL Moran's I test with the parameter restriction test for spatial error dependence in the SARAR(0,q) model based on the moment vector $\frac{1}{n}[V_n' M_{1n}V_n, \ldots, V_n' M_{qn}V_n, V_n' X_n]'$. By Theorems 4.2 and 4.5, these test statistics have the same asymptotic distribution under the same Pitman drift.

The above GEL Moran's I test using the estimated moment vector $\sqrt{n}\hat{g}_n = \frac{1}{\sqrt{n}}[\hat{V}_n' M_{1n}\hat{V}_n, \ldots, \hat{V}_n' M_{qn}\hat{V}_n]'$ relies on the null model being a linear regression model. Consider the case that the model is an SARAR(p,q) and the test is for spatial dependence in disturbances. Then, with consistently estimated residual vector $\hat{V}_n$, which can be estimated residuals from a 2SLS or QML estimated SAR equation, $\sqrt{n}\hat{g}_n$ may not have the same asymptotic distribution as

[20] Note that $\hat{g}_{n1} = g_{n1} = 0$ by the convention of the summation notation. We define $\hat{g}_{n1}$ and g_{n1} for convenience.

$\sqrt{n}g_n$ and the test statistic would not be asymptotically chi-squared distributed. Neither would be the GEL test version. This problem occurs due to the issue that the consistent estimator used to construct those moments for testing has an impact on the asymptotic distribution of the moments.[21] To overcome this problem in the GEL framework, we may consider a corresponding $C(\alpha)$-type statistic as suggested in Jin and Lee (2017). Let $\theta = (\alpha, \phi')'$, where α is the spatial error dependence parameter vector τ and the test is on whether $\alpha_0 = 0$. Denote $\hat{\theta}_n = (0, \hat{\phi}_n')'$ for any $\sqrt{n}$-consistent estimator $\hat{\phi}_n$ of ϕ_0. Instead of the moment $g_{1n}(\theta) = \frac{1}{n}[V_n'(\theta)M_{1n}V_n(\theta), \ldots, V_n'(\theta)M_{qn}V_n(\theta)]'$, where $V_n(\theta) = R_n(\tau)[S_n(\kappa)Y_n - X_n\beta]$, we may use the moment $g_n(\theta) = g_{1n}(\theta) - \frac{\partial g_{1n}(\theta)}{\partial \phi'}(\frac{\partial g_{2n}(\theta)}{\partial \phi'})^{-1}g_{2n}(\theta)$, where $g_{2n}(\theta)$ is a $(k_\theta - q) \times 1$ vector of linear and quadratic moments. As $g_{1n}(\theta)$ and $g_{2n}(\theta)$ are linear and quadratic moments, $g_n(\theta)$ can be written as $g_n(\theta) = \frac{1}{n}\sum_{i=1}^n g_{ni}(\theta)$, where $g_{ni}(\theta) = g_{1n,i}(\theta) - \frac{\partial g_{1n}(\theta)}{\partial \phi'}(\frac{\partial g_{2n}(\theta)}{\partial \phi'})^{-1}g_{2n,i}(\theta)$ with $g_{1n,i}(\theta_0)$'s and $g_{2n,i}(\theta_0)$'s being martingale differences. In place of the estimated moment $g_{1n}(\hat{\theta}_n)$, we would consider the alternative $g_n(\hat{\theta}_n)$. By the mean value theorem, we can see that $\sqrt{n}g_n(\hat{\theta}_n)$ has the same asymptotic distribution as $\sqrt{n}g_n(\theta_0)$.

Theorem 4.6. *For model* (1) *with $\tau_0 = 0$, suppose that Assumptions 1–3 and 6 hold, $\sqrt{n}(\hat{\phi}_n - \phi_0) = O_p(1)$, and $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathrm{E}[g_{ni}(\theta_0)g_{ni}'(\theta_0)]$ is positive definite. Then,*

$$2\Big[\max_{\lambda \in \Lambda_n(\hat{\theta}_n)} \sum_{i=1}^n \rho(\lambda' g_{ni}(\hat{\theta}_n)) - n\rho(0)\Big] = \Big[\sum_{i=1}^n g_{ni}(\hat{\theta}_n)\Big]'\Big[\sum_{i=1}^n g_{ni}(\hat{\theta}_n)g_{ni}'(\hat{\theta}_n)\Big]^{-1}\Big[\sum_{i=1}^n g_{ni}(\hat{\theta}_n)\Big] + o_p(1) \xrightarrow{d} \chi^2(q).$$

The test statistic is readily available with the GEL estimate of the Lagrangian multiplier λ. It is robust to unknown heteroskedasticity if quadratic matrices in the quadratic moments of $g_{2n}(\theta)$ have zero diagonals. This GEL test can use any $\sqrt{n}$-consistent estimator $\hat{\phi}_n$. However, it is desirable to choose $g_{2n}(\theta)$ and its moment estimator $\hat{\theta}_n = (0, \hat{\phi}_n')'$ such that $g_{2n}(0, \hat{\phi}_n) = 0$. Because with such moments, the estimated moment vector $g_n(\hat{\theta}_n)$ is exactly the same as the estimated moment $g_{1n}(\hat{\theta}_n)$ and we do not change the basic moments $g_{1n}(\theta)$ for testing. However, the individual $g_{ni}(\hat{\theta}_n)$ and $g_{1n,i}(\hat{\theta}_n)$ are different even though their summations over i are the same. The direct use of $g_{1n,i}(\hat{\theta}_n)$ in a GEL test would not overcome the impact of $\hat{\theta}_n$ on the asymptotic distribution of that GEL test statistic while the former can, because $g_n(\theta)$ has an orthogonality property while $g_{1n}(\theta)$ does not.

4.4 Spatial J test

Empirical researchers often face the problem on how to specify econometric models. In spatial econometrics, since an economic theory may be ambiguous on spatial weights matrices, their specifications are frequently challenged. Thus we may have possible specifications of SARAR models with different spatial weights matrices. For testing and model selection, SARAR models with different spatial weights matrices are non-nested. A popular testing procedure is based on the spatial J test (Kelejian, 2008; Kelejian and Piras, 2011).[22] In this section, we formulate the spatial J test in the GEL framework.

[21] For Moran's I test, the orthogonality holds because $\frac{1}{\sqrt{n}}(Y_n - X_n\hat{\beta}_n)'M_{jn}(Y_n - X_n\hat{\beta}_n) = \frac{1}{\sqrt{n}}(Y_n - X_n\beta_0)'M_{jn}(Y_n - X_n\beta_0) + o_p(1)$ due to $\hat{\beta}_n$ being the OLS estimator.

[22] Cox-type tests for SARAR models are developed in Jin and Lee (2013). Delgado and Robinson (2015) propose non-nested tests in a general spatial, spatio-temporal or panel data context.

Suppose that we are interested in testing model (1) against an alternative SARAR(p_1,q_1) model:

$$Y_n = \sum_{j=1}^{p_1} \boldsymbol{\kappa}_j \mathcal{W}_{jn} Y_n + \mathcal{X}_n \boldsymbol{\beta} + \mathcal{U}_n, \quad \mathcal{U}_n = \sum_{k=1}^{q_1} \boldsymbol{\tau}_k \mathcal{M}_{kn} \mathcal{U}_n + \mathcal{V}_n, \tag{10}$$

where $\mathcal{W}_{jn}$, $\mathcal{M}_{jn}$, $\mathcal{X}_n$ and $\mathcal{V}_n$ have similar properties as those in model (1). The parameters are in boldface to distinguish them from those in model (1).[23] The J test is originated in Davidson and MacKinnon (1981) and is based on whether the alternative model can significantly improve the prediction of the dependent variable vector Y_n. Let $\hat{\boldsymbol{\kappa}}_n$ and $\hat{\boldsymbol{\beta}}_n$ be, respectively, estimators of $\boldsymbol{\kappa}$ and $\boldsymbol{\beta}$ in (10), which are consistent if model (10) were the true model. The $\hat{\boldsymbol{\kappa}}_n$ and $\hat{\boldsymbol{\beta}}_n$ can be the QML, GMM or even GEL estimators.[24] A predictor of Y_n from the alternative model can be either $\hat{Y}_n = \sum_{j=1}^{p_1} \hat{\boldsymbol{\kappa}}_{jn} \mathcal{W}_{jn} Y_n + \mathcal{X}_n \hat{\boldsymbol{\beta}}_n$ using the main equation of (10) or $\hat{Y}_n = \mathcal{S}_n^{-1}(\hat{\boldsymbol{\kappa}}_n) \mathcal{X}_n \hat{\boldsymbol{\beta}}_n$ using the reduced form of Y_n under (10), where $\mathcal{S}_n(\hat{\boldsymbol{\kappa}}_n) = I_n - \sum_{j=1}^{p_1} \hat{\boldsymbol{\kappa}}_{jn} \mathcal{W}_{jn}$. The difference of using the two versions has been discussed in Kelejian and Piras (2011). As Y_n is on the right hand side of the first prediction version, that $\hat{Y}_n$ would be endogenous, while the second one is exogenous. The spatial J test of (1) against (10) is based on an augmented model:

$$Y_n = \sum_{j=1}^{p} \kappa_j W_{jn} Y_n + X_n \beta + \eta \hat{Y}_n + U_n, \quad U_n = \sum_{k=1}^{q} \tau_k M_{kn} U_n + V_n, \tag{11}$$

where $\hat{Y}_n$ is added in the null model (1) to predict Y_n. We test whether the coefficient η is significantly different from zero or not. If it is not significant, we do not reject the null model; otherwise, we reject it. In Kelejian and Piras (2011), the spatial J test uses the GS2SLS to estimate the augmented model.[25] When $\hat{Y}_n$ is exogenous, it can be included in the IV matrix Q_n and no extra IV is needed for $\hat{Y}_n$. For the version that $\hat{Y}_n$ is endogenous, extra IVs would be needed for $\hat{Y}_n$. The GS2SLS uses only linear IV moments but does not utilize quadratic moments for the main equation of (11). Thus it may lead to a relatively inefficient estimator and a less powerful test (Jin and Lee, 2013). Here as a generalization, we consider the GEL estimation of model (11) with both linear and quadratic moments.

For the augmented model (11), let $V_n(\vartheta) = R_n(\tau)[S_n(\kappa) Y_n - X_n \beta - \eta \hat{Y}_n]$, where $\vartheta = (\theta', \eta)'$. The moment vector can be

$$g_n(\vartheta) = [V_n'(\vartheta) P_{1n} V_n(\vartheta) - \sigma^2 \operatorname{tr}(P_{1n}), \ldots, V_n'(\vartheta) P_{k_p n} V_n(\vartheta) - \sigma^2 \operatorname{tr}(P_{k_p n}), Q_n' V_n(\vartheta)]$$

in the homoskedastic case, and

$$g_n(\vartheta) = [V_n'(\vartheta) P_{1n} V_n(\vartheta), \ldots, V_n'(\vartheta) P_{k_p n} V_n(\vartheta), Q_n' V_n(\vartheta)]$$

where each P_{ln}, $l = 1, \ldots, k_p$, has a zero diagonal in the heteroskedastic case. Define $g_{ni}(\vartheta)$ such that $g_n(\vartheta) =$

[23] While it is possible to test one model against several alternatives simultaneously, we only consider one alternative model for simplicity.

[24] Large sample properties of the GEL estimators $\hat{\boldsymbol{\kappa}}_n$ and $\hat{\boldsymbol{\beta}}_n$ are presented in Appendix B under regularity conditions for misspecified models.

[25] For the original spatial J test, which uses the GS2SLS to estimate the augmented model, the main equation of (11) is transformed by pre-multiplying it with $R_n(\hat{\tau}_n)$ before estimation, where $\hat{\tau}_n$ is a consistent estimator of τ (Kelejian and Piras, 2011).

$\frac{1}{n}\sum_{i=1}^n g_{ni}(\vartheta)$, and $g_{ni}(\vartheta_0)$'s are martingale differences under the null, where $\vartheta_0 = (\theta_0', 0)'$. The GEL estimator is

$$\hat{\vartheta}_n = \arg\min_{\vartheta \in \Theta} \max_{\lambda \in \Lambda_n(\vartheta)} \sum_{i=1}^n \rho(\lambda' g_{ni}(\vartheta)),$$

where $\Lambda_n(\vartheta) = \{\lambda : \lambda' g_{ni}(\vartheta) \in \mathcal{V}, i = 1, \ldots, n\}$ and Θ is the parameter space of ϑ. With the identification and regularity conditions in Appendix B, the spatial J test statistic can be formulated as a GEL ratio. This GEL ratio test is essentially a test on the parameter restriction that $\eta = 0$ in (11). It differs from the one in Section 4.1 in that here $\hat{Y}_n$ on the right hand side of (11) is a generated regressor. As the following theorem will show, the initial estimate in $\hat{Y}_n$ does not have an asymptotic impact on the GEL statistic under the null, because, at $\eta = 0$, $\hat{Y}_n$ is an irrelevant variable.

Theorem 4.7. *Suppose that Assumptions 1–7, 9 and 12 hold and ϑ_0 is in the interior of the compact parameter space Θ. Then, under H_0, $2[\sum_{i=1}^n \rho(\hat{\lambda}_n' g_{ni}(\hat{\theta}_n)) - \max_{\lambda \in \Lambda_n(\hat{\vartheta}_n)} \sum_{i=1}^n \rho(\lambda' g_{ni}(\hat{\vartheta}_n))] \xrightarrow{d} \chi^2(1)$, where $(\hat{\theta}_n', \hat{\lambda}_n')'$ is the GEL estimator for model (1), i.e., it is the restricted GEL estimator for (11) with the restriction $\eta = 0$ imposed.*

5 Monte Carlo

In this section, we report Monte Carlo results on the GEL estimator and test statistics considered in this paper.

5.1 Estimation

For estimation, we consider the following SARAR(2,0) model:[26]

$$Y_n = \kappa_1 W_{1n} Y_n + \kappa_2 W_{2n} Y_n + X_n \beta + V_n. \tag{12}$$

The first spatial weights matrix W_{1n} is based on the circular world matrix in Arraiz et al. (2010). This matrix has spatial units equally spaced on a circle, one third of which are connected to ten nearest neighbors and the rest are connected to two nearest neighbors. The second spatial weights matrix W_{2n} is based on the queen criterion. These matrices are normalized to have row sums equal to one. There are three exogenous variables in X_n: an intercept term, a variable randomly drawn from the standard normal distribution $N(0,1)$ and a variable from the uniform distribution $U[0,\sqrt{12}]$. The true value β_0 of $\beta = (\beta_1, \beta_2, \beta_3)'$ is $[0.5, 0.5, 0.5]'$. The disturbances v_{ni}'s are randomly drawn from the normal distribution $N(0, \sigma_0^2)$ in the homoskedastic case, or $N(0, \sigma_0^2 c_i^2)$ in the heteroskedastic case, where c_i is the number of nonzero elements in the ith row of W_{1n}, and σ_0^2 is chosen such that $R^2 \equiv \mathrm{var}(X_n\beta_0)/[\mathrm{var}(X_n\beta_0) + \bar{\sigma}_n^2]$ is either 0.4 or 0.8, where $\bar{\sigma}_n^2$ is the average variance of all v_{ni}'s. For the estimation of model (12), the IV matrix $Q_n = [X_n, W_{1n}X_n^*, W_{2n}X_n^*, W_{1n}^2 X_n^*, W_{2n}^2 X_n^*]$, where X_n^* is a submatrix of X_n that excludes the intercept term so that Q_n only contains one intercept. In the homoskedastic case, we use the moment vector $\frac{1}{n}[V_n'V_n - n\sigma_0^2, V_n'W_{1n}V_n, V_n'W_{2n}V_n, V_n'W_{1n}^2 V_n - \sigma_0^2 \operatorname{tr}(W_{1n}^2), V_n'W_{2n}^2 V_n - \sigma_0^2 \operatorname{tr}(W_{2n}^2), V_n'Q_n]'$;

[26]To illustrate the performance of GEL estimation and tests for high order SARAR models, we report results for models with two spatial lags of the dependent variable or disturbances. Some Monte Carlo results for the SARAR(1,1) model are reported in the supplementary file.

in the heteroskedastic case, we use the moment vector $\frac{1}{n}[V_n'W_{1n}V_n, V_n'W_{2n}V_n, V_n'(W_{1n}^2 - \text{diag}(W_{1n}^2))V_n, V_n'(W_{2n}^2 - \text{diag}(W_{2n}^2))V_n, V_n'Q_n]'$. The number of Monte Carlo repetitions for each case is $1,000$.

Table 1 reports biases, standard errors, and root mean square errors (RMSE) of the GMM, EL and ET estimators in the homoskedastic case.[27] The GMM estimator is a FOGMM estimator where in the first step the identity matrix is used as the weighting matrix to derive a consistent estimator $\tilde{\theta}_n$ and in the second step $\Omega_n(\tilde{\theta}_n)$ is used as the weighting matrix. Overall, the three estimators all have relatively small biases. The biases of the EL and ET estimators for β_2, β_3 and σ^2 are generally smaller than those of the GMM estimators, and the biases for other parameters are slightly larger. In particular, for σ^2, the bias of the GMM estimator is significantly larger than that of the ET estimator, while the latter is larger than that of the EL estimator. For the comparison of the EL and ET, except for the variance parameter σ^2, they have similar biases in most cases and neither the EL nor the ET would dominate each other. In terms of standard errors, the three estimators have similar performance. Since standard errors of estimates dominate biases for parameters other than σ^2, the RMSEs display an order in magnitude similar to that of standard errors. For σ^2, the EL estimator has the smallest RMSE, and the ET estimator has a smaller RMSE than that of the GMM estimator. As the sample size increases from 144 to 400, biases generally decrease, and standard errors decrease approximately at the theoretical rate.

[Table 1 about here.]

Table 2 shows summary statistics of the estimators in the heteroskedastic case. While all biases are small, the bias of the GMM estimator for κ_1 with a small R^2 is larger than those of the EL and ET, and it is slightly smaller in other cases. The GMM estimator has slightly smaller standard errors than the EL and ET estimators and thus smaller RMSEs. The EL estimator is observed to have larger biases, standard errors and RMSEs than those of the ET estimator.

[Table 2 about here.]

To further compare the performance of estimators, Table 3 reports coverage probabilities (CP) of 95% confidence intervals for model parameters. In the homoskedastic case, for $n = 144$, the GMM CPs are below 95%, and those for σ^2 are much smaller than 95%; the EL and ET CPs are significantly closer to 95% than GMM ones, and those for σ^2 are at least ten percentage points higher than corresponding GMM CPs. The ET CPs are higher than EL ones except for σ^2. With a larger sample size $n = 400$, the CPs are closer to 95%, but the patterns are similar. In the heteroskedastic case, the EL and ET CPs are still closer to 95% than GMM ones in general, though the differences are smaller.

[Table 3 about here.]

[27] We do not consider the continuous updating GMM estimator because it is often observed to possess multiple modes and thus generally considered to be less desirable than the EL and ET estimators (Hansen et al., 1996; Imbens et al., 1998).

5.2 Tests for spatial dependence

For various tests for spatial dependence, the data generating process (DGP) is the following SARAR(0,2) model:

$$Y_n = X_n\beta + U_n, \quad U_n = \tau_1 M_{1n} U_n + \tau_2 M_{2n} U_n + V_n, \tag{13}$$

where M_{1n} and M_{2n} are equal to, respectively, W_{1n} and W_{2n} in model (12), and other settings are the same as for (12). The null hypothesis is $\tau_{10} = \tau_{20} = 0$. We consider nine tests in the homoskedastic case: "PT_{GMM}", "PT_{EL}" and "PT_{ET}" denote parameter restriction tests implemented with, respectively, the GMM distance difference, EL ratio and ET ratio based on the moment vector $\frac{1}{n}[V_n'V_n - n\sigma_0^2, V_n'M_{1n}V_n, V_n'M_{1n}^2V_n - \sigma_0^2\,\text{tr}(M_{1n}^2), V_n'M_{2n}V_n, V_n'M_{2n}^2V_n - \sigma_0^2\,\text{tr}(M_{2n}^2), V_n'Q_n]'$; "$\text{OT}_{\text{GMM}}$", "$\text{OT}_{\text{EL}}$" and "$\text{OT}_{\text{ET}}$" denote, respectively, the GMM, EL and ET overidentification tests based on the moment vector $\frac{1}{n}[V_n'M_{1n}V_n, V_n'M_{2n}V_n, V_n'X_n]'$; "Moran" denotes Moran's I test with a robust variance estimator, and "Moran_{EL}" and "Moran_{ET}" denote, respectively, EL and ET Moran's I tests. The three Moran's I tests are based on the moment vector $\frac{1}{n}[V_n'M_{1n}V_n, V_n'M_{2n}V_n]'$, and OLS residuals are used to formulate test statistics. In the heteroskedastic case, the above tests are also considered, among which parameter restriction tests are based on the moment vector $\frac{1}{n}[V_n'M_{1n}V_n, V_n'(M_{1n}^2 - \text{diag}(M_{1n}^2))V_n, V_n'M_{2n}V_n, V_n'(M_{2n}^2 - \text{diag}(M_{2n}^2))V_n, V_n'Q_n]'$ which is robust to unknown heteroskedasticity. In addition, we consider two tests which do not take into account unknown heteroskedasticity: the GMM parameter restriction test "PT^*_{GMM}" based on the moment vector $\frac{1}{n}[V_n'V_n - n\sigma_0^2, V_n'M_{1n}V_n, V_n'M_{1n}^2V_n - \sigma_0^2\,\text{tr}(M_{1n}^2), V_n'M_{2n}V_n, V_n'M_{2n}^2V_n - \sigma_0^2\,\text{tr}(M_{2n}^2), V_n'Q_n]'$, and Moran's I test "Moran^*" based on the moment vector $\frac{1}{n}[V_n'M_{1n}V_n, V_n'M_{2n}V_n]'$, for which the involved variance is computed using its analytical form as if the disturbances were homoskedastic.

Table 4 presents empirical sizes for a nominal size of 5%. In the homoskedastic case, PT_{EL} and PT_{ET} have relatively large sizes for small sample cases and have improved sizes for larger sample sizes. Other tests have relatively small size distortions. In the heteroskedastic case, size distortions are larger. As expected, PT^*_{GMM} and Moran^* have large size distortions and the distortions do not improve with larger sample sizes. Powers of these tests except for PT^*_{GMM} and Moran^* are presented in Table 5 and Table 6. Their powers are generally similar. R^2 does not have much impact on powers. These tests are powerful in cases with larger τ_{01} and τ_{02} and a larger sample size.

[Table 4 about here.]

[Table 5 about here.]

[Table 6 about here.]

5.3 Spatial *J* tests

For spatial J tests, one SARAR(1,1) model

$$Y_n = \kappa W_n Y_n + X_n\beta + U_n, \quad U_n = \tau M_n U_n + V_n \tag{14}$$

is tested against another one. We set $W_n = M_n$. The null and alternative models only differ in W_n; specifically, the circular world matrix and the one based on the queen criterion are tested against each other. To estimate the null and alternative models, we use moment vectors similar to those for model (12). To estimate the augmented model (11), if $\hat{Y}_n = \mathcal{S}_n^{-1}(\hat{\boldsymbol{\kappa}}_n)\mathcal{X}_n\hat{\boldsymbol{\beta}}_n$ is used as the augmented explanatory variable, $\hat{Y}_n$ is added to the IV matrix in the above moment vectors; on the other hand, if $\hat{Y}_n = \hat{\boldsymbol{\kappa}}_n\mathcal{W}_n y_n + \mathcal{X}_n\hat{\boldsymbol{\beta}}_n$ is the augmented explanatory variable, $\mathcal{W}_n X_n^*$ is added to the IV matrix.

Tables 7 and 8 report empirical sizes and powers of spatial J tests for the SARAR model (14). "GMM_1" denotes the spatial J test implemented with the GMM distance difference test using the predictor $\hat{Y}_n = \hat{\boldsymbol{\kappa}}_n\mathcal{W}_n Y_n + \mathcal{X}_n\hat{\boldsymbol{\beta}}_n$, and "$\text{GMM}_2$" uses $\hat{Y}_n = \mathcal{S}_n^{-1}(\hat{\boldsymbol{\kappa}}_n)\mathcal{X}_n\hat{\boldsymbol{\beta}}_n$. Correspondingly, we have EL and ET ratio tests "EL_1", "EL_2", "ET_1" and "ET_2". The EL_1, EL_2, ET_1 and ET_2 have relatively larger size distortions for a small sample size, but are reasonably adequate for a larger sample size. Powers of these tests are similar. With larger R^2, κ_0 and sample sizes, these tests are more powerful.

[Table 7 about here.]

[Table 8 about here.]

6 Conclusion

By exploiting the martingale structure of linear and quadratic empirical moments of the high order SARAR model, this paper considers its GEL estimation and tests in both homoskedastic and heteroskedastic cases. We show that the GEL estimator is consistent and has the same asymptotic normal distribution as the optimal GMM estimator based on the same moment conditions. But the GEL avoids a first step estimation of the optimal weighting matrix with a preliminary estimator and can be robust to unknown heteroskedasticity without the computation of possibly higher order moment parameters of disturbances. The GEL is free from the asymptotic bias of the preliminary estimator and partially removes the bias due to the correlation between the moment conditions and their Jacobian. The EL further partially removes the bias from estimating the second moment matrix. We also investigate the GEL overidentification test, Moran's I test, and GEL ratio tests for parameter restrictions and non-nested hypotheses. These tests do not involve estimation of variances and higher order moment parameters, and can be robust to unknown heteroskedasticity. Our Monte Carlo results show that GEL estimators and tests perform well compared with GMM estimators and tests when the latter GMM estimates and tests take into account properly their variances and/or moment parameters of disturbances. The GMM tests are not robust while GEL tests are much better at dealing with the extra complexity of spatial models.

In a future research, it is of interest to investigate various optimality properties of EL tests for the SARAR model as in Kitamura (2001) and Otsu (2010), and their Bartlett correctability. The latter is expected by Mykland

(1995). However, Bartlett correctability is based on Edgeworth expansions, for which it is not known how to show general pointwise results on martingales.[28]

Acknowledgments

We are grateful to the editor Oliver Linton, an Associate Editor and two anonymous referees for their insightful comments that lead to substantial improvement of this paper. We also would like to thank seminar participants of NUS, Xiamen University, the 2017 Asian Meeting of the Econometric Society at CUHK, and 2017 China Meeting of the Econometric Society at Wuhan University, China, for helpful comments. The first author gratefully acknowledges the financial support from the National Natural Science Foundation of China (No. 71501119) and Shanghai Pujiang Program (No. 15PJC043).

Appendix A Proofs

Proof of Theorem 3.1. By the reduced form of Y_n, $V_n(\theta)$ is linear in V_n and quadratic in θ. Then each element of $g_n(\theta)$ can be expanded as a linear-quadratic form of V_n and is a polynomial of θ. By Lemma 2(i), $\sup_{\theta\in\Theta}\|g_n(\theta) - \mathrm{E}[g_n(\theta)]\| \xrightarrow{p} 0$. By Lemma 12, $g_n(\hat{\theta}_{n,\text{GEL}}) = O_p(n^{-1/2})$. Let $\bar{g}_n(\theta) = \mathrm{E}[g_n(\theta)]$, then $\|\bar{g}_n(\hat{\theta}_{n,\text{GEL}})\| = \|\bar{g}_n(\hat{\theta}_{n,\text{GEL}}) - g_n(\hat{\theta}_{n,\text{GEL}}) + g_n(\hat{\theta}_{n,\text{GEL}})\| \leq \|\bar{g}_n(\hat{\theta}_{n,\text{GEL}}) - g_n(\hat{\theta}_{n,\text{GEL}})\| + \|g_n(\hat{\theta}_{n,\text{GEL}})\| = o_p(1)$. Since $\lim_{n\to\infty}\bar{g}_n(\theta)$ is uniquely zero at θ_0, $\|\bar{g}_n(\theta)\|$ must be bounded away from zero outside of any neighborhood of θ_0. Therefore $\hat{\theta}_{n,\text{GEL}}$ must be inside any neighborhood of θ_0 with probability approaching one (w.p.a.1), i.e., $\hat{\theta}_{n,\text{GEL}} \xrightarrow{p} \theta_0$. As $g_n(\hat{\theta}_{n,\text{GEL}}) = O_p(n^{-1/2})$, Lemma 11 holds for $\bar{\theta}_n = \hat{\theta}_{n,\text{GEL}}$. Hence, $\hat{\lambda}_{n,\text{GEL}} = \arg\max_{\lambda\in\Lambda_n(\hat{\theta}_{n,\text{GEL}})}\frac{1}{n}\sum_{i=1}^n \rho(\lambda' g_{ni}(\hat{\theta}_{n,\text{GEL}}))$ exists w.p.a.1, and $\hat{\lambda}_{n,\text{GEL}} = O_p(n^{-1/2})$. □

Proof of Theorem 3.2. By Theorem 3.1, $\hat{\lambda}_{n,\text{GEL}} = O_p(n^{-1/2})$. Then by Lemma 10, $\max_{1\leq i\leq n}|\hat{\lambda}'_{n,\text{GEL}} g_{ni}(\hat{\theta}_{n,\text{GEL}})| \xrightarrow{p} 0$. Hence, the first order condition

$$\sum_{i=1}^n \rho_1(\hat{\lambda}'_{n,\text{GEL}} g_{ni}(\hat{\theta}_{n,\text{GEL}}))g_{ni}(\hat{\theta}_{n,\text{GEL}}) = 0$$

is satisfied w.p.a.1. By the implicit function theorem, there is a neighborhood of $\hat{\theta}_{n,\text{GEL}}$ where the solution $\lambda(\theta)$ to $\sum_{i=1}^n \rho_1(\lambda' g_{ni}(\theta))g_{ni}(\theta) = 0$ exists and is continuously differentiable. Then by the envelope theorem, the first order conditions for the GEL are

$$\sum_{i=1}^n \rho_1(\hat{\lambda}'_{n,\text{GEL}} g_{ni}(\hat{\theta}_{n,\text{GEL}}))G'_{ni}(\hat{\theta}_{n,\text{GEL}})\hat{\lambda}_{n,\text{GEL}} = 0 \text{ and } \sum_{i=1}^n \rho_1(\hat{\lambda}'_{n,\text{GEL}} g_{ni}(\hat{\theta}_{n,\text{GEL}}))g_{ni}(\hat{\theta}_{n,\text{GEL}}) = 0.$$

Applying the mean value theorem to these first order conditions, we have

$$0 = \begin{pmatrix} 0 \\ -\frac{1}{n}\sum_{i=1}^n g_{ni}(\theta_0) \end{pmatrix} + \Delta_n(\hat{\gamma}_{n,\text{GEL}} - \gamma_0),$$

[28] For SAR models, a "smoothed" (instead of pointwise) asymptotic expansion based on martingales in Mykland (1993) is shown in Jin and Lee (2015).

where

$$\Delta_n = \frac{1}{n}\sum_{i=1}^n \begin{pmatrix} \rho_2(\bar\lambda_n' g_{ni}(\bar\theta_n))G_{ni}'(\bar\theta_n)\bar\lambda_n\bar\lambda_n' G_{ni}(\bar\theta_n) + \rho_1(\bar\lambda_n' g_{ni}(\bar\theta_n))[G_{ni}^{(1)'}(\bar\theta_n)\bar\lambda_n, \ldots, G_{ni}^{(k_\theta)'}(\bar\theta_n)\bar\lambda_n] & * \\ \rho_2(\bar\lambda_n' g_{ni}(\bar\theta_n))g_{ni}(\bar\theta_n)\bar\lambda_n' G_{ni}(\bar\theta_n) + \rho_1(\bar\lambda_n' g_{ni}(\bar\theta_n))G_{ni}(\bar\theta_n) & \rho_2(\bar\lambda_n' g_{ni}(\bar\theta_n))g_{ni}(\bar\theta_n)g_{ni}'(\bar\theta_n)\end{pmatrix}$$

and $(\bar\theta_n', \bar\lambda_n')'$ is between $\hat\gamma_{n,\text{GEL}}$ and γ_0 elementwise. As $\max_{1\le i\le n}|\hat\lambda_{n,\text{GEL}}' g_{ni}(\hat\theta_{n,\text{GEL}})| \xrightarrow{p} 0$, by the twice continuous differentiability of $\rho(v)$, $\max_{1\le i\le n}|\rho_l(\bar\lambda_n' g_{ni}(\bar\theta_n)) + 1| = o_p(1)$ for $l = 1$ and 2. Then by Lemma 5 and the mean value theorem,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \rho_1(\bar\lambda_n' g_{ni}(\bar\theta_n))G_{ni}(\bar\theta_n) &= \frac{1}{n}\sum_{i=1}^n[\rho_1(\bar\lambda_n' g_{ni}(\bar\theta_n)) + 1]G_{ni}(\bar\theta_n) - \frac{1}{n}\sum_{i=1}^n G_{ni}(\theta_0) - \frac{1}{n}\sum_{i=1}^n\sum_{l=1}^{k_\theta} G_{ni}^{(l)}(\check\theta_n)(\bar\theta_{nl} - \theta_{0l}) \\ &= -\bar G_n + o_p(1),\end{aligned}$$

where $\check\theta_n$ lies between $\bar\theta_n$ and θ_0. Similarly, by Lemmas 6 and 7, $\frac{1}{n}\sum_{i=1}^n \rho_2(\bar\lambda_n' g_{ni}(\bar\theta_n))g_{ni}(\bar\theta_n)g_{ni}'(\bar\theta_n) = -\bar\Omega_n + o_p(1)$, $\frac{1}{n}\sum_{i=1}^n \rho_2(\bar\lambda_n' g_{ni}(\bar\theta_n))G_{ni}'(\bar\theta_n)\bar\lambda_n\bar\lambda_n' G_{ni}(\bar\theta_n) = o_p(1)$, $\frac{1}{n}\sum_{i=1}^n \rho_1(\bar\lambda_n' g_{ni}(\bar\theta_n))[G_{ni}^{(1)'}(\bar\theta_n)\bar\lambda_n, \ldots, G_{ni}^{(k_\theta)'}(\bar\theta_n)\bar\lambda_n] = o_p(1)$, and $\frac{1}{n}\sum_{i=1}^n \rho_2(\bar\lambda_n' g_{ni}(\bar\theta_n))g_{ni}(\bar\theta_n)\bar\lambda_n' G_{ni}(\bar\theta_n) = o_p(1)$. Thus, $\Delta_n = -\bar K_n + o_p(1)$, where $\bar K_n = \begin{pmatrix} 0 & \bar G_n' \\ \bar G_n & \bar\Omega_n\end{pmatrix}$. Hence,

$$\sqrt{n}(\hat\gamma_{n,\text{GEL}} - \gamma_0) = -\bar K_n^{-1}\begin{pmatrix} 0 \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n g_{ni}(\theta_0)\end{pmatrix} + o_p(1).$$

Since $\bar K_n^{-1} = \begin{pmatrix} -\bar\Sigma_n & \bar H_n \\ \bar H_n' & \bar D_n\end{pmatrix}$,

$$\sqrt{n}(\hat\gamma_{n,\text{GEL}} - \gamma_0) = -\begin{pmatrix}\bar H_n \\ \bar D_n\end{pmatrix}\frac{1}{\sqrt{n}}\sum_{i=1}^n g_{ni}(\theta_0) + o_p(1). \tag{15}$$

Then the asymptotic distribution of $\sqrt{n}(\hat\gamma_{n,\text{GEL}} - \gamma_0)$ follows by the central limit theorem in Kelejian and Prucha (2001, Theorem 1). □

Proof of Theorem 3.3. Since $\hat\lambda_{n,\text{GMM}} = -\Omega_n^{-1}(\tilde\theta_n)g_n(\hat\theta_{n,\text{GMM}}) = O_p(n^{-1/2})$, by Lemma 9, the first order condition (9) can be written as

$$0 = -\begin{pmatrix} G_n'(\hat\theta_{n,\text{GMM}})\hat\lambda_{n,\text{GMM}} \\ g_n(\hat\theta_{n,\text{GMM}}) + (\bar\Omega_n + n^{-1/2}\xi_n^\Omega)\hat\lambda_{n,\text{GMM}}\end{pmatrix} + O_p(n^{-3/2}). \tag{16}$$

By a second order Taylor expansion and Lemma 5,

$$0 = -\begin{pmatrix} 0 \\ g_n(\theta_0)\end{pmatrix} - K_n^\Omega(\theta_0)(\hat\gamma_{n,\text{GMM}} - \gamma_0) - \frac{1}{2}\sum_{j=1}^{k_\theta+k_g}(\hat\gamma_{nj,\text{GMM}} - \gamma_{0j})K_{nj}(\hat\gamma_{n,\text{GMM}} - \gamma_0) + O_p(n^{-3/2}),$$

where $K_n^\Omega(\theta) = \begin{pmatrix}[G_n^{(1)'}(\theta)\lambda, \ldots, G_n^{(k_\theta)'}(\theta)\lambda] & G_n'(\theta) \\ G_n(\theta) & \bar\Omega_n + n^{-1/2}\xi_n^\Omega\end{pmatrix}$, $K_{nj} = \begin{pmatrix} 0 & G_n^{(j)'}(\theta_0) \\ G_n^{(j)}(\theta_0) & 0\end{pmatrix}$ for $1 \le j \le k_\theta$, and $K_{nj} = \begin{pmatrix}[G_n^{(1)'}(\theta_0)e_{k_g,j-k_\theta}, \ldots, G_n^{(k_\theta)'}(\theta_0)e_{k_g,j-k_\theta}] & 0 \\ 0 & 0\end{pmatrix}$ for $k_\theta + 1 \le j \le k_\theta + k_g$. Let $\bar K_n = \begin{pmatrix} 0 & \bar G_n' \\ \bar G_n & \bar\Omega_n\end{pmatrix}$. Then,

$$\begin{aligned}\sqrt{n}(\hat\gamma_{n,\text{GMM}} - \gamma_0) &= -\bar K_n^{-1}\begin{pmatrix} 0 \\ \sqrt{n}g_n(\theta_0)\end{pmatrix} - \bar K_n^{-1}[K_n^\Omega(\theta_0) - \bar K_n]\sqrt{n}(\hat\gamma_{n,\text{GMM}} - \gamma_0) \\ &\quad - \frac{\sqrt{n}}{2}\sum_{j=1}^{k_\theta+k_g}\bar K_n^{-1}K_{nj}(\hat\gamma_{n,\text{GMM}} - \gamma_0)(\hat\gamma_{nj,\text{GMM}} - \gamma_{0j}) + O_p(n^{-1}).\end{aligned} \tag{17}$$

By (17), we have

$$\sqrt{n}(\hat{\gamma}_{n,\text{GMM}} - \gamma_0) = \xi_n + O_p(n^{-1/2}), \tag{18}$$

where $\xi_n = -\bar{K}_n^{-1}\binom{0}{\sqrt{n}g_n(\theta_0)} = O_p(1)$. Substituting (18) into the second and third terms of (17) yields $\sqrt{n}(\hat{\gamma}_{n,\text{GMM}} - \gamma_0) = \xi_n + n^{-1/2}\psi_n + O_p(n^{-1})$, where

$$\psi_n = -\bar{K}_n^{-1}\begin{pmatrix} 0 & \sqrt{n}(G_n' - \bar{G}_n') \\ \sqrt{n}(G_n - \bar{G}_n) & \xi_n^{\Omega} \end{pmatrix}\xi_n - \frac{1}{2}\sum_{j=1}^{k_\theta+k_g}\bar{K}_n^{-1}\bar{K}_{nj}\xi_n\xi_{nj} = O_p(1), \tag{19}$$

with $\bar{K}_{nj} = \begin{pmatrix} 0 & \bar{G}_n^{(j)'} \\ \bar{G}_n^{(j)} & 0 \end{pmatrix}$ for $1 \le j \le k_\theta$, and $\bar{K}_{nj} = \begin{pmatrix} [\bar{G}_n^{(1)'}e_{k_g,j-k_\theta}, \ldots, \bar{G}_n^{(k_\theta)'}e_{k_g,j-k_\theta}] & 0 \\ 0 & 0 \end{pmatrix}$ for $k_\theta + 1 \le j \le k_\theta + k_g$. □

Proof of Theorem 3.4. Let $v_{ni}(\gamma) = \lambda' g_{ni}(\theta)$, $h_{ni}(\gamma) = \frac{\partial v_{ni}(\gamma)}{\partial \gamma} = \binom{G_{ni}'(\theta)\lambda}{g_{ni}(\theta)}$, and $m_{ni}(\gamma) = \rho_1(v_{ni}(\gamma))h_{ni}(\gamma)$. Then the first order condition of the GEL estimator is:

$$\frac{1}{n}\sum_{i=1}^{n} m_{ni}(\hat{\gamma}_{n,\text{GEL}}) = 0. \tag{20}$$

Expressions for derivatives of $h_{ni}(\gamma)$ and $m_{ni}(\gamma)$ are provided in the supplementary file. By a second order Taylor expansion of (20),

$$0 = \frac{1}{n}\sum_{i=1}^{n} m_{ni}(\gamma_0) + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial m_{ni}(\gamma_0)}{\partial \gamma'}(\hat{\gamma}_{n,\text{GEL}} - \gamma_0) + \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{k_\theta+k_g}(\hat{\gamma}_{nj,\text{GEL}} - \gamma_{0j})\frac{\partial^2 m_{ni}(\gamma_0)}{\partial \gamma_j \partial \gamma'}(\hat{\gamma}_{n,\text{GEL}} - \gamma_0)$$
$$+ O_p(\|\hat{\gamma}_{n,\text{GEL}} - \gamma_0\|^3),$$

where the order of the remainder is derived by using the Liptchitz hypothesis of $\rho(v)$ and Lemma 5. Hence,

$$\begin{aligned} &\sqrt{n}(\hat{\gamma}_{n,\text{GEL}} - \gamma_0) \\ &= -\Big[\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\big(\frac{\partial m_{ni}(\gamma_0)}{\partial \gamma'}\big)\Big]^{-1}\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m_{ni}(\gamma_0) + \frac{1}{\sqrt{n}}\Big[\sum_{i=1}^{n}\frac{\partial m_{ni}(\gamma_0)}{\partial \gamma'} - \mathrm{E}\big(\sum_{i=1}^{n}\frac{\partial m_{ni}(\gamma_0)}{\partial \gamma'}\big)\Big](\hat{\gamma}_{n,\text{GEL}} - \gamma_0) \\ &\quad + \frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{k_\theta+k_g}(\hat{\gamma}_{nj,\text{GEL}} - \gamma_{0j})\frac{\partial^2 m_{ni}(\gamma_0)}{\partial \gamma_j \partial \gamma'}(\hat{\gamma}_{n,\text{GEL}} - \gamma_0)\Big\} + O_p(\sqrt{n}\|\hat{\gamma}_{n,\text{GEL}} - \gamma_0\|^3). \end{aligned} \tag{21}$$

Thus,

$$\sqrt{n}(\hat{\gamma}_{n,\text{GEL}} - \gamma_0) = \xi_n + O_p(n^{-1/2}), \tag{22}$$

where $\xi_n = -[\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(\frac{\partial m_{ni}(\gamma_0)}{\partial \gamma'})]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m_{ni}(\gamma_0) = -\bar{K}_n^{-1}\binom{0}{\sqrt{n}g_n(\theta_0)} = -\binom{\bar{H}_n}{\bar{D}_n}\sqrt{n}g_n(\theta_0)$. Substituting (22) into the second and third terms of (21) yields $\sqrt{n}(\hat{\gamma}_{n,\text{GEL}}-\gamma_0)=\xi_n+n^{-1/2}\psi_n+O_p(n^{-1})$, where

$$\begin{aligned}
\psi_n &= -\Big[\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\big(\frac{\partial m_{ni}(\gamma_0)}{\partial\gamma'}\big)\Big]^{-1}\Big\{\frac{1}{\sqrt{n}}\Big[\sum_{i=1}^{n}\frac{\partial m_{ni}(\gamma_0)}{\partial\gamma'}-\mathrm{E}\big(\sum_{i=1}^{n}\frac{\partial m_{ni}(\gamma_0)}{\partial\gamma'}\big)\Big]\xi_n+\frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{k_\theta+k_g}\xi_{nj}\Big[\mathrm{E}\Big(\frac{\partial^2 m_{ni}(\gamma_0)}{\partial\gamma_j\partial\gamma'}\Big)\Big]\xi_n\Big\} \\
&= -\sqrt{n}\bar{K}_n^{-1}\begin{pmatrix}0 & G_n'-\bar{G}_n' \\ G_n-\bar{G}_n & \Omega_n-\bar{\Omega}_n\end{pmatrix}\xi_n+\frac{1}{2n}\bar{K}_n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{k_\theta+k_g}\xi_{nj}\Big[\mathrm{E}\Big(\frac{\partial^2 m_{ni}(\gamma_0)}{\partial\gamma_j\partial\gamma'}\Big)\Big]\xi_n \\
&= -\sqrt{n}\bar{K}_n^{-1}\begin{pmatrix}0 & G_n'-\bar{G}_n' \\ G_n-\bar{G}_n & \Omega_n-\bar{\Omega}_n\end{pmatrix}\xi_n-\frac{1}{2n}\bar{K}_n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{k_\theta}\xi_{nj}\,\mathrm{E}\begin{pmatrix}0 & G_{ni}^{(j)\prime} \\ G_{ni}^{(j)} & g_{ni}^{(j)}g_{ni}'+g_{ni}g_{ni}^{(j)\prime}\end{pmatrix}\xi_n \\
&\quad -\frac{1}{2n}\bar{K}_n^{-1}\sum_{i=1}^{n}\sum_{s=1}^{k_g}\xi_{n,k_\theta+s}\,\mathrm{E}\begin{pmatrix}[G_{ni}^{(1)\prime}e_{k_g,s},\dots,G_{ni}^{(k_\theta)\prime}e_{k_g,s}] & G_{ni}'e_{k_g,s}g_{ni}'+g_{nis}G_{ni}' \\ g_{ni}e_{k_g,s}'G_{ni}+g_{nis}G_{ni} & -\rho_3 g_{nis}g_{ni}g_{ni}'\end{pmatrix}\xi_n
\end{aligned}\tag{23}$$

with g_{nis} being the sth element of g_{ni}. □

Proof of Theorem 3.5. Note that $\xi_n=-\binom{\bar{H}_n}{\bar{D}_n}\sqrt{n}g_n(\theta_0)$ and $\mathrm{E}(\xi_n\xi_n')=\mathrm{diag}(\bar{\Sigma}_n,\bar{D}_n)$. Then by Theorem 3.3, with ψ_n in (19),

$$\begin{aligned}
\frac{1}{n}\mathrm{E}(\psi_n) &= \bar{K}_n^{-1}\mathrm{E}\Big[\begin{pmatrix}0 & G_n' \\ G_n & n^{-1/2}\xi_n^{\Omega}\end{pmatrix}\binom{\bar{H}_n}{\bar{D}_n}g_n\Big]-\frac{1}{2n}\sum_{j=1}^{k_\theta}\bar{K}_n^{-1}\begin{pmatrix}0 & \bar{G}_n^{(j)\prime} \\ \bar{G}_n^{(j)} & 0\end{pmatrix}\mathrm{diag}(\bar{\Sigma}_n,\bar{D}_n)e_{k_\theta+k_g,j} \\
&\quad -\frac{1}{2n}\sum_{j=k_\theta+1}^{k_\theta+k_g}\bar{K}_n^{-1}\begin{pmatrix}[\bar{G}_n^{(1)\prime}e_{k_g,j-k_\theta},\dots,\bar{G}_n^{(k_\theta)\prime}e_{k_g,j-k_\theta}] & 0 \\ 0 & 0\end{pmatrix}\mathrm{diag}(\bar{\Sigma}_n,\bar{D}_n)e_{k_\theta+k_g,j} \\
&= \bar{K}_n^{-1}\begin{pmatrix}\mathrm{E}(G_n'\bar{D}_n g_n) \\ \mathrm{E}(G_n\bar{H}_n g_n+n^{-1/2}\xi_n^{\Omega}\bar{D}_n g_n)-\frac{1}{2n}\sum_{j=1}^{k_\theta}\bar{G}_n^{(j)}\bar{\Sigma}_n e_{k_\theta,j}\end{pmatrix}.
\end{aligned}$$

Since $\xi_n^{\Omega}=\sqrt{n}(\Omega_n-\bar{\Omega}_n)+\sum_{j=1}^{k_\theta}\frac{1}{n}\sum_{i=1}^{n}[\mathrm{E}(g_{ni}g_{ni}^{(j)\prime}+g_{ni}^{(j)}g_{ni}')]\tilde{\xi}_{nj}$ by Lemma 9, where $\tilde{\xi}_n=-\binom{\bar{H}_n^J}{\bar{D}_n^J}\sqrt{n}g_n(\theta_0)$ by Lemma 8, $\mathrm{E}(n^{-1/2}\xi_n^{\Omega}\bar{D}_n g_n)=\mathrm{E}(\Omega_n\bar{D}_n g_n)-\sum_{j=1}^{k_\theta}\frac{1}{n^2}\sum_{i=1}^{n}[\mathrm{E}(g_{ni}g_{ni}^{(j)\prime}+g_{ni}^{(j)}g_{ni}')]\bar{D}_n\bar{\Omega}_n\bar{H}_n^{J\prime}e_{k_\theta,j}$. Since $\bar{K}_n^{-1}=\begin{pmatrix}-\bar{\Sigma}_n & \bar{H}_n \\ \bar{H}_n' & \bar{D}_n\end{pmatrix}$ and $\bar{D}_n\bar{\Omega}_n\bar{H}_n^{J\prime}=\bar{H}_n^{J\prime}-\bar{H}_n'$, the leading bias of the GMM estimator $\hat{\theta}_n$ is the first k_θ components of $\frac{1}{n}\mathrm{E}(\psi_n)$, which is

$$\begin{aligned}
&-\bar{\Sigma}_n\mathrm{E}(G_n'\bar{D}_n g_n)+\bar{H}_n\mathrm{E}(G_n\bar{H}_n g_n)+\bar{H}_n\mathrm{E}(\Omega_n\bar{D}_n g_n) \\
&\quad -\sum_{j=1}^{k_\theta}\frac{1}{n^2}\sum_{i=1}^{n}\bar{H}_n[\mathrm{E}(g_{ni}g_{ni}^{(j)\prime}+g_{ni}^{(j)}g_{ni}')](\bar{H}_n^J-\bar{H}_n)'e_{k_\theta,j}-\frac{1}{2n}\sum_{j=1}^{k_\theta}\bar{H}_n\bar{G}_n^{(j)}\bar{\Sigma}_n e_{k_\theta,j}.
\end{aligned}$$

□

Proof of Theorem 3.6. As in the proof of Theorem 3.5, $\mathrm{E}(\xi_n\xi_n')=\mathrm{diag}(\bar{\Sigma}_n,\bar{D}_n)$. Then by Theorem 3.4, with ψ_n in (23),

$$\begin{aligned}
\frac{1}{n}\mathrm{E}(\psi_n) &= \bar{K}_n^{-1}\mathrm{E}\Big[\begin{pmatrix}0 & G_n'-\bar{G}_n' \\ G_n-\bar{G}_n & \Omega_n-\bar{\Omega}_n\end{pmatrix}\binom{\bar{H}_n}{\bar{D}_n}g_n(\theta_0)\Big] \\
&\quad -\frac{1}{2n}\bar{K}_n^{-1}\sum_{j=1}^{k_\theta}\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}0 & G_{ni}^{(j)\prime} \\ G_{ni}^{(j)} & g_{ni}^{(j)}g_{ni}'+g_{ni}g_{ni}^{(j)\prime}\end{pmatrix}\mathrm{diag}(\bar{\Sigma}_n,\bar{D}_n)e_{k_\theta+k_g,j} \\
&\quad -\frac{1}{2n}\bar{K}_n^{-1}\sum_{s=1}^{k_g}\sum_{i=1}^{n}\begin{pmatrix}[G_{ni}^{(1)\prime}e_{k_g,s},\dots,G_{ni}^{(k_\theta)\prime}e_{k_g,s}] & G_{ni}'e_{k_g,s}g_{ni}'+g_{nis}G_{ni}' \\ g_{ni}e_{k_g,s}'G_{ni}+g_{nis}G_{ni} & -\rho_3 g_{nis}g_{ni}g_{ni}'\end{pmatrix}\mathrm{diag}(\bar{\Sigma}_n,\bar{D}_n)e_{k_\theta+k_g,k_\theta+s}
\end{aligned}$$

$$= \bar{K}_n^{-1}\begin{pmatrix} \mathrm{E}(G_n'\bar{D}_n g_n) \\ \mathrm{E}(G_n\bar{H}_n g_n) + \mathrm{E}(\Omega_n\bar{D}_n g_n)\end{pmatrix} - \frac{1}{2n}\bar{K}_n^{-1}\sum_{j=1}^{k_\theta}\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 0 \\ \bar{G}_{ni}^{(j)}\bar{\Sigma}_n e_{k_\theta,j}\end{pmatrix}$$

$$- \frac{1}{2n}\bar{K}_n^{-1}\sum_{s=1}^{k_g}\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \mathrm{E}(G_{ni}'e_{k_g,s}g_{ni}' + g_{nis}G_{ni}')\bar{D}_n e_{k_g,s} \\ -\rho_3\,\mathrm{E}(g_{nis}g_{ni}g_{ni}')\bar{D}_n e_{k_g,s}\end{pmatrix}.$$

The leading bias of the GEL estimator $\hat{\theta}_{n,\mathrm{GEL}}$ is the first k_θ elements of $\mathrm{E}(\psi_n)$, and, as $\bar{K}_n^{-1} = \begin{pmatrix} -\bar{\Sigma}_n & \bar{H}_n \\ \bar{H}_n' & \bar{D}_n\end{pmatrix}$, it is

$$- \bar{\Sigma}_n\,\mathrm{E}(G_n'\bar{D}_n g_n) + \frac{1}{2n}\sum_{s=1}^{k_g}\frac{1}{n}\sum_{i=1}^{n}\bar{\Sigma}_n\,\mathrm{E}(G_{ni}'e_{k_g,s}g_{ni}' + g_{nis}G_{ni}')\bar{D}_n e_{k_g,s} + \bar{H}_n\,\mathrm{E}(G_n\bar{H}_n g_n) + \bar{H}_n\,\mathrm{E}(\Omega_n\bar{D}_n g_n)$$

$$- \frac{1}{2n}\sum_{j=1}^{k_\theta}\frac{1}{n}\sum_{i=1}^{n}\bar{H}_n\bar{G}_{ni}^{(j)}\bar{\Sigma}_n e_{k_\theta,j} + \frac{1}{2n}\rho_3\sum_{s=1}^{k_g}\frac{1}{n}\sum_{i=1}^{n}\bar{H}_n\,\mathrm{E}(g_{nis}g_{ni}g_{ni}')\bar{D}_n e_{k_g,s}.$$

Note that $\sum_{s=1}^{k_g}\mathrm{E}(G_{ni}'e_{k_g,s}g_{ni}')\bar{D}_n e_{k_g,s} = \sum_{s=1}^{k_g}\mathrm{E}(G_{ni}'e_{k_g,s}e_{k_g,s}'\bar{D}_n g_{ni}) = \mathrm{E}(G_{ni}'\bar{D}_n g_{ni})$,

$$\sum_{s=1}^{k_g}\mathrm{E}(g_{nis}G_{ni}')\bar{D}_n e_{k_g,s} = \sum_{s=1}^{k_g}\mathrm{E}(G_{ni}'\bar{D}_n g_{nis}e_{k_g,s}) = \mathrm{E}(G_{ni}'\bar{D}_n g_{ni}),$$

and $\sum_{s=1}^{k_g}\mathrm{E}(g_{nis}g_{ni}g_{ni}')\bar{D}_n e_{k_g,s} = \sum_{s=1}^{k_g}\mathrm{E}(g_{ni}g_{ni}'\bar{D}_n e_{k_g,s}g_{nis}) = \mathrm{E}(g_{ni}g_{ni}'\bar{D}_n g_{ni})$. Thus, the bias is

$$- \bar{\Sigma}_n\,\mathrm{E}(G_n'\bar{D}_n g_n) + \frac{1}{n^2}\bar{\Sigma}_n\sum_{i=1}^{n}\mathrm{E}(G_{ni}'\bar{D}_n g_{ni}) + \bar{H}_n\,\mathrm{E}(G_n\bar{H}_n g_n) + \bar{H}_n\,\mathrm{E}(\Omega_n\bar{D}_n g_n) - \frac{1}{2n}\sum_{j=1}^{k_\theta}\bar{H}_n\bar{G}_n^{(j)}\bar{\Sigma}_n e_{k_\theta,j}$$

$$+ \frac{\rho_3}{2n^2}\sum_{i=1}^{n}\bar{H}_n\,\mathrm{E}(g_{ni}g_{ni}'\bar{D}_n g_{ni}). \qquad \square$$

Proof of Theorem 4.1. The unconstrained GEL estimator $\hat{\lambda}_n$ is a maximizer, so $\frac{1}{n}\sum_{i=1}^{n}\rho_1(\hat{\lambda}_n' g_{ni}(\hat{\theta}_n))g_{ni}(\hat{\theta}_n) = 0$. As $\rho(0) = \frac{1}{n}\sum_{i=1}^{n}\rho(0\cdot g_{ni}(\hat{\theta}_n))$, by a first order Taylor expansion of $\frac{1}{n}\sum_{i=1}^{n}\rho(0\cdot g_{ni}(\hat{\theta}_n))$ at $\hat{\lambda}_n$, and using $\rho_2(0) = -1$, Lemma 5 and (15) successively, we have

$$\begin{aligned} -2n\Big[\rho(0) - \frac{1}{n}\sum_{i=1}^{n}\rho(\hat{\lambda}_n' g_{ni}(\hat{\theta}_n))\Big] &= -\sum_{i=1}^{n}\rho_2(\check{\lambda}_n' g_{ni}(\hat{\theta}_n))\hat{\lambda}_n' g_{ni}(\hat{\theta}_n)g_{ni}'(\hat{\theta}_n)\hat{\lambda}_n \\ &= \sum_{i=1}^{n}\hat{\lambda}_n' g_{ni}(\hat{\theta}_n)g_{ni}'(\hat{\theta}_n)\hat{\lambda}_n + o_p(1) \\ &= n\hat{\lambda}_n'\bar{\Omega}_n\hat{\lambda}_n + o_p(1) \\ &= [\bar{\Omega}_n^{-1/2}\sqrt{n}g_n(\theta_0)]'\bar{\Omega}_n^{1/2}\bar{D}_n\bar{\Omega}_n^{1/2}[\bar{\Omega}_n^{-1/2}\sqrt{n}g_n(\theta_0)] + o_p(1), \end{aligned} \tag{24}$$

where $\check{\lambda}_n$ lies between 0 and $\hat{\lambda}_n$, because $\sqrt{n}\hat{\lambda}_n = -\bar{D}_n\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_{ni}(\theta_0) + o_p(1)$ and $\bar{D}_n\bar{\Omega}_n\bar{D}_n' = \bar{D}_n$.

For the restricted GEL estimators $\dot{\theta}_n$ and $\dot{\lambda}_n$, the results in Theorem 3.1 hold under the null by similar arguments. In particular, $\dot{\theta}_n = \theta_0 + o_p(1)$, and $\dot{\lambda}_n = O_p(n^{-1/2})$. With these results, as in the proof of Theorem 3.2, we can apply the mean value theorem to the first order conditions of the restricted GEL estimation

$$\sum_{i=1}^{n}\rho_1(\dot{\lambda}_n' g_{ni}(\dot{\theta}_n))\frac{\partial g_{ni}'(\dot{\theta}_n)}{\partial\phi}\dot{\lambda}_n = 0 \text{ and } \sum_{i=1}^{n}\rho_1(\dot{\lambda}_n' g_{ni}(\dot{\theta}_n))g_{ni}(\dot{\theta}_n) = 0$$

to obtain

$$\sqrt{n}(\dot{\delta}_n - \delta_0) = -\begin{pmatrix}\bar{H}_{n\phi} \\ \bar{D}_{n\phi}\end{pmatrix}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g_{ni}(\theta_0) + o_p(1).$$

where $\dot{\delta}_n = (\dot{\phi}_n', \dot{\lambda}_n')'$, $\bar{G}_{n\phi} = \mathrm{E}(\frac{\partial g_n(\theta_0)}{\partial \phi'})$, $\bar{\Sigma}_{n\phi} = (\bar{G}_{n\phi}'\bar{\Omega}_n^{-1}\bar{G}_{n\phi})^{-1}$, $\bar{H}_{n\phi} = \bar{\Sigma}_{n\phi}\bar{G}_{n\phi}'\bar{\Omega}_n^{-1}$, and $\bar{D}_{n\phi} = \bar{\Omega}_n^{-1} - \bar{\Omega}_n^{-1}\bar{G}_{n\phi}\bar{\Sigma}_{n\phi}\bar{G}_{n\phi}'\bar{\Omega}_n^{-1}$. Then we can obtain the following expression analogous to (24) above:

$$-2n\Big[\rho(0) - \frac{1}{n}\sum_{i=1}^n \rho(\dot{\lambda}_n' g_{ni}(\dot{\theta}_n))\Big] = [\bar{\Omega}_n^{-1/2}\sqrt{n}g_n(\theta_0)]'\bar{\Omega}_n^{1/2}\bar{D}_{n\phi}\bar{\Omega}_n^{1/2}[\bar{\Omega}_n^{-1/2}\sqrt{n}g_n(\theta_0)] + o_p(1). \quad (25)$$

Combining (24) and (25) yields

$$2\Big[\sum_{i=1}^n \rho(\dot{\lambda}_n' g_{ni}(\dot{\theta}_n)) - \sum_{i=1}^n \rho(\hat{\lambda}_n' g_{ni}(\hat{\theta}_n))\Big] = [\bar{\Omega}_n^{-1/2}\sqrt{n}g_n(\theta_0)]'\bar{\Omega}_n^{1/2}(\bar{D}_{n\phi} - \bar{D}_n)\bar{\Omega}_n^{1/2}[\bar{\Omega}_n^{-1/2}\sqrt{n}g_n(\theta_0)] + o_p(1). \quad (26)$$

Since $\bar{G}_{n\phi}$ is a submatrix of $\bar{G}_n$ and $\mathrm{plim}_{n\to\infty}\bar{G}_n$ has full rank,

$$\begin{aligned}\bar{\Omega}_n^{1/2}(\bar{D}_{n\phi} - \bar{D}_n)\bar{\Omega}_n^{1/2} &= \bar{\Omega}_n^{-1/2}\bar{G}_n(\bar{G}_n'\bar{\Omega}_n^{-1}\bar{G}_n)^{-1}\bar{G}_n'\bar{\Omega}_n^{-1/2} - \bar{\Omega}_n^{-1/2}\bar{G}_{n\phi}(\bar{G}_{n\phi}'\bar{\Omega}_n^{-1}\bar{G}_{n\phi})^{-1}\bar{G}_{n\phi}'\bar{\Omega}_n^{-1/2} \\ &= \mathbb{M}_n\bar{\Omega}_n^{-1/2}\bar{G}_{n\alpha}(\bar{G}_{n\alpha}'\bar{\Omega}_n^{-1/2}\mathbb{M}_n\bar{\Omega}_n^{-1/2}\bar{G}_{n\alpha})^{-1}\bar{G}_{n\alpha}'\bar{\Omega}_n^{-1/2}\mathbb{M}_n\end{aligned}$$

is a projection matrix with rank k_α, where $\mathbb{M}_n = I_{k_g} - \bar{\Omega}_n^{-1/2}\bar{G}_{n\phi}(\bar{G}_{n\phi}'\bar{\Omega}_n^{-1}\bar{G}_{n\phi})^{-1}\bar{G}_{n\phi}'\bar{\Omega}_n^{-1/2}$ (Ruud, 2000, p. 60, (3.13)). Hence the theorem follows. □

Proof of Theorem 4.2. With the Pitman drift in the theorem, we still have the consistency that $\hat{\theta}_n = \theta_0 + o_p(1)$ and $\dot{\theta}_n = \theta_0 + o_p(1)$. This is because $V_n(\theta)$ is quadratic in θ and linear in V_n by the reduced form of Y_n, which implies that $g_n(\theta)$ can be expanded as a polynomial of $\theta_n = (\alpha_n', \phi_0)'$. Then under the Pitman drift, Lemmas 3–7 and 10–12 all hold by similar arguments. Hence, as in the proof of Theorem 4.1, we have (26). By the mean value theorem, $\sqrt{n}g_n(\theta_0) = \sqrt{n}g_n(\theta_n) + \frac{\partial g_n(\check{\theta}_n)}{\partial \alpha'}\sqrt{n}(\alpha_0 - \alpha_n) = \sqrt{n}g_n(\theta_n) - \bar{G}_{n\alpha}d_\alpha + o_p(1)$, where $\check{\theta}_n$ lies between θ_0 and θ_n elementwise. Under the Pitman drift, $\sqrt{n}g_n(\theta_n) = \frac{1}{\sqrt{n}}[V_n'P_{1n}V_n - \mathrm{E}(V_n'P_{1n}V_n), \ldots, V_n'P_{k_pn}V_n - \mathrm{E}(V_n'P_{k_pn}V_n), V_n'Q_n]' \xrightarrow{d} N(0, \lim_{n\to\infty}\bar{\Omega}_n)$. Since $\bar{D}_n\bar{G}_{n\alpha} = 0$ and $(\bar{G}_{n\alpha}d_\alpha)'(\bar{D}_{n\phi} - \bar{D}_n)\bar{G}_{n\alpha}d_\alpha = (\bar{G}_{n\alpha}d_\alpha)'\bar{D}_{n\phi}\bar{G}_{n\alpha}d_\alpha$, the theorem holds by (26). □

Proof of Theorem 4.3. The asymptotic distribution follows by (24) in the proof of Theorem 4.1. Because $\bar{\Omega}_n^{1/2}\bar{D}_n\bar{\Omega}_n^{1/2}$ is a projection matrix with rank $(k_g - k_\theta)$ and $\bar{\Omega}_n^{-1/2}\sqrt{n}g_n(\theta_0)$ is asymptotically standard multivariate normal, $-2n[\rho(0) - \frac{1}{n}\sum_{i=1}^n \rho(\hat{\lambda}_n' g_{ni}(\hat{\theta}_n))] \xrightarrow{d} \chi^2(k_g - k_\theta)$. □

Proof of Theorem 4.4. Explicitly, $\frac{1}{n}\hat{V}_n'M_{jn}\hat{V}_n = \frac{1}{n}V_n'M_{jn}V_n - \frac{1}{n}V_n'(M_{jn} + M_{jn}')\mathbb{P}_nV_n + \frac{1}{n}V_n'\mathbb{P}_nM_{jn}\mathbb{P}_nV_n$, where $\mathbb{P}_n = X_n(X_n'X_n)^{-1}X_n'$. Note that $\frac{1}{n}V_n'(M_{jn}+M_{jn}')\mathbb{P}_nV_n = \frac{1}{n\sqrt{n}}V_n'(M_{jn}+M_{jn}')X_n(\frac{1}{n}X_n'X_n)^{-1}\frac{1}{\sqrt{n}}X_n'V_n = O_p(n^{-1})$. Similarly, $\frac{1}{n}V_n'\mathbb{P}_nM_{jn}\mathbb{P}_nV_n = O_p(n^{-1})$. As $\frac{1}{n}V_n'M_{jn}V_n$ has mean zero under both homoskedasticity and unknown heteroskedasticity, $\frac{1}{n}\hat{V}_n'M_{jn}\hat{V}_n = \frac{1}{n}V_n'M_{jn}V_n + O_p(n^{-1}) = O_p(n^{-1/2})$. It follows that $\hat{g}_n = g_n + O_p(n^{-1}) = O_p(n^{-1/2})$. Then by Lemma 11, $\hat{\lambda}_n = \arg\max_{\lambda\in\hat{\Lambda}_n}\sum_{i=1}^n \rho(\lambda'\hat{g}_{ni})$ exists w.p.a.1, and its first order condition is $\sum_{i=1}^n \rho_1(\hat{\lambda}_n'\hat{g}_{ni})\hat{g}_{ni} = 0$. Applying the mean value theorem to this first order condition at $\lambda = 0$, we have $0 = \sum_{i=1}^n \rho_1(0)\hat{g}_{ni} + \sum_{i=1}^n \rho_2(\check{\lambda}_n'\hat{g}_{ni})\hat{g}_{ni}\hat{g}_{ni}'\hat{\lambda}_n$, where $\check{\lambda}_n$ lies between 0 and $\hat{\lambda}_n$ elementwise. Then, because $\sqrt{n}\hat{g}_n = \sqrt{n}g_n + O_p(n^{-1/2})$,

$$\begin{aligned}\sqrt{n}\hat{\lambda}_n &= \Big[\frac{1}{n}\sum_{i=1}^n \rho_2(\check{\lambda}_n'\hat{g}_{ni})\hat{g}_{ni}\hat{g}_{ni}'\Big]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_{ni} = -\Big[\frac{1}{n}\sum_{i=1}^n \mathrm{E}(g_{ni}g_{ni}')\Big]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n g_{ni} + o_p(1) \\ &\xrightarrow{d} N\Big(0, \lim_{n\to\infty}\Big[\frac{1}{n}\sum_{i=1}^n \mathrm{E}(g_{ni}g_{ni}')\Big]^{-1}\Big),\end{aligned}$$

where the second equality holds because $\frac{1}{n}\sum_{i=1}^n \rho_2(\check{\lambda}_n'\hat{g}_{ni})\hat{g}_{ni}\hat{g}_{ni}' = -\frac{1}{n}\sum_{i=1}^n \hat{g}_{ni}\hat{g}_{ni}' + o_p(1)$ as in the proof of Theorem 3.2 and $\frac{1}{n}\sum_{i=1}^n \hat{g}_{ni}\hat{g}_{ni}' = \frac{1}{n}\sum_{i=1}^n \mathrm{E}(g_{ni}g_{ni}') + o_p(1)$ by Lemma 3. Because $\rho(0) = \frac{1}{n}\sum_{i=1}^n \rho(0\cdot\hat{g}_{ni})$, by a first order Taylor expansion of $\frac{1}{n}\sum_{i=1}^n \rho(0\cdot\hat{g}_{ni})$ at $\hat{\lambda}_n$ and using the first order condition of $\hat{\lambda}_n$, $\rho(0) = \frac{1}{n}\sum_{i=1}^n \rho(\hat{\lambda}_n'\hat{g}_{ni}) + \frac{1}{2n}\sum_{i=1}^n \rho_2(\check{\lambda}_n'\hat{g}_{ni})\hat{\lambda}_n'\hat{g}_{ni}\hat{g}_{ni}'\hat{\lambda}_n$, where $\check{\lambda}_n$ lies between 0 and $\hat{\lambda}_n$. Hence,

$$2n\Big[\frac{1}{n}\sum_{i=1}^n \rho(\hat{\lambda}_n'\hat{g}_{ni})-\rho(0)\Big] = -(\sqrt{n}\hat{\lambda}_n)'\frac{1}{n}\sum_{i=1}^n \rho_2(\check{\lambda}_n'\hat{g}_{ni})\hat{g}_{ni}\hat{g}_{ni}'\sqrt{n}\hat{\lambda}_n = (\sqrt{n}\hat{\lambda}_n)'\frac{1}{n}\sum_{i=1}^n \mathrm{E}(g_{ni}g_{ni}')\sqrt{n}\hat{\lambda}_n + o_p(1) \xrightarrow{d} \chi^2(q). \tag{27}$$

□

Proof of Theorem 4.5. Let $\theta = (\tau', \beta')'$, $\theta_0 = (0, \beta_0')'$, $\theta_n = (n^{-1/2}d_\tau', \beta_0')'$, $\hat{\theta}_n = (0, \hat{\beta}_n')'$, and $V_n(\theta) = R_n(\tau)(Y_n - X_n\beta)$, where $\hat{\beta}_n = (X_n'X_n)^{-1}X_n'Y_n$ is the OLS estimate. Then, as in the proof of Theorem 4.4,

$$\begin{aligned}\sqrt{n}\hat{\lambda}_n &= \Big[\frac{1}{n}\sum_{i=1}^n \rho_1(\check{\lambda}_n'\hat{g}_{ni})\hat{g}_{ni}\hat{g}_{ni}'\Big]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_{ni} \\ &= -\Big[\frac{1}{n}\sum_{i=1}^n \mathrm{E}(g_{ni}g_{ni}')\Big]^{-1}\frac{1}{\sqrt{n}}[V_n'(\theta_0)M_{1n}V_n(\theta_0), \ldots, V_n'(\theta_0)M_{qn}V_n(\theta_0)]' + o_p(1).\end{aligned}$$

By the mean value theorem,

$$\frac{1}{\sqrt{n}}V_n'(\theta_0)M_{jn}V_n(\theta_0) = \frac{1}{\sqrt{n}}V_n'(\theta_n)M_{jn}V_n(\theta_n) + \frac{1}{n}\frac{\partial[V_n'(\check{\theta}_n)M_{jn}V_n(\check{\theta}_n)]}{\partial\tau'}\sqrt{n}(\tau_0 - \tau_n),$$

where $\check{\theta}_n$ lies between θ_0 and θ_n. Since $\frac{1}{\sqrt{n}}V_n'(\theta_n)M_{jn}V_n(\theta_n) = \frac{1}{\sqrt{n}}V_n'M_{jn}V_n$, and $\frac{1}{n}\frac{\partial[V_n'(\check{\theta}_n)M_{jn}V_n(\check{\theta}_n)]}{\partial\tau'} = -\frac{1}{n}\mathrm{E}[V_n'(M_{jn}+M_{jn}')M_{1n}V_n, \ldots, V_n'(M_{jn}+M_{jn}')M_{qn}V_n] + o_p(1)$, the result in the theorem follows by the expansion in (27). □

Proof of Theorem 4.6. By the mean value theorem and Lemma 5,

$$\begin{aligned}\sqrt{n}g_n(\hat{\theta}_n) &= \sqrt{n}g_{1n}(\theta_0) - \frac{\partial g_{1n}(\hat{\theta}_n)}{\partial\phi'}\Big(\frac{\partial g_{2n}(\hat{\theta}_n)}{\partial\phi'}\Big)^{-1}\sqrt{n}g_{2n}(\theta_0) \\ &\quad + \frac{\partial g_{1n}(\check{\theta}_n)}{\partial\phi'}\sqrt{n}(\hat{\phi}_n - \phi_0) - \frac{\partial g_{1n}(\hat{\theta}_n)}{\partial\phi'}\Big(\frac{\partial g_{2n}(\hat{\theta}_n)}{\partial\phi'}\Big)^{-1}\frac{\partial g_{2n}(\check{\theta}_n)}{\partial\phi'}\sqrt{n}(\hat{\phi}_n - \phi_0) \\ &= \sqrt{n}g_{1n}(\theta_0) - \frac{\partial g_{1n}(\theta_0)}{\partial\phi'}\Big(\frac{\partial g_{2n}(\theta_0)}{\partial\phi'}\Big)^{-1}\sqrt{n}g_{2n}(\theta_0) + o_p(1),\end{aligned}$$

where $\check{\theta}_n$ lies between $\hat{\theta}_n$ and θ_0 elementwise. Thus $\sqrt{n}g_n(\hat{\theta}_n)$ has the same asymptotic distribution as $\sqrt{n}g_n(\theta_0)$. The rest of the proof is similar to that for Theorem 4.4. □

Proof of Theorem 4.7. We only prove the consistency of $\hat{\vartheta}_n$, as the rest of the proof is similar to that of Theorem 4.1 on tests of parameter restrictions.

If $\hat{Y}_n = \mathcal{S}_n^{-1}(\hat{\boldsymbol{\kappa}}_n)\mathcal{X}_n\hat{\boldsymbol{\beta}}_n$, let $Y_n^* = \mathcal{S}_n^{-1}(\boldsymbol{\kappa}_n^*)\mathcal{X}_n\boldsymbol{\beta}_n^*$; if $\hat{Y}_n = \sum_{j=1}^{p_1}\hat{\boldsymbol{\kappa}}_{jn}\mathcal{W}_{jn}Y_n + \mathcal{X}_n\hat{\boldsymbol{\beta}}_n$, let $Y_n^* = \sum_{j=1}^{p_1}\boldsymbol{\kappa}_{jn}^*\mathcal{W}_{jn}S_n^{-1}X_n\beta_0 + \mathcal{X}_n\boldsymbol{\beta}_n^* + \sum_{j=1}^{p_1}\boldsymbol{\kappa}_{jn}^*\mathcal{W}_{jn}S_n^{-1}R_n^{-1}V_n$. Y_n^* is the leading order term of $\hat{Y}_n$. Then we have the following results: (i) $\frac{1}{n}C_n'(\hat{Y}_n - Y_n^*) = o_p(1)$, (ii) $\frac{1}{n}V_n'A_n(\hat{Y}_n - Y_n^*) = o_p(1)$, and (iii) $\frac{1}{n}(\hat{Y}_n - Y_n^*)'A_n(\hat{Y}_n - Y_n^*) = o_p(1)$, where C_n is an $n\times 1$ vector of uniformly bounded constants and A_n is an $n\times n$ nonstochastic matrix which is bounded in both row and column sum norms. If $\hat{Y}_n = \mathcal{S}_n^{-1}(\hat{\boldsymbol{\kappa}}_n)\mathcal{X}_n\hat{\boldsymbol{\beta}}_n$, by the mean value theorem, $\hat{Y}_n - Y_n^* = \sum_{j=1}^{p_1}\mathcal{S}_n^{-1}(\check{\boldsymbol{\kappa}}_n)\mathcal{W}_{jn}\mathcal{S}_n^{-1}(\check{\boldsymbol{\kappa}}_n)\mathcal{X}_n\check{\boldsymbol{\beta}}_n(\hat{\boldsymbol{\kappa}}_{jn} - \boldsymbol{\kappa}_{jn}^*) + \mathcal{S}_n^{-1}(\check{\boldsymbol{\kappa}}_n)\mathcal{X}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n^*)$, where $\check{\boldsymbol{\theta}}_n$ lies between $\hat{\boldsymbol{\theta}}_n$ and $\boldsymbol{\theta}_n^*$ elementwise. Since $\|\mathcal{S}_n^{-1}(\boldsymbol{\kappa})\mathcal{W}_{jn}\mathcal{S}_n^{-1}(\boldsymbol{\kappa})\mathcal{X}_n\|_\infty$

and $\|\mathcal{S}_n^{-1}(\boldsymbol{\kappa})\mathcal{X}_n\|_\infty$ are $O_p(1)$ uniformly in $\boldsymbol{\kappa}$ of a neighborhood of $\boldsymbol{\kappa}_n^*$ under Assumption 12(ii), $\|\hat{Y}_n - Y_n^*\|_\infty = o_p(1)$. Then (i)–(iii) hold by the submultiplicity of the row sum norm. If $\hat{Y}_n = \sum_{j=1}^{p_1} \hat{\boldsymbol{\kappa}}_{jn}\mathcal{W}_{jn}Y_n + \mathcal{X}_n\hat{\boldsymbol{\beta}}_n$, note that $\hat{Y}_n - Y_n^* = \sum_{j=1}^{p_1}(\hat{\boldsymbol{\kappa}}_{jn} - \boldsymbol{\kappa}_{jn}^*)\mathcal{W}_{jn}S_n^{-1}X_n\beta_0 + \mathcal{X}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n^*) + \sum_{j=1}^{p_1}(\hat{\boldsymbol{\kappa}}_{jn} - \boldsymbol{\kappa}_{jn}^*)\mathcal{W}_{jn}S_n^{-1}R_n^{-1}V_n$. Substituting this expression into the terms in (i)–(iii), we can see that the results hold by Lemma 1(iii). In addition, $\frac{1}{n}(Q_n - Q_n^*)'C_n = o_p(1)$ and $\frac{1}{n}(Q_n - Q_n^*)'A_nV_n = o_p(1)$ by the mean value theorem, where Q_n^* is the matrix obtained by replacing $\mathcal{S}_n^{-1}(\hat{\boldsymbol{\kappa}}_n)\mathcal{X}_n\hat{\boldsymbol{\beta}}_n$ with $\mathcal{S}_n^{-1}(\boldsymbol{\kappa}_n^*)\mathcal{X}_n\boldsymbol{\beta}_n^*$ if Q_n includes $\mathcal{S}_n^{-1}(\hat{\boldsymbol{\kappa}}_n)\mathcal{X}_n\hat{\boldsymbol{\beta}}_n$, and $Q_n^* = Q_n$ otherwise. Thus, $\sup_{\vartheta\in\Theta}\|g_n(\vartheta) - g_n^*(\vartheta)\| = o_p(1)$ for $g_n^*(\vartheta)$ defined in Appendix B. By Lemma 2, $\sup_{\vartheta\in\Theta}\|g_n^*(\vartheta) - \mathrm{E}[g_n^*(\vartheta)]\| = o_p(1)$. It follows that $\sup_{\vartheta\in\Theta}\|g_n(\vartheta) - \mathrm{E}[g_n^*(\vartheta)]\| = o_p(1)$.

Next we can show that Lemmas 4 and 10–12 hold if $g_{ni}(\theta)$ and $g_n(\theta)$ are replaced by, respectively, $g_{ni}(\vartheta)$ and $g_n(\vartheta)$. For Lemma 4, we first show the result when $Q_n = Q_n^*$ so it is nonstochastic. If $\hat{Y}_n = \mathcal{S}_n^{-1}(\hat{\boldsymbol{\kappa}}_n)\mathcal{X}_n\hat{\boldsymbol{\beta}}_n$, $V_n(\vartheta) = V_n(\theta) - \eta R_n(\tau)Y_n^* - \eta R_n(\tau)(\hat{Y}_n - Y_n^*)$, where $\|Y_n^*\|_\infty = O(1)$ and $\|-\eta R_n(\tau)(\hat{Y}_n - Y_n^*)\|_\infty \le |\eta|\cdot\|R_n(\tau)\|_\infty\|\hat{Y}_n - Y_n^*\|_\infty = o_p(1)$. Y_n^* behaves like X_n. Then, if $g_{ni}(\theta)$ in Lemma 4 is replaced by $g_{ni}(\vartheta)$, we may modify its proof by taking out any element of $-\eta R_n(\tau)(\hat{Y}_n - Y_n^*)$ when expanding $g_{ni}(\vartheta)$ and applying Lemma 1(i). Thus, the lemma still holds. If $\hat{Y}_n = \sum_{j=1}^{p_1}\hat{\boldsymbol{\kappa}}_{jn}\mathcal{W}_{jn}Y_n + \mathcal{X}_n\hat{\boldsymbol{\beta}}_n$, then $V_n(\vartheta) = [V_n(\theta) - \eta R_n(\tau)Y_n^*] - \sum_{j=1}^{p_1}(\hat{\boldsymbol{\kappa}}_{jn} - \boldsymbol{\kappa}_{jn}^*)\eta R_n(\tau)\mathcal{W}_{jn}S_n^{-1}R_n^{-1}V_n - \eta R_n(\tau)\{\sum_{j=1}^{p_1}(\hat{\boldsymbol{\kappa}}_{jn} - \boldsymbol{\kappa}_{jn}^*)\mathcal{W}_{jn}S_n^{-1}X_n\beta_0 + \mathcal{X}_n(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n^*)\}$. Thus elements of $g_{ni}(\vartheta)$ can be expanded as polynomials of $(\hat{\boldsymbol{\kappa}}_{jn} - \boldsymbol{\kappa}_{jn}^*)$ and $(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n^*)$. Applying Lemma 1$(i)$ to the coefficients of these polynomials implies that Lemma 4 holds if $g_{ni}(\theta)$ is replaced by $g_{ni}(\vartheta)$. If $Q_n \neq Q_n^*$, then $\|Q_n - Q_n^*\|_\infty = o_p(1)$ by the mean value theorem. Then similar to the above argument for the case with $\hat{Y}_n = \mathcal{S}_n^{-1}(\hat{\boldsymbol{\kappa}}_n)\mathcal{X}_n\hat{\boldsymbol{\beta}}_n$, Lemma 4 still holds if $g_{ni}(\theta)$ is replaced by $g_{ni}(\vartheta)$. With Lemma 4, Lemmas 10–12 also hold. By these lemmas, the consistency of $\hat{\vartheta}_n$ follows as in the proof of Theorem 3.1.

With the consistency that $\hat{\vartheta}_n = \vartheta_0 + o_p(1)$, similar to the proof of Theorem 3.2, $\hat{\vartheta}_n$ and the estimated λ can be shown to be asymptotically normal. Then the asymptotic distribution of the GEL ratio follows as in the proof of Theorem 4.1. □

Appendix B Identification condition for the spatial J test

In this appendix, we provide an identification condition of the augmented model (11) for the spatial J test. The identification condition is in terms of pseudo-true values for an alternative model while the null model is the DGP. In general, we expect that parameter estimates for the alternative model would converge to their pseudo true values. For GS2SLS estimates of the alternative model, relevant studies are in Kelejian (2008) and Kelejian and Piras (2011). We show the convergence result under regularity conditions if the alternative model is estimated by the GEL in the second part of this section.

B.1 Identification condition

For the estimator $\hat{\boldsymbol{\theta}}_n$ of an alternative model while the null model is the DGP, assume that $\{\boldsymbol{\theta}_n^*\}$ is a sequence of nonstochastic pseudo-true values such that $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n^* = o_p(1)$. Since $\hat{Y}_n$ is a generated regressor, we first give its leading order term Y_n^*. If $\hat{Y}_n = \mathcal{S}_n^{-1}(\hat{\boldsymbol{\kappa}}_n)\mathcal{X}_n\hat{\boldsymbol{\beta}}_n$, $Y_n^* = \mathcal{S}_n^{-1}(\boldsymbol{\kappa}_n^*)\mathcal{X}_n\boldsymbol{\beta}_n^*$; if $\hat{Y}_n = \sum_{j=1}^{p_1}\hat{\boldsymbol{\kappa}}_{jn}\mathcal{W}_{jn}Y_n + \mathcal{X}_n\hat{\boldsymbol{\beta}}_n$, $Y_n^* = \sum_{j=1}^{p_1}\boldsymbol{\kappa}_{jn}^*\mathcal{W}_{jn}S_n^{-1}X_n\beta_0 + \mathcal{X}_n\boldsymbol{\beta}_n^* + \sum_{j=1}^{p_1}\boldsymbol{\kappa}_{jn}^*\mathcal{W}_{jn}S_n^{-1}R_n^{-1}V_n$. The first $\hat{Y}_n$ is asymptotically exogenous since Y_n^* is exogenous, but the second one is not due to the presence of $\sum_{j=1}^{p_1}\boldsymbol{\kappa}_{jn}^*\mathcal{W}_{jn}S_n^{-1}R_n^{-1}V_n$ in Y_n^*. Since the first $\hat{Y}_n$ can be directly used as an IV, if Q_n includes the first $\hat{Y}_n$, we define Q_n^* to be the matrix obtained by replacing this $\hat{Y}_n$ by its leading order term $\mathcal{S}_n^{-1}(\boldsymbol{\kappa}_n^*)\mathcal{X}_n\boldsymbol{\beta}_n^*$. If Q_n does not include the first $\hat{Y}_n$, define $Q_n^* = Q_n$. Let $g_n^*(\vartheta)$ be the vector obtained by replacing every $\hat{Y}_n$ in $g_n(\vartheta)$ by Y_n^* and Q_n by Q_n^*. Then we may prove that $g_n(\vartheta)$ converges in probability to $\lim_{n\to\infty}\mathrm{E}[g_n^*(\vartheta)]$ uniformly in ϑ. The identification of ϑ_0 requires ϑ_0 to be the unique solution to $\lim_{n\to\infty}\mathrm{E}[g_n^*(\vartheta)] = 0$. Lower level conditions are provided in the supplementary file.

Assumption 12. *(i) $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n^* = o_p(1)$; (ii) if $\hat{Y}_n = \mathcal{S}_n^{-1}(\hat{\boldsymbol{\kappa}}_n)\mathcal{X}_n\hat{\boldsymbol{\beta}}_n$, $\{\mathcal{S}_n(\boldsymbol{\kappa}_n^*)\}$ are invertible and $\{\mathcal{S}_n^{-1}(\boldsymbol{\kappa}_n^*)\}$ are bounded in both row and column sum norms; (iii) $\vartheta_0 \in \mathrm{int}(\Theta)$ is the unique solution to $\lim_{n\to\infty}\mathrm{E}[g_n^*(\vartheta)] = 0$.*

B.2 GEL estimation of the alternative model

For the alternative model (10), let $\mathcal{V}_n(\boldsymbol{\theta}) = \mathcal{R}_n(\boldsymbol{\tau})[\mathcal{S}_n(\boldsymbol{\kappa})Y_n - \mathcal{X}_n\boldsymbol{\beta}]$, where $\mathcal{R}_n(\boldsymbol{\tau}) = I_n - \sum_{k=1}^{q_1}\boldsymbol{\tau}_k\mathcal{M}_{kn}$. If elements of $\mathcal{V}_n$ were assumed to be i.i.d., the moment vector can be

$$\mathfrak{g}_n(\boldsymbol{\theta}) = \frac{1}{n}[\mathcal{V}_n'(\boldsymbol{\theta})\mathcal{P}_{1n}\mathcal{V}_n(\boldsymbol{\theta}) - \boldsymbol{\sigma}^2\,\mathrm{tr}(\mathcal{P}_{1n}), \ldots, \mathcal{V}_n'(\boldsymbol{\theta})\mathcal{P}_{k_pn}\mathcal{V}_n(\boldsymbol{\theta}) - \boldsymbol{\sigma}^2\,\mathrm{tr}(\mathcal{P}_{k_pn}), \mathcal{Q}_n'\mathcal{V}_n(\boldsymbol{\theta})],$$

where $\mathcal{P}_{1n}$, $\ldots$, $\mathcal{P}_{k_pn}$ are $n\times n$ spatial weights matrices and $\mathcal{Q}_n$ is an $n\times k_q$ IV matrix. On the other hand, if elements of $\mathcal{V}_n$ were independent but heteroskedastic, the moment vector for consistent estimation would be

$$\mathfrak{g}_n(\boldsymbol{\theta}) = \frac{1}{n}[\mathcal{V}_n'(\boldsymbol{\theta})\mathcal{P}_{1n}\mathcal{V}_n(\boldsymbol{\theta}), \ldots, \mathcal{V}_n'(\boldsymbol{\theta})\mathcal{P}_{k_pn}\mathcal{V}_n(\boldsymbol{\theta}), \mathcal{Q}_n'\mathcal{V}_n(\boldsymbol{\theta})],$$

where $\mathcal{P}_{jn}$'s now have zero diagonals. With the moment vector $\mathfrak{g}_n(\boldsymbol{\theta})$, define $\mathfrak{g}_{ni}(\boldsymbol{\theta})$ in a way similar to $g_{ni}(\theta)$ in Section 2 with the intension to capture the martingale difference property. The GEL estimators are

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\max_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}}\sum_{i=1}^n \rho(\boldsymbol{\lambda}'\mathfrak{g}_{ni}(\boldsymbol{\theta})), \text{ and } \hat{\boldsymbol{\lambda}}_n = \arg\max_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}}\sum_{i=1}^n \rho(\boldsymbol{\lambda}'\mathfrak{g}_{ni}(\hat{\boldsymbol{\theta}}_n)),$$

where $\boldsymbol{\Theta}$ and $\boldsymbol{\Lambda}$ are compact.[29] Suppose that there exist pseudo-true values $\boldsymbol{\theta}_n^* \in \boldsymbol{\Theta}$ and $\boldsymbol{\lambda}_n^* \in \boldsymbol{\Lambda}$ such that

$$\mathrm{E}\sum_{i=1}^n \rho(\boldsymbol{\lambda}_n^{*'}\mathfrak{g}_{ni}(\boldsymbol{\theta}_n^*)) = \min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\max_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}}\mathrm{E}\sum_{i=1}^n \rho(\boldsymbol{\lambda}'\mathfrak{g}_{ni}(\boldsymbol{\theta})).$$

Under regularity conditions, the pseudo-true values would satisfy $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n^* = o_p(1)$ and $\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_n^* = o_p(1)$.

Assumption 13. *(i) $\boldsymbol{\Theta}$ and $\boldsymbol{\Lambda}$ are compact, and $\mathcal{V}$ includes all realizations of $\boldsymbol{\lambda}'\mathfrak{g}_{ni}(\boldsymbol{\theta})$ for all $1\le i\le n$, $\boldsymbol{\lambda}\in\boldsymbol{\Lambda}$ and $\boldsymbol{\theta}\in\boldsymbol{\Theta}$; (ii) $\sup_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda},\boldsymbol{\theta}\in\boldsymbol{\Theta}}\frac{1}{n}|\sum_{i=1}^n\rho(\boldsymbol{\lambda}'\mathfrak{g}_{ni}(\boldsymbol{\theta})) - \mathrm{E}\sum_{i=1}^n\rho(\boldsymbol{\lambda}'\mathfrak{g}_{ni}(\boldsymbol{\theta}))| = o_p(1)$; (iii) $\frac{1}{n}\mathrm{E}\sum_{i=1}^n\rho(\boldsymbol{\lambda}'\mathfrak{g}_{ni}(\boldsymbol{\theta}))$ is uniformly equicontinuous on $(\boldsymbol{\Theta},\boldsymbol{\Lambda})$; (iv) for each $\boldsymbol{\theta}\in\boldsymbol{\Theta}$, the identifiably unique maximizer $\boldsymbol{\lambda}_n^*(\boldsymbol{\theta})\in\boldsymbol{\Lambda}$ of $\max_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}}\mathrm{E}\sum_{i=1}^n\rho(\boldsymbol{\lambda}'\mathfrak{g}_{ni}(\boldsymbol{\theta}))$ is equicontinuous in $\boldsymbol{\theta}$;[30] (v) $\mathrm{E}\sum_{i=1}^n\rho(\boldsymbol{\lambda}_n^{*'}(\boldsymbol{\theta})\mathfrak{g}_{ni}(\boldsymbol{\theta}))$ has identifiably unique minimizer $\boldsymbol{\theta}_n^*\in\boldsymbol{\Theta}$.*

[29] For analytical convenience, the parameter space of $\boldsymbol{\lambda}$ for the alternative model is assumed to be compact, unlike the case of the null model where the compactness assumption can be avoided by the concavity of $\rho(\cdot)$.

[30] $\boldsymbol{\lambda}_n^*(\boldsymbol{\theta})$ is identifiably unique if for all $\epsilon>0$, $\limsup_{n\to\infty}[\max_{\boldsymbol{\lambda}\in B_n^c(\epsilon)}\frac{1}{n}\mathrm{E}\sum_{i=1}^n\rho(\boldsymbol{\lambda}'\mathfrak{g}_{ni}(\boldsymbol{\theta})) - \frac{1}{n}\mathrm{E}\sum_{i=1}^n\rho(\boldsymbol{\lambda}_n^{*'}(\boldsymbol{\theta})\mathfrak{g}_{ni}(\boldsymbol{\theta}))] < 0$, where $B_n^c(\epsilon)$ is the complement in $\boldsymbol{\Lambda}$ of an open ball $B_n(\epsilon)$ centered at $\boldsymbol{\lambda}_n^*(\boldsymbol{\theta})$ with radius ϵ (White, 1994).

The above assumption gives high level conditions similar to those in Hong et al. (2003).[31] Some conditions might be relaxed, e.g., the uniform convergence condition in Assumption 13(ii) follows by pointwise convergence and stochastic equicontinuity, while the latter holds if the first order derivative of $\rho(\cdot)$ is bounded on its domain. If the null model is the DGP, Assumption 13 provides sufficient conditions for the convergence of the GEL estimates for the alternative model to their pseudo-true values.

Theorem B.1. *Under Assumption 13,* $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n^* = o_p(1)$ *and* $\hat{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_n^* = o_p(1)$.

Proof. Let $\hat{\boldsymbol{\lambda}}_n(\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\lambda}\in\boldsymbol{\Lambda}} \sum_{i=1}^n \rho(\boldsymbol{\lambda}'\mathfrak{g}_{ni}(\boldsymbol{\theta}))$. Under Assumption 13, as in the proof of Lemma 1 in Hong et al. (2003), $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \|\hat{\boldsymbol{\lambda}}_n(\boldsymbol{\theta}) - \boldsymbol{\lambda}_n^*(\boldsymbol{\theta})\| = o_p(1)$. Then $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{n}|\operatorname{E}\sum_{i=1}^n \rho(\hat{\boldsymbol{\lambda}}_n'(\boldsymbol{\theta})\mathfrak{g}_{ni}(\boldsymbol{\theta})) - \operatorname{E}\sum_{i=1}^n \rho(\boldsymbol{\lambda}_n^{*'}(\boldsymbol{\theta})\mathfrak{g}_{ni}(\boldsymbol{\theta}))| = o_p(1)$ under Assumption 13(iii). By Assumption 13(ii), $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{n}\Big|\sum_{i=1}^n \rho\big(\hat{\boldsymbol{\lambda}}_n'(\boldsymbol{\theta})\mathfrak{g}_{ni}(\boldsymbol{\theta})\big) - \operatorname{E}\sum_{i=1}^n \rho\big(\hat{\boldsymbol{\lambda}}_n'(\boldsymbol{\theta})\mathfrak{g}_{ni}(\boldsymbol{\theta})\big)\Big| = o_p(1)$. Thus, $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \frac{1}{n}|\sum_{i=1}^n \rho(\hat{\boldsymbol{\lambda}}_n'(\boldsymbol{\theta})\mathfrak{g}_{ni}(\boldsymbol{\theta})) - \operatorname{E}\sum_{i=1}^n \rho((\boldsymbol{\lambda}_n^{*'}(\boldsymbol{\theta})\mathfrak{g}_{ni}(\boldsymbol{\theta}))| = o_p(1)$. Hence, $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n^* = o_p(1)$ (White, 1994, Theorem 3.4 on p. 28). If follows that $\hat{\boldsymbol{\lambda}}_n = \hat{\boldsymbol{\lambda}}_n(\hat{\boldsymbol{\theta}}_n) = \boldsymbol{\lambda}_n^*(\hat{\boldsymbol{\theta}}_n) + o_p(1) = \boldsymbol{\lambda}_n^*(\boldsymbol{\theta}_n^*) + o_p(1) = \boldsymbol{\lambda}_n^* + o_p(1)$. □

Appendix C Lemmas

C.1 General lemmas

The martingale differences generated by linear-quadratic forms can be seen as linear forms. The following lemma on products of linear forms is useful in deriving relevant orders of products of martingale differences.

Lemma 1. *For* $l = 1, \dots, s$, *let* $D_{ln}(\theta) = [d_{ln,ij}(\theta)]$ *be* $n\times n$ *matrices which are bounded in row sum norm uniformly in* $\theta \in \Theta$, *and* $u_{ln} = [u_{ln,i}]$ *be* $n \times 1$ *vectors such that* $\sup_{1\le l\le s}\sup_{1\le j\le n} \operatorname{E}|u_{ln,j}|^{a_l} = O(1)$ *for* $a_l > 1$. *Then,*

(i) $\sup_{\theta\in\Theta}\sup_{1\le i\le n} |\prod_{l=1}^s \sum_{j=1}^n d_{ln,ij}(\theta)u_{ln,j}| = O_p(n^{\sum_{l=1}^s \frac{1}{a_l}})$;

(ii) $\sup_{\theta\in\Theta}\sup_{1\le i\le n} \operatorname{E}(|\prod_{l=1}^s \sum_{j=1}^n d_{ln,ij}(\theta)u_{ln,j}|^{1/\sum_{l=1}^s \frac{1}{a_l}}) = O(1)$ *if* $D_{ln}(\theta)$*'s are nonstochastic;*

(iii) $\frac{1}{n}\sum_{i=1}^n \prod_{l=1}^s \sum_{j=1}^n d_{ln,ij}(\theta)u_{ln,j} = O_p(1)$ *if* $\sum_{l=1}^s \frac{1}{a_l} \le 1$, *and* $\sup_{\theta\in\Theta} \frac{1}{n}|\sum_{i=1}^n \prod_{l=1}^s \sum_{j=1}^n d_{ln,ij}(\theta)u_{ln,j}| = O_p(1)$ *if* $\sum_{l=1}^s \frac{1}{a_l} \le 1$ *and* $D_{ln}(\theta)$*'s are also bounded in column sum norm uniformly in* $\theta \in \Theta$.

The following lemma involves the specific forms related to the martingale differences of linear-quadratic forms and is useful to show orders of terms in Nagar-type expansions of GMM and GEL estimators. In particular, it is used to prove Lemma 7.

Lemma 2. *Suppose that* v_{ni}*'s are independent with zero mean and* $\operatorname{E}(v_{ni}^2) = \sigma_{ni}^2$ *for* $i = 1, \dots, n$, *and* $[a_{ln,ij}]$, $[b_{ln,ij}]$, $[c_{ln,ij}]$, $[d_{ln,ij}]$, $[e_{ln,ij}]$, $[f_{ln,ij}]$, $[g_{ln,ij}]$ *and* $[h_{ln,ij}]$ *for* $l = 1, 2$ *are* $n \times n$ *nonstochastic matrices with bounded row sum norms. Then,*

[31] Among the regularity conditions, uniform convergence of the GEL objective function is assumed. With a misspecified model, the proof strategy of Theorem 3.1 for a correctly specified model might not be applicable and also the GEL objective function is not a sum of martingale differences. Thus, other low level conditions for uniform convergence might be needed. We assume uniform convergence for simplicity.

(i) *for* $r_{ni}^{(l)} = a_{ln,ii}(v_{ni}^2 - \sigma_{ni}^2) + b_{ln,ii}v_{ni} + (c_{ln,ii} + d_{ln,ii}v_{ni})\sum_{j=1}^{i-1} e_{ln,ij}v_{nj} + \sum_{j=1}^{i-1}(g_{ln,ij}v_{nj}\sum_{k=1}^{j-1} h_{ln,ik}v_{nk})$ *with* $l = 1$ *and 2, if* $\sup_n \sup_{1\le i\le n} \mathrm{E}(v_{ni}^4) < \infty$*, then* $\frac{1}{n}\sum_{i=1}^n \mathrm{E}(r_{ni}^{(1)}r_{ni}^{(2)}) = O(1)$ *and* $\frac{1}{n}\sum_{i=1}^n [r_{ni}^{(1)}r_{ni}^{(2)} - \mathrm{E}(r_{ni}^{(1)}r_{ni}^{(2)})] = o_p(1)$*;*

(ii) *for*

$$r_{ni}^{(l)} = a_{ln,ii}(v_{ni}^2 - \sigma_{ni}^2) + b_{ln,ii}v_{ni} + (c_{ln,ii} + d_{ln,ii}v_{ni})\sum_{j=1}^{i-1} e_{ln,ij}v_{nj} + \sum_{j=1}^{i-1} f_{ln,ij}(v_{nj}^2 - \sigma_{nj}^2) + \sum_{j=1}^{i-1}\Big(g_{ln,ij}v_{nj}\sum_{k=1}^{j-1} h_{ln,ik}v_{nk}\Big)$$

with $l = 1$ *and* 2*, if* $\sup_n \sup_{1\le i\le n} \mathrm{E}(v_{ni}^8) < \infty$*, then* $\frac{1}{n}\sum_{i=1}^n [r_{ni}^{(1)}r_{ni}^{(2)} - \mathrm{E}(r_{ni}^{(1)}r_{ni}^{(2)})] = O_p(n^{-1/2})$.

C.2 Lemmas for the SARAR(p,q) model

All lemmas below accommodate both the homoskedastic and heteroskedastic cases for the SARAR(p,q) model (1), where $\theta = (\tau, \kappa, \beta', \sigma^2)'$ for the homoskedastic case, and $\theta = (\tau, \kappa, \beta')'$ for the heteroskedastic case. Let k_θ be the dimension of θ. The next lemma shows the consistency of an estimator of the covariance of two linear-quadratic forms, where the estimator is formed with estimated martingale differences.

Lemma 3. *Suppose that* $A_{rn}(\theta) = [a_{rn,ij}(\theta)]$ *for* $r = 1, 2$ *are square matrices of dimension* n*,* $b_{rn}(\theta) = [b_{rn,i}(\theta)]$ *for* $r = 1, 2$ *are column vectors of dimension* n*, and their elements are nonstochastic functions of* $\theta \in \Theta$*. Assume that elements of* $A_{rn}(\theta)$ *and* $b_{rn}(\theta)$ *are differentiable with respect to* θ*, the sequences* $\{A_{rn}(\theta)\}$ *and* $\{\frac{\partial A_{rn}(\theta)}{\partial\theta_j}\}$ *for* $r = 1, 2$ *and* $j = 1, \ldots, k_\theta$ *are bounded in both row and column sum norms, and* $\{b_{rn}(\theta)\}$ *and* $\{\frac{\partial b_{rn}(\theta)}{\partial\theta_j}\}$ *for* $r = 1, 2$ *and* $j = 1, \ldots, k_\theta$ *are bounded in row sum norm, uniformly in a neighborhood of* θ_0*.*

Let $\xi_{rn,i}(\theta) = a_{rn,ii}(\theta)[v_{ni}^2(\theta) - \sigma^2] + 2v_{ni}(\theta)\sum_{j=1}^{i-1} a_{rn,ij}(\theta)v_{nj}(\theta) + b_{rn,i}(\theta)v_{ni}(\theta)$ *for* $r = 1, 2$ *if the disturbances* v_{ni}*'s are homoskedastic, and* $\xi_{rn,i}(\theta) = 2v_{ni}(\theta)\sum_{j=1}^{i-1} a_{rn,ij}(\theta)v_{nj}(\theta) + b_{rn,i}(\theta)v_{ni}(\theta)$ *for* $r = 1, 2$ *if* v_{ni}*'s are heteroskedastic. Assume that* $\hat\theta_n = \theta_0 + o_p(1)$*. Then, under Assumptions 1–4,* $\frac{1}{n}\sum_{i=1}^n \xi_{1n,i}(\hat\theta_n)\xi_{2n,i}(\hat\theta_n) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}[\xi_{1n,i}(\theta_0)\xi_{2n,i}(\theta_0)] + o_p(1)$*.*

Lemma 4. *Under Assumptions 1–4,* $\sup_{\theta\in\Theta}\sup_{1\le i\le n}\|g_{ni}(\theta)\| = O_p(n^{2/(4+\iota)})$*, where* $\|\cdot\|$ *denotes the Euclidean norm.*

Let $G_n^{(j)}(\theta) = \frac{\partial G_n(\theta)}{\partial\theta_j}$, $G_n^{(jk)}(\theta) = \frac{\partial^2 G_n(\theta)}{\partial\theta_j\partial\theta_k}$, $G_n^{(jkl)}(\theta) = \frac{\partial^3 G_n(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l}$ and $G_{ni}^{(j)}(\theta) = \frac{\partial G_{ni}(\theta)}{\partial\theta_j}$, where $G_{ni}(\theta) = \frac{\partial g_{ni}(\theta)}{\partial\theta'}$.

Lemma 5. *Under Assumptions 1–4,* $\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^n\|g_{ni}(\theta)\|^2$, $\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^n\|G_{ni}(\theta)\|$, $\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^n\|G_{ni}^{(j)}(\theta)\|$, $\sup_{\theta\in\Theta}\|g_n(\theta)\|$, $\sup_{\theta\in\Theta}\|G_n(\theta)\|$, $\sup_{\theta\in\Theta}\|G_n^{(j)}(\theta)\|$, $\sup_{\theta\in\Theta}\|G_n^{(jk)}(\theta)\|$, *and* $\sup_{\theta\in\Theta}\|G_n^{(jkl)}(\theta)\|$ *are all of order* $O_p(1)$*.*

Let $g_{ni}^{(j)}(\theta) = \frac{\partial g_{ni}(\theta)}{\partial\theta_j}$, $g_{ni}^{(jk)}(\theta) = \frac{\partial^2 g_{ni}(\theta)}{\partial\theta_j\partial\theta_k}$, $g_{ni}^{(jkl)}(\theta) = \frac{\partial^3 g_{ni}(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l}$, and $g_{ni}^{(jklr)}(\theta) = \frac{\partial^4 g_{ni}(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l\partial\theta_r}$.

Lemma 6. *Under Assumptions 1–4,* $\sup_{\theta\in\Theta}\|\frac{1}{n}\sum_{i=1}^n g_{ni}(\theta)g_{ni}'(\theta)\|$, $\sup_{\theta\in\Theta}\|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(j)}(\theta)g_{ni}'(\theta)\|$, $\sup_{\theta\in\Theta}\|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(jk)}(\theta)g_{ni}'(\theta)\|$, $\sup_{\theta\in\Theta}\|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(j)}(\theta)g_{ni}^{(k)'}(\theta)\|$, $\sup_{\theta\in\Theta}\|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(jkl)}(\theta)g_{ni}'(\theta)\|$, $\sup_{\theta\in\Theta}\|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(jk)}(\theta)g_{ni}^{(l)'}(\theta)\|$, $\sup_{\theta\in\Theta}\|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(jklr)}(\theta)g_{ni}'(\theta)\|$, $\sup_{\theta\in\Theta}\|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(jkl)}(\theta)g_{ni}^{(r)'}(\theta)\|$ *and* $\sup_{\theta\in\Theta}\|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(jk)}(\theta)g_{ni}^{(lr)'}(\theta)\|$ *have order* $O_p(1)$.

Lemma 7. *Under Assumptions 1–4, (i)* $\frac{1}{n}\sum_{i=1}^n g_{ni}(\theta_0)g'_{ni}(\theta_0) = \bar{\Omega}_n + o_p(1)$, *(ii)* $\frac{1}{n}\sum_{i=1}^n [\mathrm{E}\, g_{ni}^{(j)}(\theta_0)]g'_{ni}(\theta_0) = O_p(n^{-1/2})$, *and (iii)* $\frac{1}{n}\sum_{i=1}^n \mathrm{E}[g_{ni}^{(j)}(\theta_0)g'_{ni}(\theta_0)] = O(1)$; *under Assumptions 1–4 and 10, (iv)* $\frac{1}{n}\sum_{i=1}^n g_{ni}(\theta_0)g'_{ni}(\theta_0) = \bar{\Omega}_n + O_p(n^{-1/2})$, *and (v)* $\frac{1}{n}\sum_{i=1}^n \{g_{ni}^{(j)}(\theta_0)g'_{ni}(\theta_0) - \mathrm{E}[g_{ni}^{(j)}(\theta_0)g'_{ni}(\theta_0)]\} = O_p(n^{-1/2})$, *for* $j = 1, \ldots, k_\theta$.

The first order condition for the initial GMM can be written as

$$0 = -\begin{pmatrix} G'_n(\tilde{\theta}_n)\tilde{\lambda}_n \\ g_n(\tilde{\theta}_n) + \hat{J}_n\tilde{\lambda}_n \end{pmatrix}, \tag{28}$$

where $\tilde{\lambda}_n = -\hat{J}_n^{-1}g_n(\tilde{\theta}_n)$. Let $\tilde{\gamma}_n = (\tilde{\theta}'_n, \tilde{\lambda}'_n)'$ and $\gamma_0 = (\theta'_0, 0_{1\times k_g})'$. Recall that in the following Lemma, $\bar{G}_n$ is the expected value of G_n, $\bar{J}_n$ is in Assumption 8, and $e_{k_g,j}$ is the jth column of the $k_g \times k_g$ identity matrix.

Lemma 8. *Under Assumptions 1–5 and 8–9,* $\sqrt{n}(\tilde{\gamma}_n - \gamma_0) = \tilde{\xi}_n + n^{-1/2}\tilde{\psi}_n + O_p(n^{-1})$, *where* $\tilde{\xi}_n = -(\bar{K}_n^J)^{-1}\binom{0}{\sqrt{n}g_n(\theta_0)} = O_p(1)$ *and* $\tilde{\psi}_n = -(\bar{K}_n^J)^{-1}\begin{pmatrix} 0 & \sqrt{n}(G'_n - \bar{G}'_n) \\ \sqrt{n}(G_n - \bar{G}_n) & \xi_n^J \end{pmatrix}\tilde{\xi}_n - \frac{1}{2}(\bar{K}_n^J)^{-1}\sum_{j=1}^{k_\theta+k_g}\tilde{\xi}_{nj}\bar{K}_{nj}\tilde{\xi}_n = O_p(1)$, *where* $\bar{K}_n^J = \begin{pmatrix} 0 & \bar{G}'_n \\ \bar{G}_n & \bar{J}_n \end{pmatrix}$, $\bar{K}_{nj} = \begin{pmatrix} 0 & \bar{G}_n^{(j)'} \\ \bar{G}_n^{(j)} & 0 \end{pmatrix}$ *for* $1 \le j \le k_\theta$, *and* $\bar{K}_{nj} = \begin{pmatrix} [\bar{G}_n^{(1)'}e_{k_g,j-k_\theta}, \ldots, \bar{G}_n^{(k_\theta)'}e_{k_g,j-k_\theta}] & 0 \\ 0 & 0 \end{pmatrix}$ *for* $k_\theta + 1 \le j \le k_\theta + k_g$.

Lemma 9. *Under Assumptions 1–5 and 8–9,* $\Omega_n(\tilde{\theta}_n) = \bar{\Omega}_n + o_p(1)$; *under the additional Assumption 10,*

$$\sqrt{n}[\Omega_n(\tilde{\theta}_n) - \bar{\Omega}_n] = \xi_n^{\Omega} + O_p(n^{-1/2}),$$

where $\xi_n^{\Omega} = \sqrt{n}[\frac{1}{n}\sum_{i=1}^n g_{ni}(\theta_0)g'_{ni}(\theta_0) - \bar{\Omega}_n] + \sum_{j=1}^{k_\theta}\{\frac{1}{n}\sum_{i=1}^n \mathrm{E}[g_{ni}(\theta_0)g_{ni}^{(j)'}(\theta_0) + g_{ni}^{(j)}(\theta_0)g'_{ni}(\theta_0)]\}\tilde{\xi}_{nj} = O_p(1)$.

Lemma 10. *Under Assumptions 1–4, for any* ζ *with* $\zeta > \frac{2}{4+\iota}$ *and* $\Lambda_n = \{\lambda : \|\lambda\| \le n^{-\zeta}\}$, $\sup_{\theta\in\Theta,\lambda\in\Lambda_n,1\le i\le n}|\lambda' g_{ni}(\theta)| \xrightarrow{p} 0$, *and* $\Lambda_n \subset \Lambda_n(\theta)$ *for all* $\theta \in \Theta$ *w.p.a.1.*

Denote $\varrho_n(\theta, \lambda) = \frac{1}{n}\sum_{i=1}^n \rho(\lambda' g_{ni}(\theta))$ for the next lemmas for simplicity.

Lemma 11. *Under Assumptions 1–4, 6 and 7, if* $\bar{\theta}_n \xrightarrow{p} \theta_0, \bar{\theta}_n \in \Theta$, *and* $g_n(\bar{\theta}_n) = O_p(n^{-1/2})$, *then* $\bar{\lambda}_n = \arg\max_{\lambda\in\Lambda_n(\bar{\theta}_n)}\varrho_n(\bar{\theta}_n, \lambda)$ *exists w.p.a.1,* $\bar{\lambda}_n = O_p(n^{-1/2})$, *and* $\sup_{\lambda\in\Lambda_n(\bar{\theta}_n)}\varrho_n(\bar{\theta}_n, \lambda) \le \rho(0) + O_p(n^{-1})$.

Lemma 12. *Under Assumptions 1–4, 6 and 7,* $\|g_n(\hat{\theta}_{n,\mathrm{GEL}})\| = O_p(n^{-1/2})$, *where* $\hat{\theta}_{n,\mathrm{GEL}}$ *is the GEL estimator.*

References

Altonji, J.G., Segal, L.M., 1996. Small-sample bias in GMM estimation of covariance structures. Journal of Business & Economic Statistics 14, 353–366.

Anatolyev, S., 2005. GMM, GEL, serial correlation, and asymptotic bias. Econometrica 73, 983–1002.

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

Bandyopadhyay, S., Lahiri, S.N., Nordman, D.J., 2015. A frequency domain empirical likelihood method for irregularly spaced spatial data. Annals of Statistics 43, 519–545.

Chen, S.X., Keilegom, I.V., 2009. A review on empirical likelihood methods for regression. Test 18, 415–447.

Chuang, C.S., Chan, N.H., 2002. Empirical likelihood for autoregressive models, with applications to unstable time series. Statistical Sinica 12, 387–407.

Corcoran, S.A., 1998. Bartlett adjustment of empirical discrepancy statistics. Biometrika 85, 967–972.

Davidson, R., MacKinnon, J., 1981. Several tests for model specification in the presence of alternative hypotheses. Econometrica 47, 781–793.

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

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

DiCiccio, T., Hall, P., Romano, J., 1991. Empirical likelihood is Bartlett-correctable. Annals of Statistics 19, 1053–1061.

Donald, S.G., Imbens, G.W., Newey, W.K., 2003. Empirical likelihood estimation and consistent tests with conditional moment restrictions. Journal of Econometrics 117, 55–93.

Fernández-Vázquez, E., Mayor-Fernández, M., Rodríguez-Vález, J., 2009. Estimating spatial autoregressive models by GME-GCE techniques. International Regional Science Review 32, 148–172.

Hall, P., Heyde, C., 1980. Martingale limit theory and its applications. Academic Press, New York.

Hall, P., La Scala, B., 1990. Methodology and algorithms of empirical likelihood. International Statistical Review 58, 109–127.

Hansen, L.P., Heaton, J., Yaron, A., 1996. Finite-sample properties of some alternative GMM estimators. Journal of Business & Economic Statistics 14, 262–280.

Hong, H., Preston, B., Shum, M., 2003. Generalized empirical likelihood–based model selection criteria for moment condition models. Econometric Theory 19, 923–943.

Imbens, G.W., 1997. One-step estimators for over-identified generalized method of moments models. Review of Economic Studies 64, 359–383.

Imbens, G.W., Spady, R.H., Johnson, P., 1998. Information theoretic approaches to inference in moment condition models. Econometrica 66, 333–357.

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

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

Jin, F., Lee, L.F., 2015. On the bootstrap for Moran's I test for spatial dependence. Journal of Econometrics 184, 295–314.

Jin, F., Lee, L.F., 2017. Outer-product-of-gradients tests for spatial autoregressive models. Forthcoming in the *Regional Science and Urban Economics*.

Kelejian, H.H., 2008. A spatial J-test for model specification against a single or a set of non-nested alternatives. Letters in Spatial and Resource Sciences 1, 3–11.

Kelejian, H.H., Piras, G., 2011. An extension of Kelejian's J-test for non-nested spatial models. Regional Science and Urban Economics 41, 281–292.

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

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

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

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

Kitamura, Y., 1997. Empirical likelihood methods with weakly dependent processes. The Annals of Statistics 25, 2084–2102.

Kitamura, Y., 2001. Asymptotic optimality of empirical likelihood for testing moment restrictions. Econometrica 69, 1661–1672.

Kitamura, Y., 2007. Empirical likelihood methods in econometrics: Theory and practice, in: Blundell, R., Newey, W.K., Personn, T. (Eds.), Advances in Economics and Econometrics: Ninth World Congress of the Econometric Society.

Kitamura, Y., Stutzer, M., 1997. An information-theoretical alternative to generalized method of moments estimation. Econometrica 65, 861–874.

Kostov, P., 2013. Empirical likelihood estimation of the spatial quantile regression. Journal of Geographical Systems 15, 51–69.

Lafontaine, F., White, K.J., 1986. Obtaining any Wald statistic you want. Economics Letters 21, 35–40.

Lazar, N.A., Mykland, P.A., 1999. Empirical likelihood in the presence of nuisance parameters. Biometrika 86, 203–211.

Lee, L.F., 2001. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Manuscript. Department of Economics, Ohio State University.

Lee, L.F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometric theory 18, 252–277.

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

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

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

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

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

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

Marsh, T.L., Mittelhammer, R.C., 2004. Generalized maximum entropy estimation of a first order spatial autoregressive model, in: Lesage, J.P., Pace, R.K. (Eds.), Spatial and Spatiotemporal Econometrics, Emerald Group Publishing Limited. pp. 199–234.

Mittelhammer, R.C., Judge, G.G., Schoenberg, R., 2005. Empirical evidence concerning the finite sample performance of EL-type structural equation estimation and inference methods, in: Andrews, D.W.K., Stock, J.H. (Eds.), Identification and inference for econometric models: Essays in honor of Thomas Rothenberg, Cambridge University Press. pp. 282–305.

Mykland, P.A., 1993. Asymptotic expansions for martingales. Annals of Probability 21, 800–818.

Mykland, P.A., 1995. Dual likelihood. Annals of Statistics 23, 396–421.

Nagar, A.L., 1959. The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. Econometrica 27, 575–595.

Newey, W.K., Ramalho, J.J., Smith, R.J., 2005. Asymptotic bias for GMM and GEL estimators with estimated nuisance parameters, in: Andrews, D.W.K., Stock, J.H. (Eds.), Identification and inference for econometric models: Essays in honor of Thomas Rothenberg, Cambridge University Press. pp. 245–281.

Newey, W.K., Smith, R.J., 2004. Higher order properties of GMM and generalized empirical likelihood estimators. Econometrica 72, 219–255.

Newey, W.K., West, K., 1987. Hypothesis testing with efficient method of moments estimation. International Economic Review 28, 777–787.

Nordman, D.J., 2008a. A blockwise empirical likelihood for spatial lattice data. Statistica Sinica 18, 1111–1129.

Nordman, D.J., 2008b. An empirical likelihood method for spatial regression. Metrika 68, 351—-363.

Otsu, T., 2010. On Bahadur efficiency of empirical likelihood. Journal of Econometrics 157, 248–256.

Owen, A., 1991. Empirical likelihood for linear models. Annals of Statistics 19, 1725–1747.

Owen, A., 2001. Empirical likelihood. Chapman & Hall, New York.

Perevodchikov, E.V., Marsh, T.L., Mittelhammer, R.C., 2012. Information theory estimators for the first-order spatial autoregressive model. Entropy 14, 1165–1185.

Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating equations. Annals of Statistics 22, 300–325.

Qu, X., Lee, L.F., 2012. LM tests for spatial correlation in spatial models with limited dependent variables. Regional Science and Urban Economics 42, 430–445.

Qu, X., Lee, L.F., 2013. Locally most powerful tests for spatial interactions in the simultaneous SAR Tobit model. Regional Science and Urban Economics 43, 307– 321.

Ramalho, J.J., 2002. Alternative estimation methods and specification tests for moment condition models. Ph.D. thesis. University of Bristol.

Rilstone, P., Srivastava, V.K., Ullah, A., 1996. The second-order bias and mean squared error of nonlinear estimators. Journal of Econometrics 75, 369–395.

Ruud, P.A., 2000. An Introduction to Classical Econometric Theory. Oxford University Press.

Smith, R.J., 1997. Alternative semi-parametric likelihood approaches to generalized method of moments estimation. Economic Journal 107, 503–519.

Van Hala, M., Nordman, D.J., Zhu, Z., 2015. Empirical likelihood for irregularly located spatial data. Statistica Sinica 25, 1399–1420.

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

Table 1: Biases, standard errors and RMSEs of estimators for the SARAR(2,0) model (12) in the homoskedastic case

R^2, κ_{10}, κ_{20}		κ_1	κ_2	β_1	β_2	β_3	σ^2
	$n = 144$						
0.8, 0.2, 0.2	GMM	-0.001[0.067]0.067	-0.007[0.095]0.096	0.024[0.209]0.211	-0.003[0.033]0.033	-0.002[0.032]0.032	-0.019[0.015]0.024
	EL	-0.005[0.069]0.069	-0.013[0.096]0.097	0.043[0.208]0.213	-0.001[0.033]0.033	-0.001[0.033]0.033	-0.011[0.015]0.019
	ET	-0.005[0.069]0.069	-0.013[0.097]0.098	0.043[0.210]0.215	-0.001[0.034]0.034	-0.001[0.033]0.033	-0.017[0.015]0.022
0.8, 0.2, 0.4	GMM	-0.002[0.068]0.068	-0.003[0.086]0.086	0.017[0.250]0.251	-0.002[0.033]0.033	-0.001[0.030]0.030	-0.018[0.015]0.023
	EL	-0.005[0.070]0.070	-0.010[0.086]0.087	0.051[0.250]0.255	-0.001[0.034]0.034	0.001[0.032]0.032	-0.010[0.015]0.018
	ET	-0.005[0.070]0.070	-0.009[0.087]0.087	0.049[0.251]0.256	-0.001[0.034]0.034	0.001[0.032]0.032	-0.016[0.015]0.022
0.8, 0.4, 0.2	GMM	-0.002[0.060]0.060	-0.001[0.084]0.084	0.015[0.253]0.253	-0.003[0.032]0.032	-0.002[0.032]0.032	-0.017[0.015]0.023
	EL	-0.007[0.061]0.062	-0.007[0.087]0.087	0.049[0.261]0.266	-0.001[0.032]0.032	-0.000[0.033]0.033	-0.009[0.015]0.018
	ET	-0.007[0.062]0.062	-0.007[0.087]0.087	0.047[0.259]0.263	-0.001[0.033]0.033	-0.000[0.033]0.033	-0.015[0.015]0.021
0.8, 0.4, 0.4	GMM	-0.001[0.058]0.058	-0.004[0.070]0.071	0.035[0.364]0.366	-0.002[0.032]0.032	-0.002[0.033]0.033	-0.017[0.014]0.023
	EL	-0.004[0.060]0.060	-0.008[0.072]0.073	0.080[0.368]0.377	-0.000[0.033]0.033	-0.001[0.033]0.033	-0.010[0.015]0.018
	ET	-0.004[0.059]0.060	-0.008[0.072]0.073	0.082[0.367]0.376	-0.000[0.033]0.033	-0.001[0.033]0.033	-0.015[0.015]0.021
0.4, 0.2, 0.2	GMM	0.000[0.104]0.104	-0.005[0.150]0.150	0.020[0.343]0.344	-0.004[0.079]0.079	-0.003[0.075]0.075	-0.113[0.089]0.144
	EL	-0.006[0.100]0.101	-0.020[0.148]0.150	0.067[0.330]0.337	-0.002[0.080]0.080	-0.001[0.077]0.077	-0.065[0.093]0.114
	ET	-0.006[0.101]0.101	-0.020[0.148]0.149	0.065[0.331]0.338	-0.002[0.080]0.080	-0.001[0.077]0.077	-0.097[0.092]0.134
0.4, 0.2, 0.4	GMM	-0.004[0.105]0.105	0.002[0.136]0.136	0.010[0.408]0.408	-0.002[0.076]0.076	-0.002[0.076]0.076	-0.111[0.090]0.143
	EL	-0.010[0.105]0.106	-0.016[0.132]0.133	0.086[0.387]0.397	0.001[0.079]0.079	0.001[0.077]0.077	-0.062[0.093]0.112
	ET	-0.010[0.105]0.106	-0.016[0.132]0.133	0.088[0.387]0.397	0.001[0.079]0.079	0.001[0.077]0.077	-0.094[0.092]0.131
0.4, 0.4, 0.2	GMM	-0.002[0.087]0.087	-0.009[0.128]0.128	0.049[0.413]0.416	-0.002[0.079]0.079	-0.004[0.080]0.081	-0.109[0.088]0.140
	EL	-0.013[0.087]0.087	-0.021[0.124]0.125	0.122[0.398]0.416	0.002[0.081]0.081	-0.002[0.083]0.083	-0.060[0.092]0.110
	ET	-0.013[0.087]0.088	-0.022[0.124]0.126	0.123[0.398]0.417	0.001[0.082]0.082	-0.002[0.083]0.083	-0.091[0.091]0.128
0.4, 0.4, 0.4	GMM	-0.001[0.086]0.086	0.001[0.115]0.115	-0.001[0.677]0.677	-0.006[0.081]0.081	0.001[0.081]0.081	-0.107[0.091]0.140
	EL	-0.011[0.088]0.089	-0.008[0.120]0.120	0.123[0.717]0.728	-0.002[0.084]0.084	0.003[0.083]0.083	-0.056[0.094]0.110
	ET	-0.011[0.088]0.089	-0.007[0.119]0.120	0.121[0.718]0.729	-0.002[0.084]0.084	0.003[0.084]0.084	-0.088[0.093]0.128
	$n = 400$						
0.8, 0.2, 0.2	GMM	-0.002[0.036]0.037	-0.006[0.055]0.055	0.020[0.115]0.117	-0.001[0.019]0.019	-0.000[0.018]0.018	-0.008[0.009]0.012
	EL	-0.003[0.036]0.036	-0.007[0.055]0.055	0.022[0.114]0.116	-0.000[0.019]0.019	0.000[0.018]0.018	-0.003[0.009]0.009
	ET	-0.003[0.036]0.036	-0.007[0.055]0.055	0.022[0.114]0.116	-0.000[0.019]0.019	0.000[0.018]0.018	-0.006[0.009]0.010
0.8, 0.2, 0.4	GMM	0.000[0.037]0.037	-0.003[0.048]0.048	0.010[0.138]0.138	-0.001[0.019]0.019	-0.001[0.018]0.018	-0.008[0.009]0.012
	EL	-0.000[0.037]0.037	-0.004[0.048]0.048	0.016[0.139]0.140	-0.000[0.019]0.019	0.000[0.018]0.018	-0.003[0.009]0.010
	ET	-0.000[0.037]0.037	-0.004[0.048]0.048	0.016[0.139]0.139	-0.000[0.019]0.019	0.000[0.018]0.018	-0.005[0.009]0.011
0.8, 0.4, 0.2	GMM	0.001[0.033]0.033	-0.004[0.044]0.044	0.008[0.131]0.131	-0.001[0.018]0.018	-0.001[0.019]0.019	-0.008[0.009]0.012
	EL	-0.000[0.033]0.033	-0.004[0.044]0.044	0.014[0.130]0.131	-0.001[0.018]0.018	0.000[0.019]0.019	-0.003[0.009]0.010
	ET	-0.000[0.033]0.033	-0.004[0.044]0.044	0.013[0.130]0.131	-0.001[0.018]0.018	0.000[0.019]0.019	-0.005[0.009]0.011
0.8, 0.4, 0.4	GMM	-0.000[0.033]0.033	-0.002[0.040]0.040	0.017[0.188]0.189	-0.002[0.018]0.018	-0.001[0.018]0.018	-0.007[0.009]0.012
	EL	-0.002[0.032]0.032	-0.002[0.039]0.040	0.027[0.187]0.189	-0.001[0.018]0.018	0.000[0.018]0.018	-0.002[0.010]0.010
	ET	-0.002[0.032]0.033	-0.002[0.039]0.039	0.027[0.187]0.189	-0.001[0.018]0.018	0.000[0.018]0.018	-0.005[0.009]0.011
0.4, 0.2, 0.2	GMM	-0.001[0.057]0.057	-0.005[0.082]0.082	0.014[0.184]0.185	-0.002[0.043]0.043	-0.001[0.045]0.045	-0.048[0.053]0.071
	EL	-0.003[0.057]0.057	-0.008[0.082]0.082	0.024[0.184]0.185	-0.001[0.043]0.043	0.000[0.045]0.045	-0.019[0.053]0.057
	ET	-0.003[0.056]0.056	-0.008[0.082]0.082	0.024[0.183]0.185	-0.001[0.043]0.043	-0.000[0.045]0.045	-0.034[0.053]0.063
0.4, 0.2, 0.4	GMM	-0.003[0.057]0.057	-0.005[0.070]0.070	0.032[0.214]0.217	-0.004[0.045]0.045	-0.002[0.044]0.044	-0.042[0.056]0.070
	EL	-0.005[0.057]0.057	-0.008[0.070]0.071	0.047[0.214]0.220	-0.003[0.045]0.045	-0.000[0.045]0.045	-0.012[0.057]0.058
	ET	-0.005[0.057]0.057	-0.008[0.070]0.070	0.048[0.214]0.219	-0.003[0.045]0.045	-0.000[0.045]0.045	-0.028[0.056]0.063
0.4, 0.4, 0.2	GMM	-0.000[0.051]0.051	-0.010[0.071]0.072	0.037[0.216]0.219	-0.000[0.044]0.044	-0.000[0.044]0.044	-0.042[0.055]0.069
	EL	-0.004[0.051]0.052	-0.012[0.071]0.072	0.053[0.218]0.224	0.002[0.045]0.045	0.001[0.044]0.044	-0.011[0.055]0.056
	ET	-0.004[0.051]0.051	-0.012[0.071]0.072	0.053[0.217]0.223	0.001[0.044]0.044	0.001[0.044]0.044	-0.027[0.055]0.061
0.4, 0.4, 0.4	GMM	-0.001[0.051]0.051	-0.003[0.061]0.061	0.028[0.303]0.304	-0.003[0.045]0.045	-0.001[0.045]0.045	-0.042[0.060]0.073
	EL	-0.005[0.051]0.051	-0.004[0.060]0.060	0.056[0.297]0.303	-0.002[0.046]0.046	0.000[0.046]0.046	-0.012[0.061]0.062
	ET	-0.004[0.051]0.051	-0.004[0.060]0.060	0.057[0.297]0.302	-0.001[0.045]0.045	0.000[0.045]0.045	-0.027[0.060]0.066

$\beta_0 = [0.5, 0.5, 0.5]'$.

Table 2: Biases, standard errors and RMSEs of estimators for the SARAR(2,0) model (12) in the heteroskedastic case

R^2, κ_{10}, κ_{20}		κ_1	κ_2	β_1	β_2	β_3
	$n = 144$					
0.8, 0.2, 0.2	GMM	0.000[0.046]0.046	-0.000[0.093]0.093	0.002[0.177]0.177	-0.001[0.020]0.020	-0.000[0.020]0.020
	EL	0.001[0.053]0.053	-0.009[0.111]0.111	0.018[0.206]0.207	-0.001[0.021]0.021	0.001[0.022]0.022
	ET	0.001[0.048]0.048	-0.007[0.102]0.103	0.016[0.191]0.192	-0.001[0.020]0.020	0.000[0.020]0.020
0.8, 0.2, 0.4	GMM	-0.000[0.048]0.048	0.001[0.084]0.084	-0.002[0.213]0.213	-0.001[0.021]0.021	-0.000[0.019]0.019
	EL	0.001[0.054]0.054	-0.006[0.098]0.098	0.019[0.248]0.249	-0.001[0.023]0.023	0.000[0.021]0.021
	ET	0.001[0.051]0.051	-0.005[0.091]0.091	0.015[0.231]0.231	-0.001[0.021]0.021	0.000[0.019]0.019
0.8, 0.4, 0.2	GMM	0.001[0.043]0.043	-0.001[0.084]0.084	0.001[0.223]0.223	-0.001[0.020]0.020	-0.001[0.020]0.020
	EL	0.002[0.050]0.050	-0.009[0.102]0.102	0.022[0.270]0.271	-0.001[0.022]0.022	-0.000[0.022]0.022
	ET	0.002[0.046]0.046	-0.008[0.095]0.095	0.019[0.250]0.250	-0.001[0.021]0.021	-0.000[0.021]0.021
0.8, 0.4, 0.4	GMM	0.002[0.045]0.045	-0.002[0.070]0.070	0.003[0.317]0.317	-0.001[0.019]0.019	-0.001[0.020]0.020
	EL	0.003[0.050]0.050	-0.010[0.082]0.083	0.046[0.375]0.378	-0.000[0.021]0.021	-0.000[0.022]0.022
	ET	0.003[0.047]0.047	-0.009[0.077]0.077	0.040[0.346]0.348	-0.000[0.020]0.020	-0.000[0.021]0.021
0.4, 0.2, 0.2	GMM	0.011[0.100]0.100	-0.004[0.183]0.183	-0.014[0.365]0.365	-0.001[0.048]0.048	-0.001[0.049]0.049
	EL	0.009[0.109]0.109	-0.022[0.198]0.199	0.033[0.391]0.393	0.001[0.052]0.052	0.000[0.054]0.054
	ET	0.008[0.102]0.102	-0.018[0.187]0.188	0.025[0.369]0.370	0.001[0.050]0.050	0.000[0.051]0.051
0.4, 0.2, 0.4	GMM	0.003[0.102]0.102	0.009[0.164]0.165	-0.039[0.449]0.450	0.000[0.047]0.047	-0.002[0.047]0.047
	EL	0.001[0.113]0.113	-0.016[0.182]0.183	0.050[0.487]0.490	0.002[0.051]0.051	-0.001[0.052]0.052
	ET	0.003[0.106]0.106	-0.015[0.172]0.173	0.042[0.454]0.456	0.001[0.048]0.048	-0.001[0.048]0.048
0.4, 0.4, 0.2	GMM	0.002[0.096]0.096	0.002[0.166]0.166	-0.010[0.464]0.464	0.000[0.048]0.048	-0.002[0.047]0.047
	EL	-0.003[0.104]0.104	-0.019[0.180]0.181	0.079[0.492]0.498	0.000[0.053]0.053	-0.001[0.052]0.052
	ET	-0.001[0.098]0.098	-0.018[0.171]0.172	0.068[0.461]0.466	0.000[0.049]0.049	-0.001[0.048]0.048
0.4, 0.4, 0.4	GMM	0.005[0.091]0.091	-0.003[0.138]0.138	-0.019[0.671]0.672	-0.002[0.049]0.049	-0.001[0.049]0.049
	EL	0.003[0.101]0.101	-0.020[0.157]0.159	0.117[0.799]0.807	-0.000[0.054]0.054	0.001[0.055]0.055
	ET	0.003[0.095]0.095	-0.019[0.150]0.152	0.108[0.755]0.763	-0.000[0.051]0.051	0.001[0.052]0.052
	$n = 400$					
0.8, 0.2, 0.2	GMM	0.000[0.028]0.028	-0.003[0.059]0.059	0.007[0.107]0.107	-0.000[0.011]0.011	0.000[0.012]0.012
	EL	0.000[0.030]0.030	-0.006[0.063]0.063	0.012[0.114]0.114	-0.000[0.012]0.012	0.000[0.013]0.013
	ET	0.000[0.028]0.028	-0.005[0.060]0.061	0.012[0.109]0.110	-0.000[0.012]0.012	0.000[0.012]0.012
0.8, 0.2, 0.4	GMM	0.001[0.029]0.029	-0.002[0.053]0.053	0.005[0.132]0.132	-0.001[0.012]0.012	-0.000[0.011]0.011
	EL	0.002[0.031]0.031	-0.007[0.057]0.058	0.017[0.143]0.144	-0.000[0.012]0.012	-0.000[0.012]0.012
	ET	0.002[0.030]0.030	-0.006[0.054]0.055	0.015[0.135]0.136	-0.000[0.012]0.012	-0.000[0.012]0.012
0.8, 0.4, 0.2	GMM	0.001[0.026]0.026	-0.002[0.051]0.051	0.007[0.129]0.129	0.000[0.012]0.012	-0.001[0.011]0.011
	EL	0.001[0.028]0.028	-0.004[0.054]0.055	0.013[0.138]0.139	0.001[0.012]0.012	-0.001[0.012]0.012
	ET	0.001[0.027]0.027	-0.004[0.052]0.052	0.012[0.132]0.132	0.000[0.012]0.012	-0.001[0.011]0.011
0.8, 0.4, 0.4	GMM	0.001[0.027]0.027	-0.001[0.042]0.042	0.002[0.176]0.176	-0.001[0.012]0.012	-0.001[0.012]0.012
	EL	0.001[0.029]0.029	-0.003[0.045]0.045	0.012[0.189]0.189	-0.000[0.013]0.013	-0.000[0.012]0.012
	ET	0.001[0.028]0.028	-0.003[0.043]0.043	0.012[0.181]0.181	-0.000[0.012]0.012	-0.000[0.012]0.012
0.4, 0.2, 0.2	GMM	0.002[0.061]0.061	-0.001[0.104]0.104	-0.001[0.217]0.217	-0.002[0.027]0.027	-0.001[0.029]0.029
	EL	0.000[0.064]0.064	-0.008[0.107]0.107	0.021[0.225]0.226	-0.003[0.029]0.029	-0.001[0.030]0.030
	ET	0.000[0.061]0.061	-0.008[0.104]0.104	0.020[0.218]0.218	-0.002[0.028]0.028	-0.001[0.029]0.029
0.4, 0.2, 0.4	GMM	0.001[0.061]0.061	-0.003[0.100]0.100	0.004[0.264]0.264	-0.001[0.028]0.028	-0.001[0.028]0.028
	EL	0.002[0.063]0.063	-0.015[0.103]0.104	0.043[0.275]0.279	-0.001[0.030]0.030	-0.001[0.030]0.030
	ET	0.002[0.061]0.061	-0.014[0.100]0.101	0.039[0.266]0.269	-0.001[0.029]0.029	-0.001[0.029]0.029
0.4, 0.4, 0.2	GMM	0.002[0.054]0.054	0.001[0.095]0.095	-0.011[0.262]0.262	-0.001[0.029]0.029	-0.001[0.028]0.028
	EL	-0.001[0.056]0.056	-0.005[0.099]0.099	0.022[0.272]0.273	-0.001[0.030]0.030	-0.001[0.029]0.029
	ET	-0.001[0.055]0.055	-0.005[0.096]0.096	0.020[0.263]0.264	-0.001[0.029]0.029	-0.001[0.028]0.028
0.4, 0.4, 0.4	GMM	0.001[0.055]0.055	-0.000[0.082]0.082	-0.005[0.372]0.372	-0.000[0.029]0.029	-0.002[0.029]0.029
	EL	0.001[0.058]0.058	-0.009[0.086]0.087	0.055[0.390]0.394	0.001[0.030]0.030	-0.001[0.030]0.030
	ET	0.001[0.056]0.056	-0.008[0.083]0.084	0.052[0.375]0.379	0.000[0.029]0.029	-0.001[0.029]0.029

$\beta_0 = [0.5, 0.5, 0.5]'$.

Table 3: Coverage probabilities of 95% confidence intervals for the SARAR(2,0) model (12)

R^2, κ_{10}, κ_{20}		Homoskedastic case						Heteroskedastic case				
		κ_1	κ_2	β_1	β_2	β_3	σ^2	κ_1	κ_2	β_1	β_2	β_3
	$n = 144$											
0.8, 0.2, 0.2	GMM	0.899	0.902	0.885	0.886	0.907	0.586	0.914	0.864	0.866	0.903	0.925
	EL	0.931	0.946	0.926	0.926	0.927	0.812	0.943	0.916	0.914	0.948	0.945
	ET	0.949	0.956	0.946	0.936	0.951	0.698	0.957	0.935	0.934	0.957	0.958
0.8, 0.2, 0.4	GMM	0.881	0.901	0.888	0.887	0.913	0.621	0.913	0.864	0.891	0.906	0.929
	EL	0.920	0.941	0.940	0.928	0.930	0.833	0.937	0.910	0.925	0.942	0.942
	ET	0.942	0.953	0.956	0.935	0.946	0.735	0.949	0.927	0.941	0.954	0.959
0.8, 0.4, 0.2	GMM	0.888	0.877	0.879	0.898	0.907	0.616	0.898	0.864	0.859	0.921	0.929
	EL	0.932	0.923	0.932	0.934	0.934	0.856	0.925	0.919	0.915	0.950	0.956
	ET	0.955	0.938	0.947	0.949	0.952	0.750	0.947	0.932	0.940	0.960	0.969
0.8, 0.4, 0.4	GMM	0.884	0.877	0.866	0.905	0.909	0.624	0.899	0.872	0.871	0.933	0.915
	EL	0.929	0.922	0.923	0.930	0.929	0.847	0.930	0.922	0.916	0.949	0.949
	ET	0.949	0.934	0.942	0.941	0.946	0.741	0.947	0.942	0.935	0.957	0.952
0.4, 0.2, 0.2	GMM	0.889	0.879	0.878	0.897	0.917	0.591	0.876	0.812	0.850	0.919	0.916
	EL	0.928	0.922	0.933	0.919	0.942	0.827	0.937	0.900	0.926	0.940	0.939
	ET	0.940	0.938	0.948	0.937	0.959	0.722	0.959	0.929	0.941	0.954	0.953
0.4, 0.2, 0.4	GMM	0.874	0.887	0.888	0.905	0.913	0.589	0.879	0.833	0.852	0.942	0.930
	EL	0.913	0.926	0.936	0.931	0.940	0.844	0.932	0.928	0.927	0.955	0.949
	ET	0.931	0.946	0.953	0.939	0.944	0.710	0.951	0.950	0.945	0.964	0.964
0.4, 0.4, 0.2	GMM	0.889	0.862	0.878	0.885	0.906	0.601	0.866	0.837	0.864	0.911	0.920
	EL	0.940	0.917	0.922	0.917	0.931	0.841	0.917	0.920	0.941	0.945	0.950
	ET	0.947	0.934	0.940	0.933	0.947	0.750	0.931	0.945	0.958	0.956	0.963
0.4, 0.4, 0.4	GMM	0.889	0.876	0.873	0.887	0.887	0.629	0.862	0.837	0.851	0.922	0.913
	EL	0.918	0.918	0.920	0.917	0.915	0.840	0.941	0.914	0.914	0.954	0.947
	ET	0.941	0.939	0.936	0.929	0.932	0.758	0.950	0.936	0.938	0.962	0.956
	$n = 400$											
0.8, 0.2, 0.2	GMM	0.938	0.931	0.930	0.938	0.942	0.796	0.933	0.914	0.908	0.942	0.949
	EL	0.954	0.942	0.945	0.948	0.946	0.919	0.943	0.918	0.916	0.939	0.953
	ET	0.962	0.949	0.953	0.955	0.956	0.868	0.947	0.935	0.929	0.947	0.955
0.8, 0.2, 0.4	GMM	0.940	0.932	0.937	0.944	0.936	0.785	0.941	0.919	0.900	0.928	0.943
	EL	0.951	0.947	0.951	0.946	0.944	0.903	0.946	0.928	0.909	0.929	0.938
	ET	0.956	0.951	0.958	0.953	0.952	0.859	0.950	0.938	0.924	0.933	0.947
0.8, 0.4, 0.2	GMM	0.914	0.952	0.936	0.929	0.941	0.791	0.927	0.928	0.911	0.939	0.945
	EL	0.923	0.955	0.949	0.935	0.939	0.922	0.929	0.938	0.921	0.935	0.944
	ET	0.931	0.957	0.956	0.947	0.949	0.878	0.939	0.950	0.931	0.942	0.948
0.8, 0.4, 0.4	GMM	0.926	0.924	0.911	0.951	0.920	0.802	0.921	0.921	0.915	0.933	0.940
	EL	0.935	0.937	0.928	0.952	0.934	0.922	0.924	0.918	0.915	0.930	0.943
	ET	0.943	0.945	0.939	0.954	0.943	0.869	0.931	0.937	0.930	0.940	0.947
0.4, 0.2, 0.2	GMM	0.928	0.928	0.927	0.941	0.915	0.781	0.935	0.899	0.903	0.941	0.944
	EL	0.941	0.932	0.936	0.948	0.921	0.912	0.945	0.932	0.929	0.941	0.948
	ET	0.946	0.944	0.947	0.952	0.932	0.860	0.949	0.942	0.939	0.947	0.951
0.4, 0.2, 0.4	GMM	0.928	0.929	0.938	0.934	0.940	0.789	0.919	0.899	0.914	0.939	0.921
	EL	0.936	0.938	0.949	0.930	0.940	0.906	0.926	0.917	0.933	0.939	0.914
	ET	0.943	0.944	0.952	0.944	0.947	0.863	0.939	0.922	0.943	0.948	0.924
0.4, 0.4, 0.2	GMM	0.933	0.924	0.923	0.932	0.932	0.824	0.933	0.917	0.914	0.949	0.947
	EL	0.938	0.933	0.926	0.936	0.936	0.930	0.935	0.939	0.931	0.950	0.940
	ET	0.944	0.941	0.934	0.944	0.943	0.883	0.948	0.954	0.945	0.955	0.954
0.4, 0.4, 0.4	GMM	0.933	0.929	0.915	0.933	0.941	0.807	0.938	0.906	0.910	0.931	0.938
	EL	0.941	0.937	0.929	0.936	0.944	0.906	0.942	0.942	0.930	0.929	0.939
	ET	0.947	0.951	0.938	0.950	0.956	0.866	0.959	0.953	0.942	0.941	0.944

The variance matrix of a GMM estimator $\hat{\theta}_n$ is computed as $\frac{1}{n}[G_n'(\hat{\theta}_n)\Omega_n^{-1}(\hat{\theta}_n)G_n(\hat{\theta}_n)]^{-1}$, and that of a GEL estimator $\dot{\gamma}_n = (\dot{\theta}_n', \dot{\lambda}_n')'$ is computed as $\frac{1}{n}\Delta_n^{-1}(\dot{\gamma}_n)\begin{pmatrix}0 & 0\\ 0 & \Omega_n(\dot{\theta}_n)\end{pmatrix}\Delta_n^{-1}(\dot{\gamma}_n)$, where $\Delta_n(\gamma)$ is the second order derivative matrix of the GEL objective function given in the proof of Theorem 3.2.

Table 4: Empirical sizes of tests for $\tau_{01} = \tau_{02} = 0$ in the SARAR(0,2) model (13)

	Homoskedastic case						Heteroskedastic case					
	$n = 144$		$n = 400$		$n = 900$		$n = 144$		$n = 400$		$n = 900$	
	$R^2 = 0.8$	$R^2 = 0.4$	$R^2 = 0.8$	$R^2 = 0.4$	$R^2 = 0.8$	$R^2 = 0.4$	$R^2 = 0.8$	$R^2 = 0.4$	$R^2 = 0.8$	$R^2 = 0.4$	$R^2 = 0.8$	$R^2 = 0.4$
PT_{GMM}	0.056	0.039	0.037	0.035	0.043	0.048	0.073	0.056	0.061	0.067	0.047	0.063
PT_{EL}	0.138	0.144	0.059	0.062	0.055	0.056	0.302	0.301	0.142	0.147	0.075	0.091
PT_{ET}	0.116	0.102	0.055	0.057	0.052	0.058	0.201	0.204	0.117	0.126	0.066	0.085
OT_{GMM}	0.049	0.036	0.042	0.035	0.046	0.047	0.040	0.048	0.053	0.050	0.047	0.058
OT_{EL}	0.062	0.049	0.047	0.041	0.047	0.053	0.094	0.095	0.068	0.066	0.055	0.062
OT_{ET}	0.064	0.048	0.044	0.038	0.046	0.052	0.078	0.088	0.071	0.066	0.054	0.062
Moran	0.052	0.040	0.041	0.032	0.044	0.047	0.046	0.053	0.054	0.047	0.047	0.058
Moran_{EL}	0.063	0.051	0.047	0.037	0.045	0.052	0.086	0.091	0.070	0.065	0.055	0.061
Moran_{ET}	0.065	0.048	0.047	0.038	0.045	0.052	0.076	0.086	0.071	0.066	0.055	0.062
PT^*_{GMM}							0.080	0.085	0.115	0.126	0.185	0.215
Moran^*							0.000	0.052	0.001	0.085	0.001	0.099

"PT_{GMM}", "PT_{EL}" and "PT_{ET}" denote, respectively, the GMM, EL and ET parameter restriction tests; "OT_{GMM}", "OT_{EL}" and "OT_{ET}" denote, respectively, the GMM, EL and ET overidentification tests; "Moran", "Moran_{EL} " and "Moran_{ET} " denote, respectively, the robust, EL and ET Moran's I tests; "PT^*_{GMM}" denotes the GMM parameter restriction test without taking into account unknown heteroskedasticity; and "Moran^*" denotes Moran's I test that does not take into account unknown heteroskedasticity. The nominal size is 5%.

Table 5: Powers of tests for $\tau_{01} = \tau_{02} = 0$ in the SARAR(0,2) model (13) in the homoskedastic case

		$\tau_{01} = 0$		$\tau_{01} = 0.2$			$\tau_{01} = 0.4$		
		$\tau_{02} = 0.2$	$\tau_{02} = 0.4$	$\tau_{02} = 0$	$\tau_{02} = 0.2$	$\tau_{02} = 0.4$	$\tau_{02} = 0$	$\tau_{02} = 0.2$	$\tau_{02} = 0.4$
$n = 144$, $R^2 = 0.8$	PT_{GMM}	0.149	0.580	0.326	0.592	0.918	0.883	0.961	0.995
	PT_{EL}	0.299	0.741	0.553	0.778	0.977	0.978	0.995	1.000
	PT_{ET}	0.255	0.702	0.501	0.742	0.968	0.965	0.993	1.000
	OT_{GMM}	0.158	0.631	0.356	0.673	0.972	0.953	0.992	1.000
	OT_{EL}	0.197	0.665	0.423	0.716	0.979	0.964	0.993	1.000
	OT_{ET}	0.186	0.661	0.419	0.714	0.978	0.966	0.993	1.000
	Moran	0.152	0.631	0.357	0.669	0.970	0.946	0.990	1.000
	Moran_{EL}	0.191	0.664	0.418	0.707	0.977	0.966	0.993	1.000
	Moran_{ET}	0.182	0.664	0.419	0.708	0.977	0.967	0.993	1.000
$n = 144$, $R^2 = 0.4$	PT_{GMM}	0.153	0.550	0.336	0.604	0.910	0.898	0.960	0.998
	PT_{EL}	0.305	0.733	0.558	0.805	0.982	0.980	0.994	0.999
	PT_{ET}	0.261	0.706	0.511	0.771	0.968	0.969	0.991	1.000
	OT_{GMM}	0.151	0.623	0.371	0.692	0.963	0.952	0.988	1.000
	OT_{EL}	0.186	0.665	0.436	0.743	0.971	0.970	0.991	1.000
	OT_{ET}	0.183	0.657	0.430	0.737	0.973	0.968	0.992	1.000
	Moran	0.150	0.624	0.371	0.693	0.963	0.950	0.989	1.000
	Moran_{EL}	0.181	0.669	0.440	0.745	0.972	0.971	0.992	1.000
	Moran_{ET}	0.181	0.663	0.437	0.741	0.974	0.967	0.992	1.000
$n = 400$, $R^2 = 0.8$	PT_{GMM}	0.420	0.981	0.822	0.990	1.000	1.000	1.000	1.000
	PT_{EL}	0.488	0.988	0.876	0.995	1.000	1.000	1.000	1.000
	PT_{ET}	0.490	0.988	0.882	0.994	1.000	1.000	1.000	1.000
	OT_{GMM}	0.428	0.985	0.861	0.994	1.000	1.000	1.000	1.000
	OT_{EL}	0.448	0.987	0.866	0.996	1.000	1.000	1.000	1.000
	OT_{ET}	0.447	0.986	0.870	0.996	1.000	1.000	1.000	1.000
	Moran	0.425	0.985	0.862	0.995	1.000	1.000	1.000	1.000
	Moran_{EL}	0.447	0.987	0.867	0.996	1.000	1.000	1.000	1.000
	Moran_{ET}	0.446	0.987	0.875	0.996	1.000	1.000	1.000	1.000
$n = 400$, $R^2 = 0.4$	PT_{GMM}	0.427	0.981	0.842	0.990	1.000	1.000	1.000	1.000
	PT_{EL}	0.509	0.991	0.881	0.992	1.000	1.000	1.000	1.000
	PT_{ET}	0.497	0.991	0.885	0.994	1.000	1.000	1.000	1.000
	OT_{GMM}	0.443	0.987	0.864	0.993	1.000	1.000	1.000	1.000
	OT_{EL}	0.463	0.988	0.870	0.994	1.000	1.000	1.000	1.000
	OT_{ET}	0.466	0.990	0.875	0.995	1.000	1.000	1.000	1.000
	Moran	0.444	0.988	0.861	0.993	1.000	1.000	1.000	1.000
	Moran_{EL}	0.457	0.990	0.871	0.994	1.000	1.000	1.000	1.000
	Moran_{ET}	0.463	0.990	0.874	0.995	1.000	1.000	1.000	1.000

"PT_{GMM}", "PT_{EL}" and "PT_{ET}" denote, respectively, the GMM, EL and ET parameter restriction tests; "OT_{GMM}", "OT_{EL}" and "OT_{ET}" denote, respectively, the GMM, EL and ET overidentification tests; "Moran", "Moran_{EL} " and "Moran_{ET} " denote, respectively, the robust, EL and ET Moran's I tests. The nominal size is 5%.

Table 6: Powers of tests for $\tau_{01} = \tau_{02} = 0$ in the SARAR(0,2) model (13) in the heteroskedastic case

		$\tau_{01} = 0$		$\tau_{01} = 0.2$			$\tau_{01} = 0.4$		
		$\tau_{02} = 0.2$	$\tau_{02} = 0.4$	$\tau_{02} = 0$	$\tau_{02} = 0.2$	$\tau_{02} = 0.4$	$\tau_{02} = 0$	$\tau_{02} = 0.2$	$\tau_{02} = 0.4$
$n = 144$, $R^2 = 0.8$	$\mathrm{PT}_{\mathrm{GMM}}$	0.139	0.348	0.156	0.314	0.669	0.509	0.701	0.953
	$\mathrm{PT}_{\mathrm{EL}}$	0.392	0.653	0.445	0.564	0.837	0.725	0.846	0.981
	$\mathrm{PT}_{\mathrm{ET}}$	0.291	0.573	0.371	0.506	0.807	0.688	0.830	0.983
	$\mathrm{OT}_{\mathrm{GMM}}$	0.091	0.359	0.105	0.220	0.623	0.400	0.607	0.943
	$\mathrm{OT}_{\mathrm{EL}}$	0.153	0.417	0.149	0.267	0.651	0.418	0.640	0.937
	$\mathrm{OT}_{\mathrm{ET}}$	0.137	0.407	0.143	0.265	0.666	0.431	0.648	0.946
	Moran	0.084	0.336	0.096	0.196	0.595	0.376	0.595	0.936
	$\mathrm{Moran}_{\mathrm{EL}}$	0.149	0.409	0.147	0.256	0.648	0.414	0.627	0.933
	$\mathrm{Moran}_{\mathrm{ET}}$	0.141	0.410	0.145	0.260	0.665	0.427	0.648	0.944
$n = 144$, $R^2 = 0.4$	$\mathrm{PT}_{\mathrm{GMM}}$	0.129	0.394	0.158	0.294	0.629	0.459	0.695	0.954
	$\mathrm{PT}_{\mathrm{EL}}$	0.418	0.664	0.434	0.566	0.825	0.688	0.859	0.986
	$\mathrm{PT}_{\mathrm{ET}}$	0.328	0.598	0.331	0.492	0.787	0.649	0.836	0.990
	$\mathrm{OT}_{\mathrm{GMM}}$	0.099	0.394	0.101	0.227	0.597	0.376	0.601	0.949
	$\mathrm{OT}_{\mathrm{EL}}$	0.155	0.446	0.138	0.269	0.624	0.421	0.632	0.944
	$\mathrm{OT}_{\mathrm{ET}}$	0.145	0.447	0.133	0.256	0.632	0.420	0.637	0.952
	Moran	0.090	0.363	0.090	0.209	0.577	0.354	0.571	0.943
	$\mathrm{Moran}_{\mathrm{EL}}$	0.152	0.436	0.134	0.258	0.612	0.410	0.622	0.941
	$\mathrm{Moran}_{\mathrm{ET}}$	0.145	0.444	0.132	0.259	0.626	0.415	0.642	0.950
$n = 400$, $R^2 = 0.8$	$\mathrm{PT}_{\mathrm{GMM}}$	0.209	0.729	0.303	0.639	0.970	0.887	0.977	1.000
	$\mathrm{PT}_{\mathrm{EL}}$	0.341	0.823	0.418	0.723	0.979	0.909	0.978	1.000
	$\mathrm{PT}_{\mathrm{ET}}$	0.301	0.822	0.398	0.710	0.979	0.918	0.985	1.000
	$\mathrm{OT}_{\mathrm{GMM}}$	0.205	0.754	0.236	0.573	0.962	0.813	0.967	1.000
	$\mathrm{OT}_{\mathrm{EL}}$	0.236	0.788	0.248	0.590	0.963	0.809	0.962	1.000
	$\mathrm{OT}_{\mathrm{ET}}$	0.244	0.789	0.254	0.604	0.963	0.826	0.968	1.000
	Moran	0.200	0.746	0.230	0.565	0.962	0.810	0.964	1.000
	$\mathrm{Moran}_{\mathrm{EL}}$	0.232	0.778	0.246	0.586	0.962	0.811	0.959	1.000
	$\mathrm{Moran}_{\mathrm{ET}}$	0.241	0.787	0.254	0.602	0.963	0.827	0.967	1.000
$n = 400$, $R^2 = 0.4$	$\mathrm{PT}_{\mathrm{GMM}}$	0.204	0.725	0.282	0.631	0.957	0.872	0.980	1.000
	$\mathrm{PT}_{\mathrm{EL}}$	0.334	0.837	0.372	0.722	0.972	0.906	0.989	1.000
	$\mathrm{PT}_{\mathrm{ET}}$	0.301	0.832	0.362	0.716	0.975	0.910	0.990	1.000
	$\mathrm{OT}_{\mathrm{GMM}}$	0.195	0.764	0.199	0.572	0.956	0.808	0.966	1.000
	$\mathrm{OT}_{\mathrm{EL}}$	0.221	0.786	0.221	0.574	0.953	0.804	0.965	1.000
	$\mathrm{OT}_{\mathrm{ET}}$	0.215	0.790	0.224	0.588	0.957	0.812	0.967	1.000
	Moran	0.196	0.758	0.191	0.559	0.955	0.798	0.964	1.000
	$\mathrm{Moran}_{\mathrm{EL}}$	0.216	0.783	0.221	0.571	0.952	0.797	0.963	1.000
	$\mathrm{Moran}_{\mathrm{ET}}$	0.215	0.790	0.223	0.591	0.959	0.810	0.967	1.000

"$\mathrm{PT}_{\mathrm{GMM}}$", "$\mathrm{PT}_{\mathrm{EL}}$" and "$\mathrm{PT}_{\mathrm{ET}}$" denote, respectively, the GMM, EL and ET parameter restriction tests; "$\mathrm{OT}_{\mathrm{GMM}}$", "$\mathrm{OT}_{\mathrm{EL}}$" and "$\mathrm{OT}_{\mathrm{ET}}$" denote, respectively, the GMM, EL and ET overidentification tests; "Moran", "$\mathrm{Moran}_{\mathrm{EL}}$ " and "$\mathrm{Moran}_{\mathrm{ET}}$ " denote, respectively, the robust, EL and ET Moran's I tests. The nominal size is 5%.

Table 7: Empirical sizes of spatial J tests for the SARAR model (14)

	$n=144$						$n=400$					
	GMM$_1$	EL$_1$	ET$_1$	GMM$_2$	EL$_2$	ET$_2$	GMM$_1$	EL$_1$	ET$_1$	GMM$_2$	EL$_2$	ET$_2$
Circular vs Queen: Homoskedastic case												
$R^2=0.8$, $\kappa_0=0.2$, $\tau_0=0.2$	0.048	0.068	0.064	0.055	0.063	0.062	0.049	0.059	0.058	0.054	0.060	0.060
$R^2=0.8$, $\kappa_0=0.2$, $\tau_0=0.4$	0.054	0.076	0.075	0.057	0.068	0.062	0.044	0.055	0.055	0.044	0.051	0.050
$R^2=0.8$, $\kappa_0=0.4$, $\tau_0=0.2$	0.054	0.081	0.081	0.058	0.076	0.064	0.057	0.065	0.067	0.056	0.063	0.062
$R^2=0.8$, $\kappa_0=0.4$, $\tau_0=0.4$	0.060	0.094	0.084	0.059	0.076	0.071	0.049	0.056	0.057	0.045	0.054	0.053
$R^2=0.4$, $\kappa_0=0.2$, $\tau_0=0.2$	0.063	0.114	0.105	0.060	0.074	0.071	0.051	0.065	0.064	0.056	0.060	0.060
$R^2=0.4$, $\kappa_0=0.2$, $\tau_0=0.4$	0.080	0.111	0.109	0.062	0.072	0.072	0.065	0.072	0.077	0.049	0.058	0.056
$R^2=0.4$, $\kappa_0=0.4$, $\tau_0=0.2$	0.070	0.103	0.093	0.055	0.081	0.074	0.052	0.063	0.060	0.038	0.047	0.044
$R^2=0.4$, $\kappa_0=0.4$, $\tau_0=0.4$	0.070	0.122	0.113	0.056	0.067	0.063	0.056	0.064	0.064	0.057	0.061	0.057
Circular vs Queen: Heteroskedastic case												
$R^2=0.8$, $\kappa_0=0.2$, $\tau_0=0.2$	0.052	0.129	0.104	0.051	0.106	0.084	0.057	0.086	0.077	0.053	0.079	0.073
$R^2=0.8$, $\kappa_0=0.2$, $\tau_0=0.4$	0.046	0.108	0.085	0.047	0.089	0.076	0.060	0.084	0.080	0.053	0.074	0.065
$R^2=0.8$, $\kappa_0=0.4$, $\tau_0=0.2$	0.052	0.147	0.117	0.063	0.122	0.100	0.056	0.087	0.082	0.056	0.082	0.079
$R^2=0.8$, $\kappa_0=0.4$, $\tau_0=0.4$	0.056	0.117	0.098	0.063	0.095	0.084	0.050	0.078	0.066	0.056	0.074	0.071
$R^2=0.4$, $\kappa_0=0.2$, $\tau_0=0.2$	0.069	0.140	0.115	0.066	0.101	0.089	0.054	0.078	0.076	0.058	0.075	0.077
$R^2=0.4$, $\kappa_0=0.2$, $\tau_0=0.4$	0.060	0.142	0.121	0.068	0.113	0.096	0.051	0.081	0.069	0.052	0.071	0.068
$R^2=0.4$, $\kappa_0=0.4$, $\tau_0=0.2$	0.064	0.151	0.123	0.061	0.110	0.100	0.053	0.079	0.071	0.061	0.076	0.069
$R^2=0.4$, $\kappa_0=0.4$, $\tau_0=0.4$	0.076	0.155	0.120	0.066	0.115	0.092	0.067	0.102	0.092	0.061	0.069	0.066
Queen vs Circular: Homoskedastic case												
$R^2=0.8$, $\kappa_0=0.2$, $\tau_0=0.2$	0.048	0.087	0.074	0.049	0.073	0.068	0.055	0.068	0.065	0.059	0.067	0.069
$R^2=0.8$, $\kappa_0=0.2$, $\tau_0=0.4$	0.055	0.085	0.082	0.054	0.089	0.081	0.049	0.060	0.057	0.051	0.062	0.064
$R^2=0.8$, $\kappa_0=0.4$, $\tau_0=0.2$	0.047	0.077	0.075	0.045	0.059	0.059	0.049	0.054	0.059	0.043	0.051	0.051
$R^2=0.8$, $\kappa_0=0.4$, $\tau_0=0.4$	0.046	0.078	0.070	0.043	0.069	0.058	0.036	0.044	0.044	0.034	0.040	0.041
$R^2=0.4$, $\kappa_0=0.2$, $\tau_0=0.2$	0.059	0.089	0.079	0.044	0.071	0.064	0.048	0.052	0.055	0.043	0.050	0.053
$R^2=0.4$, $\kappa_0=0.2$, $\tau_0=0.4$	0.066	0.097	0.090	0.060	0.073	0.068	0.047	0.058	0.055	0.050	0.056	0.056
$R^2=0.4$, $\kappa_0=0.4$, $\tau_0=0.2$	0.072	0.112	0.093	0.068	0.093	0.080	0.055	0.062	0.056	0.060	0.060	0.060
$R^2=0.4$, $\kappa_0=0.4$, $\tau_0=0.4$	0.076	0.117	0.104	0.069	0.080	0.075	0.058	0.068	0.069	0.057	0.057	0.059
Queen vs Circular: Heteroskedastic case												
$R^2=0.8$, $\kappa_0=0.2$, $\tau_0=0.2$	0.050	0.081	0.075	0.054	0.082	0.074	0.045	0.050	0.051	0.043	0.047	0.048
$R^2=0.8$, $\kappa_0=0.2$, $\tau_0=0.4$	0.063	0.094	0.087	0.064	0.090	0.090	0.050	0.062	0.063	0.054	0.062	0.060
$R^2=0.8$, $\kappa_0=0.4$, $\tau_0=0.2$	0.047	0.082	0.073	0.041	0.067	0.062	0.046	0.054	0.055	0.049	0.056	0.056
$R^2=0.8$, $\kappa_0=0.4$, $\tau_0=0.4$	0.046	0.076	0.069	0.044	0.072	0.067	0.042	0.055	0.051	0.053	0.061	0.060
$R^2=0.4$, $\kappa_0=0.2$, $\tau_0=0.2$	0.059	0.092	0.082	0.043	0.069	0.064	0.059	0.075	0.072	0.062	0.068	0.068
$R^2=0.4$, $\kappa_0=0.2$, $\tau_0=0.4$	0.052	0.081	0.076	0.056	0.059	0.056	0.049	0.055	0.054	0.055	0.058	0.060
$R^2=0.4$, $\kappa_0=0.4$, $\tau_0=0.2$	0.063	0.103	0.090	0.061	0.091	0.082	0.054	0.060	0.061	0.049	0.064	0.063
$R^2=0.4$, $\kappa_0=0.4$, $\tau_0=0.4$	0.066	0.101	0.096	0.062	0.073	0.071	0.062	0.064	0.067	0.066	0.070	0.071

"GMM$_1$" denotes the spatial J test implemented with the GMM distance difference test using the predictor $\hat{Y}_n = \hat{\boldsymbol{\kappa}}_n \mathcal{W}_n Y_n + \mathcal{X}_n \hat{\boldsymbol{\beta}}_n$, and "GMM$_2$" uses $\hat{Y}_n = (I_n - \hat{\boldsymbol{\kappa}}_n \mathcal{W}_n)^{-1} \mathcal{X}_n \hat{\boldsymbol{\beta}}_n$. Correspondingly, we have EL and ET ratio tests "EL$_1$", "EL$_2$", "ET$_1$" and "ET$_2$". "Circular vs Queen" means that an SARAR model with the circular world matrix is tested against one with the queen matrix. "Queen vs Circular" has a similar meaning. The nominal size is 5%.

Table 8: Powers of spatial J tests for the SARAR model (14)

	$n = 144$						$n = 400$					
	GMM_1	EL_1	ET_1	GMM_2	EL_2	ET_2	GMM_1	EL_1	ET_1	GMM_2	EL_2	ET_2
Circular vs Queen: Homoskedastic case												
$R^2 = 0.8$, $\kappa_0 = 0.2$, $\tau_0 = 0.2$	0.361	0.449	0.431	0.367	0.426	0.413	0.731	0.760	0.755	0.732	0.745	0.747
$R^2 = 0.8$, $\kappa_0 = 0.2$, $\tau_0 = 0.4$	0.334	0.426	0.403	0.330	0.391	0.370	0.680	0.710	0.713	0.669	0.693	0.696
$R^2 = 0.8$, $\kappa_0 = 0.4$, $\tau_0 = 0.2$	0.840	0.898	0.890	0.853	0.893	0.882	1.000	1.000	1.000	0.998	0.999	0.999
$R^2 = 0.8$, $\kappa_0 = 0.4$, $\tau_0 = 0.4$	0.758	0.823	0.812	0.728	0.786	0.771	0.990	0.993	0.992	0.987	0.989	0.990
$R^2 = 0.4$, $\kappa_0 = 0.2$, $\tau_0 = 0.2$	0.127	0.188	0.176	0.126	0.146	0.135	0.210	0.223	0.224	0.200	0.216	0.214
$R^2 = 0.4$, $\kappa_0 = 0.2$, $\tau_0 = 0.4$	0.173	0.237	0.225	0.168	0.191	0.180	0.244	0.265	0.261	0.230	0.232	0.235
$R^2 = 0.4$, $\kappa_0 = 0.4$, $\tau_0 = 0.2$	0.363	0.421	0.411	0.325	0.382	0.369	0.586	0.608	0.608	0.570	0.572	0.573
$R^2 = 0.4$, $\kappa_0 = 0.4$, $\tau_0 = 0.4$	0.331	0.418	0.398	0.286	0.325	0.314	0.555	0.591	0.589	0.520	0.526	0.528
Circular vs Queen: Heteroskedastic case												
$R^2 = 0.8$, $\kappa_0 = 0.2$, $\tau_0 = 0.2$	0.414	0.508	0.493	0.414	0.477	0.468	0.747	0.753	0.766	0.742	0.745	0.757
$R^2 = 0.8$, $\kappa_0 = 0.2$, $\tau_0 = 0.4$	0.404	0.492	0.480	0.405	0.445	0.443	0.699	0.710	0.726	0.683	0.695	0.706
$R^2 = 0.8$, $\kappa_0 = 0.4$, $\tau_0 = 0.2$	0.886	0.917	0.918	0.888	0.901	0.905	0.998	0.997	0.998	0.998	0.996	0.997
$R^2 = 0.8$, $\kappa_0 = 0.4$, $\tau_0 = 0.4$	0.774	0.832	0.831	0.765	0.807	0.806	0.991	0.992	0.993	0.989	0.991	0.991
$R^2 = 0.4$, $\kappa_0 = 0.2$, $\tau_0 = 0.2$	0.162	0.248	0.220	0.167	0.200	0.184	0.237	0.266	0.268	0.231	0.248	0.253
$R^2 = 0.4$, $\kappa_0 = 0.2$, $\tau_0 = 0.4$	0.188	0.289	0.251	0.176	0.237	0.218	0.255	0.283	0.292	0.225	0.252	0.252
$R^2 = 0.4$, $\kappa_0 = 0.4$, $\tau_0 = 0.2$	0.365	0.471	0.458	0.339	0.411	0.397	0.636	0.651	0.665	0.617	0.633	0.642
$R^2 = 0.4$, $\kappa_0 = 0.4$, $\tau_0 = 0.4$	0.364	0.444	0.440	0.337	0.408	0.393	0.576	0.581	0.598	0.540	0.547	0.558
Queen vs Circular: Homoskedastic case												
$R^2 = 0.8$, $\kappa_0 = 0.2$, $\tau_0 = 0.2$	0.634	0.736	0.712	0.632	0.711	0.694	0.981	0.984	0.984	0.969	0.976	0.976
$R^2 = 0.8$, $\kappa_0 = 0.2$, $\tau_0 = 0.4$	0.579	0.670	0.654	0.538	0.614	0.599	0.955	0.964	0.966	0.930	0.936	0.940
$R^2 = 0.8$, $\kappa_0 = 0.4$, $\tau_0 = 0.2$	0.975	0.986	0.987	0.965	0.979	0.978	1.000	1.000	1.000	1.000	1.000	1.000
$R^2 = 0.8$, $\kappa_0 = 0.4$, $\tau_0 = 0.4$	0.945	0.974	0.963	0.903	0.944	0.940	1.000	1.000	1.000	0.997	1.000	1.000
$R^2 = 0.4$, $\kappa_0 = 0.2$, $\tau_0 = 0.2$	0.243	0.290	0.276	0.219	0.270	0.261	0.498	0.520	0.518	0.410	0.441	0.444
$R^2 = 0.4$, $\kappa_0 = 0.2$, $\tau_0 = 0.4$	0.251	0.293	0.278	0.211	0.237	0.231	0.539	0.565	0.571	0.376	0.412	0.403
$R^2 = 0.4$, $\kappa_0 = 0.4$, $\tau_0 = 0.2$	0.539	0.605	0.591	0.461	0.528	0.521	0.932	0.934	0.936	0.806	0.829	0.831
$R^2 = 0.4$, $\kappa_0 = 0.4$, $\tau_0 = 0.4$	0.536	0.578	0.554	0.385	0.448	0.435	0.882	0.878	0.877	0.636	0.689	0.683
Queen vs Circular: Heteroskedastic case												
$R^2 = 0.8$, $\kappa_0 = 0.2$, $\tau_0 = 0.2$	0.696	0.757	0.757	0.675	0.711	0.717	0.977	0.978	0.979	0.964	0.968	0.972
$R^2 = 0.8$, $\kappa_0 = 0.2$, $\tau_0 = 0.4$	0.618	0.689	0.682	0.558	0.633	0.620	0.959	0.953	0.959	0.896	0.894	0.906
$R^2 = 0.8$, $\kappa_0 = 0.4$, $\tau_0 = 0.2$	0.983	0.987	0.990	0.972	0.980	0.982	1.000	1.000	1.000	1.000	1.000	1.000
$R^2 = 0.8$, $\kappa_0 = 0.4$, $\tau_0 = 0.4$	0.953	0.977	0.976	0.914	0.942	0.946	1.000	1.000	1.000	0.998	0.996	0.997
$R^2 = 0.4$, $\kappa_0 = 0.2$, $\tau_0 = 0.2$	0.297	0.348	0.341	0.252	0.322	0.300	0.540	0.555	0.560	0.456	0.476	0.484
$R^2 = 0.4$, $\kappa_0 = 0.2$, $\tau_0 = 0.4$	0.283	0.343	0.316	0.199	0.265	0.242	0.541	0.557	0.560	0.389	0.422	0.417
$R^2 = 0.4$, $\kappa_0 = 0.4$, $\tau_0 = 0.2$	0.600	0.634	0.627	0.507	0.576	0.573	0.909	0.901	0.911	0.779	0.796	0.811
$R^2 = 0.4$, $\kappa_0 = 0.4$, $\tau_0 = 0.4$	0.563	0.600	0.591	0.412	0.499	0.481	0.885	0.873	0.882	0.646	0.682	0.705

"GMM_1" denotes the spatial J test implemented with the GMM distance difference test using the predictor $\hat{Y}_n = \hat{\boldsymbol{\kappa}}_n \mathcal{W}_n Y_n + \mathcal{X}_n \hat{\boldsymbol{\beta}}_n$, and "$GMM_2$" uses $\hat{Y}_n = (I_n - \hat{\boldsymbol{\kappa}}_n \mathcal{W}_n)^{-1} \mathcal{X}_n \hat{\boldsymbol{\beta}}_n$. Correspondingly, we have EL and ET ratio tests "EL_1", "EL_2", "ET_1" and "ET_2". "Circular vs Queen" means that an SARAR model with the circular world matrix is tested against one with the queen matrix. "Queen vs Circular" has a similar meaning.