GEL estimation and tests of spatial autoregressive models

Fei Jin[a] and Lung-fei Lee[b*]

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

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

February 9, 2017

Abstract

This paper considers the generalized empirical likelihood (GEL) estimation and tests of 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 a general 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.

Keywords: Spatial autoregressive, martingale, empirical likelihood, higher order asymptotic bias, unknown heteroskedasticity, non-normality, robustness, over-identification test, Moran's I test for spatial dependence, J 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 the popular spatial autoregressive (SAR) model 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

*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).

EL, 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 procedure in estimation and testing. As alternatives to the two-step optimal GMM estimator 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. A nonparametric analog of the parametric likelihood 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. But the EL and GEL approaches have not be considered for estimation and testing with spatially dependent data. These motivate our investigation of the use of EL and GEL for estimation and hypothesis testing with spatial data. We realize that many popular spatial econometric models and hence spatially correlated variables can be characterized by martingale processes under proper filtrations. The importance of martingale processes for spatial random variables has been recognized by Kelejian and Prucha (2001). They develop a central limit theorem (CLT) for linear-quadratic forms of independent disturbances by exploring the martingale structure of a linear-quadratic form.[2] This CLT can be applied to a large class of spatial econometric models such as the SAR model, the spatial error (SE) model, the spatial moving average model, the spatial Durbin model, the spatial error components model, and the SAR model with SAR disturbances (SARAR model).

Various estimation methods for the SARAR model, which includes the SAR and SE models as special cases, have been proposed in the literature, e.g., the generalized spatial two-stage least squares (GS2SLS) estimation (Kelejian and Prucha, 1998), the quasi maximum likelihood (QML) estimation (Lee, 2004), and the GMM estimation (Lee,

[2] In the time series literature, quadratic statistics have long been written as martingales. See, e.g., Hall and Heyde (1980).

2007).[3] The GS2SLS estimates the equation by the two stage least squares (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 SAR models with high order spatial lags. The GMM can be computationally simpler than the QML and can be as efficient as the QML.[4] The GMM may employ not only linear moments in disturbances but also quadratic ones. Quadratic moments can be motivated from the QML and Moran's I test (Moran, 1950), which capture spatial dependence. In the presence of unknown heteroskedasticity, by selecting quadratic matrices with zero diagonals, the quadratic moments can obtain 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. We consider the GEL estimation of the SARAR model under both homoskedasticity and unknown heteroskedasticity in this paper. For spatial data, original sample observations are not martingale differences, so the EL and GEL cannot be applied directly to them. However, as noted in Kelejian and Prucha (2001), linear-quadratic forms of independent disturbances can be written as a sum of martingale differences. For linear and quadratic moments, treating each martingale difference as if it was a data observation, we can set up EL and GEL objective functions to derive corresponding estimates and relevant test statistics.

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 under mixing conditions. 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 2SLS or GS2LS estimation, 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. As far as we know, this may be the first paper that explores the GEL estimation and tests of models with spatial data.[5]

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

[4] 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 QML estimator for the SARAR model with non-normal disturbances (Liu et al., 2010; Lee and Liu, 2010).

[5] Although we only focus on the SARAR model, by exploring martingale structures, other spatial econometric models may be possibly

This paper is organized as follows. Section 2 introduces the SARAR model, and the GEL and GMM estimation based on its martingale structure. Section 3 shows the consistency and asymptotic normality of the GEL estimator and compare 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 desirable finite sample performance of GEL estimators and test statistics. Section 6 concludes. All lemmas and proofs are collected in appendices.

2 The SARAR model and GEL estimation

Consider the SARAR model:

$$Y_n = \kappa W_n Y_n + X_n \beta + U_n, \quad U_n = \tau M_n U_n + V_n, \tag{1}$$

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 parameter vector β, $W_n = (w_{n,ij})$ and $M_n = (m_{n,ij})$ are $n \times n$ nonstochastic spatial weights matrices with zero diagonals, κ and τ 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 two cases on the variance of v_{ni}. In the first case v_{ni}'s are homoskedastic, and in the second case v_{ni}'s have heteroskedastic variances with unknown form. Let $S_n(\kappa) = I_n - \kappa W_n$, and $R_n(\tau) = I_n - \tau M_n$, with I_n being the $n \times n$ identity matrix, and $(\kappa_0, \tau_0, \beta_0')'$ be the true value of $(\kappa, \tau, \beta')'$. 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).[6]

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 estimation can be

$$g_n(\theta) = \frac{1}{n}[V_n'(\theta) P_{n1} V_n(\theta) - \sigma^2 \operatorname{tr}(P_{n1}), \ldots, V_n'(\theta) P_{n,k_p} V_n(\theta) - \sigma^2 \operatorname{tr}(P_{n,k_p}), V_n'(\theta) Q_n]', \tag{2}$$

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 $k_\theta = k_x + 3$, P_{nl} for $l = 1, \ldots, k_p$ are $n \times n$ nonstochastic matrices, and Q_n is an $n \times k_q$ matrix of instrumental variables (IV) with full column rank k_q. Without loss of generality, assume that P_{nl}, for $l = 1, \ldots, k_p$, are symmetric and linearly independent.[7] The quadratic moments are valid since $\mathrm{E}(V_n' P_{nl} V_n) = \sigma_0^2 \operatorname{tr}(P_{nl})$. The IV matrix Q_n may consist of independent columns of $X_n, W_n X_n$ and so on, and P_{ni}'s can be functions of W_n and M_n such as W_n, M_n, W_n^2 and M_n^2. The total number of elements in $g_n(\theta)$ is $k_g = k_p + k_q$, which is greater than or equal to k_θ.

As each moment condition of $g_n(\theta)$ at the true parameter vector θ_0 is either linear or quadratic in V_n, we may also consider a more general vector moment conditions of linear-quadratic forms, which are linearly independent, such as

$$\Xi_n = \frac{1}{n}[V_n' A_{n1} V_n - \sigma_0^2 \operatorname{tr}(A_{n1}) + b_{n1}' V_n, \ldots, V_n' A_{np} V_n - \sigma_0^2 \operatorname{tr}(A_{np}) + b_{np}' V_n]'$$

studied.

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

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

for some finite p, where $A_{nr} = (a_{nr,ij})$ for $r = 1, \ldots, p$ are $n \times n$ symmetric matrices and $b_{nr} = (b_{nr,i})$ for $r = 1, \ldots, p$ are $n \times 1$ vectors. We can rewrite Ξ_n as a sum of martingale differences. Specifically $\Xi_n = \frac{1}{n}\sum_{i=1}^n \xi_{ni}$, where

$$\xi_{ni} = [a_{n1,ii}(v_{ni}^2 - \sigma_0^2) + 2v_{ni}\sum_{j=1}^{i-1} a_{n1,ij}v_{nj} + b_{n1,i}v_{ni}, \ldots, a_{np,ii}(v_{ni}^2 - \sigma_0^2) + 2v_{ni}\sum_{j=1}^{i-1} a_{np,ij}v_{nj} + b_{np,i}v_{ni}]'$$

is a p-dimensional column vector. Consider the σ-fields $\mathscr{F}_{n0} = \{\emptyset, \Omega\}$, $\mathscr{F}_{ni} = \sigma(v_{n1}, \ldots, v_{ni})$, $1 \le i \le n$. As $\mathscr{F}_{n,i-1} \subset \mathscr{F}_{ni}$ and $\mathrm{E}(\xi_{ni}|\mathscr{F}_{n,i-1}) = 0$, $\{\xi_{ni}, \mathscr{F}_{ni}, 1 \le i \le n, n \ge 1\}$ forms a martingale difference array. Thus ξ_{ni}'s are uncorrelated and the variance of Ξ_n is $\frac{1}{n^2}\sum_{i=1}^n \mathrm{E}(\xi_{ni}\xi_{ni}')$. Let $\varphi_n = (\xi_{n1}, \ldots, \xi_{nn})'$ be an $n \times p$ matrix of martingale differences. Then, the variance of Ξ_n is $\frac{1}{n^2}\mathrm{E}(\varphi_n'\varphi_n)$. The moment vector $g_n(\theta)$ in (2) can be equivalent to Ξ_n above when relevant quadratic matrices and linear functions including zeros are properly chosen. Let $Q_n = [Q_{n1}, \ldots, Q_{nn}]'$, $V_n'(\theta)P_{nl}V_n(\theta) - \sigma^2\,\mathrm{tr}(P_{nl}) = \sum_{i=1}^n \omega_{nl,i}(\theta)$ for $l = 1, \ldots, k_p$, where

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

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

$$g_{ni}(\theta) = [\omega_{n1,i}(\theta), \ldots, \omega_{n,k_p,i}(\theta), Q_{ni}'v_{ni}(\theta)]'. \tag{4}$$

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. 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).[8]

In the case that there is unknown heteroskedasticity, we may select all P_{nl}'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_{ni}'s can be W_n, M_n, $W_n^2 - \mathrm{diag}(W_n^2)$, $M_n^2 - \mathrm{diag}(M_n^2)$ and so on, where $\mathrm{diag}(A)$ for a square matrix A denotes a diagonal matrix formed by the diagonal elements of A. Let the moment vector be

$$g_n(\theta) = \frac{1}{n}[V_n'(\theta)P_{n1}V_n(\theta), \ldots, V_n'(\theta)P_{n,k_p}V_n(\theta), V_n'(\theta)Q_n]', \tag{5}$$

where P_{nl}'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 θ is k_θ-dimensional for $k_\theta = k_x + 2$.[9] Then $\omega_{nl,i}(\theta)$ and $g_{ni}(\theta)$ can still have the forms in (3) and (4), as the first term on the r.h.s. of (3) is zero.

[8]Note that $\mathrm{E}(V_n'P_{nl}V_n) - \sigma_0^2\,\mathrm{tr}(P_{nl}) = \mathrm{E}\{V_n'[P_{nl} - \mathrm{tr}(P_{nl})I_n/n]V_n\} + [\mathrm{E}(V_n'V_n)\,\mathrm{tr}(P_{nl})/n - \sigma_0^2\,\mathrm{tr}(P_{nl})] = 0$, where $\mathrm{E}\{V_n'[P_{nl} - \mathrm{tr}(P_{nl})I_n/n]V_n\} = 0$ as $P_{nl} - \mathrm{tr}(P_{nl})I_n/n$ is a matrix with a zero trace. Lee (2001) and Lee (2007) use quadratic moments of the form $\mathrm{E}\{V_n'[P_{nl} - \mathrm{tr}(P_{nl})I_n/n]V_n\} = 0$ to formulate the GMM estimation, which do not involve σ^2. The zero trace quadratic matrices would not be appropriate to be used here due to the required martingale difference property. However the two sets of quadratic moments can be asymptotically equivalent as shown in Liu et al. (2010) for GMM estimation. As we show in Appendix A, in the case that v_{ni}'s are normal, for the GMM in Lee (2001, 2007) and that considered here, there are moment vectors with which the resulting GMM estimators are as efficient as the ML estimator.

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

We consider the GEL estimator:

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

where $\Lambda_n(\theta) = \{\lambda : \lambda' g_{ni}(\theta) \in \mathcal{V}, i = 1, \dots, n\}$ for an open interval $\mathcal{V}$ containing 0, and $\rho(v)$ is a twice continuously differentiable concave function of a scalar v on $\mathcal{V}$.[10] 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 $\sup_n \sup_{1\le i\le n} \mathrm{E}(|v_{ni}|^{4+\iota}) < \infty$.*

Assumption 2. *The elements of X_n are uniformly bounded constants, X_n has full column rank, and $\lim_{n\to\infty}\frac{1}{n}X_n'X_n$ exists and is nonsingular.*

Assumption 3. *(i) W_n and M_n have zero diagonals; (ii) S_n and R_n are nonsingular; and (iii) the sequences of matrices $\{W_n\}$, $\{M_n\}$, $\{S_n^{-1}\}$ and $\{R_n^{-1}\}$ are bounded in both row and column sum norms.*

Assumption 4. *θ_0 is in the interior of a compact parameter space Θ in the k_θ-dimensional Euclidean space.*

Assumption 5. *$\rho(v)$ 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 are the same as those in Lee (2007); and the additional conditions on M_n are similar to those on W_n. Assumption 1(ii) for the heteroskedastic case is the same as that in Lin and Lee (2010). The existence of moments higher than the fourth order in Assumption 1 is needed for the application of the CLT on linear-quadratic forms as in Kelejian and Prucha (2001). In Assumption 2, explanatory variables are assumed to be constants for convenience and multicollinearity is ruled out. Assumption 3 restricts the degree of spatial dependence to be manageable. Assumption 4 is a standard assumption on extremum estimation. Assumption 5 is a smoothness condition on $\rho(\cdot)$ as in Newey and Smith (2004).

We have the interest to compare asymptotic properties of GEL estimation with GMM estimation. Let $\Omega_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} g_{ni}(\theta)g_{ni}'(\theta)$, then $\mathrm{var}[\sqrt{n}g_n(\theta_0)] = \mathrm{E}[\Omega_n(\theta_0)]$. Denote $\bar{\Omega}_n = \mathrm{E}[\Omega_n(\theta_0)]$, which can be estimated 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:

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

[10]In practice, λ can be chosen from $\mathbb{R}^{k_g}$. If for some θ, 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 the λ is not the solution of the problem. This is because $\hat{\lambda}_n = O_p(n^{-1/2})$ by Proposition 3.1, and with probability approaching one, $\lambda' g_{ni}(\theta) \in \mathcal{V}$ for all $1 \le i \le n$, $\theta \in \Theta$ and $\|\lambda\| \le n^{-\zeta}$, where ζ is a positive number, by Lemma C.10.

We shall compare this FOGMM estimator with the GEL estimator. For these estimators, $\bar{\Omega}_n$ is required to be nonsingular in the limit. The nonsingularity of $\bar{\Omega}_n$ will be guaranteed by the linear independence of the linear-quadratic moment conditions. In the limit, we just require such linear independence properties not to vanish.

Assumption 6. $\lim_{n\to\infty}\bar{\Omega}_n$ *exists and is nonsingular.*

For the initial estimator $\tilde{\theta}_n$ for the FOGMM, one may suppose that it is 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 7. $\hat{J}_n=\bar{J}_n+n^{-1/2}\xi_n^J+O_p(n^{-1})$, *where* $\bar{J}_n$ *is a nonstochastic positive definite matrix,* $\lim_{n\to\infty}\bar{J}_n$ *is nonsingular,* $\xi_n^J=O_p(1)$ *and* $\mathrm{E}(\xi_n^J)=0$.

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 orders 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. Under the identification assumptions 11 and 12 in Appendix A, the following proposition 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.

Proposition 3.1. *Under Assumptions 1(i), 2, 3, 5, 6 and 11 in the homoskedastic case, or under Assumptions 1(ii), 2, 3, 5, 6 and 12 in the heteroskedastic case,* $\hat{\theta}_{n,\mathrm{GEL}}\xrightarrow{p}\theta_0$, *and* $g_n(\hat{\theta}_{n,\mathrm{GEL}})=O_p(n^{-1/2})$*; furthermore,* $\hat{\lambda}_{n,\mathrm{GEL}}=\arg\max_{\lambda\in\Lambda_n(\hat{\theta}_{n,\mathrm{GEL}})}\frac{1}{n}\sum_{i=1}^n\rho(\lambda' g_{ni}(\hat{\theta}_{n,\mathrm{GEL}}))$ *exists with probability approaching one (w.p.a.1.), and* $\hat{\lambda}_{n,\mathrm{GEL}}=O_p(n^{-1/2})$.

With the consistency of the GEL estimator, its asymptotic distribution can be derived as usual. Let $\bar{G}_n=\mathrm{E}(\frac{\partial g_n(\theta_0)}{\partial\theta'})$, $\gamma=(\theta',\lambda')'$, and $\gamma_0=(\theta_0',0_{1\times k_g})'$. Furthermore, denote $\bar{\Sigma}_n=(\bar{G}_n'\bar{\Omega}_n^{-1}\bar{G}_n)^{-1}$, $\bar{H}_n=(\bar{G}_n'\bar{\Omega}_n^{-1}\bar{G}_n)^{-1}\bar{G}_n'\bar{\Omega}_n^{-1}$, and $\bar{D}_n=\bar{\Omega}_n^{-1}-\bar{\Omega}_n^{-1}\bar{G}_n(\bar{G}_n'\bar{\Omega}_n^{-1}\bar{G}_n)^{-1}\bar{G}_n'\bar{\Omega}_n^{-1}$.

Assumption 8. $\lim_{n\to\infty}\bar{G}_n$ *has full rank.*

As usual, Assumption 8 rules out functionally dependent moments. The next proposition shows that $\hat{\gamma}_{n,\mathrm{GEL}}=(\hat{\theta}_{n,\mathrm{GEL}}',\hat{\lambda}_{n,\mathrm{GEL}}')'$ is asymptotically normal.

Proposition 3.2. *Under Assumptions 1(i), 2–6, 8 and 11 in the homoskedastic case, or under Assumptions 1(ii), 2–6, 8 and 12 in the heteroskedastic case,* $\sqrt{n}(\hat{\gamma}_{n,\mathrm{GEL}}-\gamma_0)\xrightarrow{d}N\big(0,\lim_{n\to\infty}\mathrm{diag}(\bar{\Sigma}_n,\bar{D}_n)\big)$, *where* $\mathrm{diag}(\bar{\Sigma}_n,\bar{D}_n)$ *is the block diagonal matrix formed by* $\bar{\Sigma}_n$ *and* $\bar{D}_n$.

We see that the GEL estimator $\hat{\theta}_{n,\mathrm{GEL}}$ of θ_0 has the same asymptotic distribution as the GMM estimator $\hat{\theta}_{n,\mathrm{GMM}}$ in (7) (see Propositions A.1 and A.2 in Appendix A).[11]

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

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

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

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

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

The stochastic expansion requires the existence of higher order moments of disturbances.

Assumption 9. $\sup_n \sup_{1\le i\le n} \mathrm{E}\,|v_{ni}|^8 < \infty$.

Proposition 3.3. *For the FOGMM estimator $\hat{\gamma}_{n,\mathrm{GMM}} = (\hat{\theta}_{n,\mathrm{GMM}}', \hat{\lambda}_{n,\mathrm{GMM}}')'$, under Assumptions 1(i), 2–4, 6–9 and 11 in the homoskedastic case, or under Assumptions 1(ii), 2–4, 6–9 and 12 in the heteroskedastic case, the expansion* (8) *holds.*

The explicit forms of ξ_n and ψ_n for the asymptotic expansion of $\hat{\gamma}_{n,\mathrm{GMM}}$ are rather complex, but can be found in Appendix D in the proof of that proposition. A similar expansion for the GEL estimator $\hat{\theta}_{n,\mathrm{GEL}}$ can also be derived as in Appendix D. The expansion requires further smoothness condition on $\rho(v)$.

Assumption 10. $\rho(v)$ *is four times continuously differentiable with Lipschitz fourth derivative in a neighborhood of zero.*

Proposition 3.4. *For the GEL estimator $\hat{\gamma}_{n,\mathrm{GEL}}$, under Assumptions 1(i), 2–6, 8–10 and 11 in the homoskedastic case, or under Assumptions 1(ii), 2–6, 8–10 and 12 in the heteroskedastic case, the expansion* (8) *holds.*

With the above two propositions, we can compute the asymptotic biases of the FOGMM and GEL estimators with the form $\frac{1}{n}\,\mathrm{E}(\psi_n)$. Let $\Omega_n = \Omega_n(\theta_0)$, $G_n = G_n(\theta_0)$, $\bar{G}_n^{(l)} = \mathrm{E}(\frac{\partial G_n(\theta_0)}{\partial \theta_l})$, $g_n = g_n(\theta_0)$, $g_{ni} = g_{ni}(\theta_0)$, $g_{ni}^{(l)} = \frac{\partial g_{ni}(\theta_0)}{\partial \theta_l}$, 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.

[11] Since the GMM estimators in both homoskedastic and heteroskedastic cases have been studied in the literature, we relegate their consistency and asymptotic distribution results to Appendix A and omit their proofs.

Proposition 3.5. *Under Assumptions 1(i), 2–4, 6–9 and 11 in the homoskedastic case, or under Assumptions 1(ii), 2–4, 6–9 and 12 in the heteroskedastic case, 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 $B_n^J = -\sum_{l=1}^{k_\theta} \frac{1}{n^2}\sum_{i=1}^n \bar{H}_n[\mathrm{E}(g_{ni} g_{ni}^{(l)\prime} + g_{ni}^{(l)} g_{ni}')]\bar{D}_n \bar{\Omega}_n (\bar{H}_n^J - \bar{H}_n)' e_{k_\theta,l}$ with $\bar{H}_n^J = (\bar{G}_n' \bar{J}_n^{-1} \bar{G}_n)^{-1} \bar{G}_n' \bar{J}_n^{-1}$.*

In Proposition 3.5, B_n^I is the asymptotic bias for a GMM estimator with the optimal linear combination $\bar{G}_n' \bar{\Omega}_n^{-1} g_n(\theta_0)$ of empirical moments $g_n(\theta_0)$; 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, $\bar{D}_n = 0$; thus, $B_n^G = B_n^\Omega = B_n^J = 0$. Let $G_{ni} = \frac{\partial g_{ni}(\theta_0)}{\partial \theta'} = [g_{ni}^{(1)}, \ldots, g_{ni}^{(k_\theta)}]$.

Proposition 3.6. *Under Assumptions 1(i), 2–6, 8–10 and 11 in the homoskedastic case, or under Assumptions 1(ii), 2–6, 8–10 and 12 in the heteroskedastic case, 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 $\tilde{B}_n^G = -\frac{1}{n^2}\bar{\Sigma}_n \sum_{i=1}^n \mathrm{E}(G_{ni}' \bar{D}_n g_{ni})$, $\rho_3 = \frac{d^3\rho(0)}{dv^3}$ is the third order derivative of $\rho(v)$ evaluated at $v = 0$, and $\tilde{B}_n^\Omega = \frac{1}{n^2}\sum_{i=1}^n \bar{H}_n \mathrm{E}(g_{ni} g_{ni}' \bar{D}_n g_{ni})$.*

Since $g_{ni}(\theta_0)$'s are not independent across i, $B_n^G \neq \tilde{B}_n^G$ and $B_n^\Omega \neq \tilde{B}_n^\Omega$ in general. 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 mixing conditions.

When $g_n(\theta)$ only contains linear moments, g_{ni} becomes $Q_{ni} v_{ni}$. Then, with only IV estimation, $B_n^\Omega = \tilde{B}_n^\Omega$ and the bias of the EL estimator reduces to $B_n^I + B_n^G - \tilde{B}_n^G$, i.e., the EL does not have a bias from estimation of the second moment matrix $\bar{\Omega}_n$. If further $\mathrm{E}(v_{ni}^3) = 0$ for $i = 1, \ldots, n$, because $B_n^\Omega = \tilde{B}_n^\Omega = 0$, 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, not just the EL estimator.

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 $B_n^\Omega = \frac{1}{n^2}\bar{H}_n \sum_{i=1}^n Q_{ni} Q_{ni}' \bar{D}_n Q_{ni} \mathrm{E}(v_{ni}^3)$.*

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 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, GEL test statistics based on quadratic moments with zero diagonal quadratic matrices can be robust to non-normal distributions, while conventional test statistics might not be so if higher order moments are not properly taken into account.

4.1 Test for parameter restrictions

We may test for parameter restrictions in the GEL framework. Let $\theta = (\alpha', \phi')'$, where α is a $k_\alpha \times 1$ subvector of θ, e.g., α might be a vector of spatial dependence parameters κ and/or τ in (1). 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. 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.

Proposition 4.1. *Suppose that Assumptions 2–6 and 8 are satisfied. Then, given Assumptions 1(i) and 11 for the homoskedastic case, or Assumptions 1(ii) and 12 for the heteroskedastic case, under the null hypothesis* $H_0 : \alpha_0 = c_\alpha$,

$$2\Big[\sum_{i=1}^n \rho(\dot{\lambda}'_n g_{ni}(\dot{\theta}_n)) - \sum_{i=1}^n \rho(\hat{\lambda}'_n g_{ni}(\hat{\theta}_n))\Big] \xrightarrow{d} \chi^2(k_\alpha).$$

The GEL ratio test is asymptotically equivalent to the distance difference test in the GMM framework (Donald et al., 2003). 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. In the regard of unknown heteroskedasticity, it has a computational advantage over a Wald test as the latter would require the use of a robust variance estimate as in White (1980).

To understand power properties of this test statistic, we investigate its power under 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, then the GEL ratio statistic can be shown to be asymptotically distributed with a noncentral chi-squared distribution, which is the same as that for a distance difference test in the GMM framework (Newey and West, 1987). 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.

Proposition 4.2. *Suppose that Assumptions 2–6 and 8 are satisfied. Then, given Assumptions 1(i) and 11 for the homoskedastic case, or Assumptions 1(ii) and 12 for the heteroskedastic case, under the Pitman drift* $\alpha_n = c_\alpha + n^{-1/2} d_\alpha$,

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

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.

Proposition 4.3. *Suppose that Assumptions 2–6 and 8 are satisfied. Then under Assumptions 1(i) and 11 in the homoskedastic case, or Assumptions 1(ii) and 12 in the heteroskedastic case,* $2[\sum_{i=1}^n \rho(\hat{\lambda}_n' g_{ni}(\hat{\theta}_n)) - n\rho(0)] \xrightarrow{d} \chi^2(k_g - k_\theta)$*, where the number of moments* k_g *is not less than the number of parameters* k_θ*.*

This GEL overidentification test is asymptotically equivalent to the GMM overidentification test. In general, misspecification of a 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 power test statistics. In a subsequent section, we consider a non-nested test, namely, a J-test, for SAR models with different specified 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 least squares (LS) residual vector $\hat{V}_n = [I_n - X_n(X_n'X_n)^{-1}X_n']Y_n$ 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 $\frac{1}{\sqrt{n}}\hat{V}_n' W_n \hat{V}_n$. 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 but not spatially correlated.[12] Here we show that such a test of spatial dependence can be conveniently implemented in the GEL framework. Such a GEL test can be robust against disturbances with unknown heteroskedasticity, while there is no need to estimate the asymptotic variance of $\frac{1}{\sqrt{n}}\hat{V}' W_n \hat{V}_n$. Let $g_{ni} = v_{ni}\sum_{j=1}^{i-1}(w_{n,ij} + w_{n,ji})v_{nj}$ and $\hat{g}_{ni} = \hat{v}_{ni}\sum_{j=1}^{i-1}(w_{n,ij} + w_{n,ji})\hat{v}_{nj}$, where $\hat{v}_{ni}$ is the ith element of $\hat{V}_n$, for $i = 1, \ldots, n$, and $\hat{\Lambda}_n = \{\lambda : \lambda\hat{g}_{ni} \in \mathcal{V}, i = 1, \ldots, n\}$.[13]

Proposition 4.4. *Suppose that in the regression model* $Y_n = X_n\beta_0 + V_n$ *with zero mean independent disturbances* v_{ni}*'s,* W_n *is an* $n \times n$ *nonstochastic matrix with a zero diagonal and bounded row and column sum norms, and* $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \mathrm{E}(g_{ni}^2) \neq 0$*. Under Assumptions 1, 2 and 5,*

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

[12]Kelejian and Prucha (2001) propose a generalized Moran's I test that cover the SARAR models 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.

[13]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.

The GEL test statistic can use the estimated $\hat{g}_{ni}$ instead of the true g_{ni}, because $\frac{1}{\sqrt{n}}\hat{V}_n'W_n\hat{V}_n$ with the OLS estimated $\hat{V}_n$ has the same asymptotic distribution as $\frac{1}{\sqrt{n}}V_n'W_nV_n$ due to an orthogonality property. Note that the GEL Moran's I test statistic is robust to unknown heteroskedasticity. A conventional Moran's I test would need to evaluate the asymptotic variance of the statistic $\frac{1}{\sqrt{n}}V_n'W_nV_n$ under the null. A robust Moran's I test can be computed as $(\sum_{i=1}^n \hat{g}_{ni}^2)^{-1}(\sum_{i=1}^n \hat{g}_{ni})^2$, given in the above proposition, if we use $\sum_{i=1}^n \hat{g}_{ni}^2$ to estimate the variance of $\sum_{i=1}^n \hat{g}_{ni}$. A GEL version of Moran I's test can bypass such calculations as the GEL takes cares of unknown heteroskedasticity internally.

For the local power of Moran's I test, we consider the alternative model being an SE model, $Y_n = X_n\beta + U_n, U_n = \tau_n U_n + V_n$, where the spatial error dependence parameter is subject to the Pitman drift $\tau_n = n^{-1/2}d_\tau$.

Proposition 4.5. *Suppose that* $Y_n = X_n\beta_0 + U_n$, $U_n = n^{-1/2}d_\tau W_n U_n + V_n$, *where* d_τ *is a constant,* W_n *is an* $n\times n$ *nonstochastic matrix with a zero diagonal and bounded row and column sum norms, and* $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \mathrm{E}(g_{ni}^2) \neq 0$. *Under Assumptions 1, 2 and 5,*

$$2\Big[\max_{\lambda\in\hat{\Lambda}_n}\sum_{i=1}^n \rho(\lambda\hat{g}_{ni}) - n\rho(0)\Big] \xrightarrow{d} \chi^2\Big(1, \lim_{n\to\infty}\{\frac{1}{n}\mathrm{E}[(V_n'W_nV_n)^2]\}^{-1}\{\frac{d_\tau}{n}\mathrm{E}[V_n'(W_n+W_n')W_nV_n]\}^2\Big).$$

We may compare this GEL Moran's I test with the parameter restriction test for spatial error dependence in the SE model based on the moment vector $\frac{1}{n}[V_n'W_nV_n, V_n'X_n]'$. By Propositions 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 uses the estimated moment condition $\frac{1}{\sqrt{n}}\hat{V}_n'W_n\hat{V}_n$, which relies on the null model being a linear regression model. If the model is an SARAR model (1), and the test is for spatial dependence in disturbances, then with consistently estimated residual vector $\hat{V}_n$ such as the estimated residuals from a 2SLS or QML estimated SAR equation, $\frac{1}{\sqrt{n}}\hat{V}_n'W_n\hat{V}_n$ may not have the same asymptotic distribution as $\frac{1}{\sqrt{n}}V_n'W_nV_n$ and the test statistic would not be asymptotically chi-squared distributed. Neither would the GEL test version. This problem occurs due to the issue that the consistent estimator used to construct the moments for testing has an impact on the asymptotic distribution of the moments.[14] To overcome this problem in the GEL framework, we may consider a corresponding $C(\alpha)$-type statistic as suggested in Jin and Lee (2016). Let $\theta = (\alpha, \phi')'$, where α is the spatial error dependence parameter 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 $g_{1n}(\theta) = \frac{1}{n}V_n'(\theta)M_nV_n(\theta)$,[15] where $V_n(\theta) = (I_n - \tau M_n)[(I_n - \kappa W_n)Y_n - X_n\beta]$, 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 $(k_\theta - 1)\times 1$ 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 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 $\sqrt{n}g_n(\theta_0)$.

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

[15]Model (1) can allow for different spatial weights matrices in the spatial lag and spatial error processes, even though in practice they are usually the same. The spatial weights matrix in the spatial error process is M_n, so we have the quadratic moment $\frac{1}{n}V_n'(\theta)M_nV_n(\theta)$.

Proposition 4.6. *For model* (1) *with* $\tau_0 = 0$, *suppose that Assumptions 1–3 and 5 hold, and* $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathrm{E}(g_{ni}^2(\theta_0)) \neq 0$. *Then* $2\Big[\max_{\lambda\in\Lambda_n(\hat\theta_n)} \sum_{i=1}^n \rho(\lambda g_{ni}(\hat\theta_n)) - n\rho(0)\Big] = (\sum_{i=1}^n g_{ni}^2(\hat\theta_n))^{-1}(\sum_{i=1}^n g_{ni}(\hat\theta_n))^2 + o_p(1) \xrightarrow{d} \chi^2(1)$.

The test statistic is readily available with the GEL estimate of λ. It is robust to unknown heteroskedasticity if quadratic matrices in the quadratic moments of $g_{n2}(\theta)$ have zero diagonals. The above GEL test can use any $\sqrt{n}$-consistent estimator $\hat\theta_n$. 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 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 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 SAR 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).[16] In this section, we formulate the spatial J test in the GEL framework.

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

$$Y_n = \kappa_1 W_{1n} Y_n + X_{1n}\beta_1 + U_{1n}, \quad U_{1n} = \tau_1 M_{1n} U_{1n} + V_{1n}, \tag{10}$$

where W_{1n}, M_{1n}, X_{1n} and V_{1n} have similar meanings to those in model (1).[17] 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 $\hat\kappa_{1n}$ and $\hat\beta_{1n}$ be, respectively, estimators of κ_1 and β_1 in (10), which are consistent if model (10) was the true model. The $\hat\kappa_{1n}$ and $\hat\beta_{1n}$ can be the QML, GMM or even GEL estimators.[18] A predictor of Y_n from the alternative model can be either $\hat Y_n = \hat\kappa_{1n} W_{1n} Y_n + X_{1n}\hat\beta_{1n}$ using the main equation of (10) or $\hat Y_n = (I_n - \hat\kappa_{1n} W_{1n})^{-1} X_{1n}\hat\beta_{1n}$ using the reduced form of Y_n under (10). 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 for (1) is based on an augmented model:

$$Y_n = \kappa W_n Y_n + X_n\beta + \eta\hat Y_n + U_n, \quad U_n = \tau M_n U_n + V_n, \tag{11}$$

[16] 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.

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

[18] Large sample properties of the GEL estimators $\hat\kappa_{1n}$ and $\hat\beta_{1n}$ are presented in Appendix B under regularity conditions for misspecified models.

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, we do not reject the alternative model; otherwise, we reject it. In Kelejian and Piras (2011), the spatial J test uses the GS2SLS to estimate the augmented model.[19] When $\hat{Y}_n$ is exogenous, it can be used directly as an extra IV for W_nY_n. For the version that $\hat{Y}_n$ is endogenous, then 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

$$g_n(\vartheta) = [V_n'(\vartheta)P_{n1}V_n(\vartheta) - \sigma^2 \operatorname{tr}(P_{n1}), \ldots, V_n'(\vartheta)P_{n,k_p}V_n(\vartheta) - \sigma^2 \operatorname{tr}(P_{n1}), Q_n'V_n(\vartheta)]$$

in the homoskedastic case, and

$$g_n(\vartheta) = [V_n'(\vartheta)P_{n1}V_n(\vartheta), \ldots, V_n'(\vartheta)P_{n,k_p}V_n(\vartheta), Q_n'V_n(\vartheta)]$$

where each P_{nl}, $l = 1, \ldots, k_p$, has a zero diagonal in the heteroskedastic case. Define $g_{ni}(\vartheta)$ such that $g_n(\vartheta) = \frac{1}{n}\sum_{i=1}^n g_{ni}(\vartheta)$. Under the null, $g_{ni}(\vartheta_0)$'s are martingale differences, 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 test is essentially a test of the parameter restriction that $\eta = 0$ in (11). It differs from the one in the preceding Section 4.1 in that here $\hat{Y}_n$ on the right hand side of (11) is a generated regressor. As the following proposition will show, the initial estimate in $\hat{Y}_n$ does not have an asymptotic impact on the GEL statistic under the null.

Proposition 4.7. *Suppose that Assumptions 2–6, 8, 13 and 14 hold and ϑ_0 is in the interior of the compact parameter space Θ. Then, under Assumptions 1(i) and 11 in the homoskedastic case, or Assumptions 1(ii) and 12 in the heteroskedastic case, $2[\sum_{i=1}^n \rho(\hat{\lambda}_n' g_{ni}(\hat{\theta}_n)) - \max_{\lambda\in\Lambda_n(\hat{\vartheta}_n)} \sum_{i=1}^n \rho(\lambda' g_{ni}(\hat{\vartheta}_n))] \xrightarrow{d} \chi^2(1)$, where $(\hat{\theta}_n', \hat{\lambda}_n')'$ is the GEL estimator for model (1), i.e., it is the restricted GEL estimator 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.

The data generating process is the SARAR model (1) or its restricted form with $\kappa = 0$ and/or $\tau = 0$. 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 $\beta = (\beta_1, \beta_2, \beta_3)'$

[19] Since the original spatial J test uses the GS2SLS to estimate the augmented model, the main equation of (11) is transformed by pre-multiplying it with $(I_n - \hat{\tau}_{1n}M_{1n})$ before estimation, where $\hat{\tau}_{1n}$ is a consistent estimator of τ_{10} (Kelejian and Piras, 2011).

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 the number of nonzero elements in the ith row of the spatial weights matrix W_n, and σ_0^2 is chosen such that $R^2 \equiv \text{var}(X_n\beta_0)/[\text{var}(X_n\beta_0) + \bar{\sigma}_n^2]$ is either 0.4 or 0.8, where $\bar{\sigma}_n^2$ is the average variance of all v_{ni}'s. We set the two spatial weights matrices W_n and M_n to be the same. For GEL estimation and tests other than the spatial J test, W_n is based on the circular world matrix in Arraiz et al. (2010). For the circular world matrix, spatial units are equally spaced on a circle. One third of them are connected to ten nearest neighbors and the rest are connected to two nearest neighbors. For the spatial J test, the null and alternative models only differ in W_n; specifically, the circular matrix and the one based on the queen criterion are tested against each other. These matrices are normalized to have row sums equal to one. For the estimation of model (1), in the homoskedastic case, we use the moment vector $\frac{1}{n}[V_n'V_n - n\sigma_0^2, V_n'W_nV_n, V_n'W_n^2V_n - \sigma_0^2\,\text{tr}(W_n^2), V_n'(X_n, W_nX_n^*, W_n^2X_n^*)]'$, where X_n^* is a submatrix of X_n that excludes the intercept term so that the IV matrix $(X_n, W_nX_n^*, W_n^2X_n^*)$ only contains one intercept; in the heteroskedastic case, we use the moment vector $\frac{1}{n}[V_n'W_nV_n, V_n'(W_n^2 - \text{diag}(W_n^2))V_n, V_n'(X_n, W_nX_n^*, W_n^2X_n^*)]'$. For the spatial J test, the null and alternative models are estimated with moment vectors similar to the above ones. To estimate the augmented model (11), if $\hat{Y}_n = (I_n - \hat{\kappa}_{1n}W_{1n})^{-1}X_{1n}\hat{\beta}_{1n}$ 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 $\hat{Y}_n = \hat{\kappa}_{1n}W_{1n}y_n + X_{1n}\hat{\beta}_{1n}$ is the augmented explanatory variable, $W_{1n}[X_{2n}, X_{3n}]$ is added to the IV matrix. The nominal size of various tests is 0.05. 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.[20] 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. The biases of the EL and ET estimators are smaller than those of the GMM estimator except for some cases, mostly for τ. 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. 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. In terms of standard errors, the ET estimator performs better than the EL estimator, and the GMM estimator generally performs the worst. 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 2 shows summary statistics of the estimators in the heteroskedastic case. The biases are small and their pattern are similar to those in the homoskedastic case. The EL estimator is observed to have larger standard errors and RMSEs than those of the ET estimator. For the parameters κ, τ and the intercept β_1, the ET estimator

[20] 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).

generally has the smallest standard errors and RMSEs, even though for the parameters β_2 and β_3 of regressors, the GMM estimator has the smallest standard errors and RMSEs in some cases.

Table 3 reports coverage probabilities (CP) of 95% confidence intervals for parameters in the SARAR model (1). 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 closer to 95% than GMM ones, and those for σ^2 are about 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.

For Monte Carlo studies on hypothesis testing, nine tests are considered 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 $\frac{1}{n}[V_n'V_n - n\sigma_0^2, V_n'W_nV_n, V_n'W_n^2V_n - \sigma_0^2\operatorname{tr}(W_n^2), V_n'(X_n, W_nX_n^*, W_n^2X_n^*)]'$; "$\text{OT}_{\text{GMM}}$", "$\text{OT}_{\text{EL}}$" and "$\text{OT}_{\text{ET}}$" denote, respectively, the GMM, EL and ET overidentification tests based on the moment vector $\frac{1}{n}[V_n'W_nV_n, V_n'X_n]'$; "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. For the latter three tests, 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 $\frac{1}{n}[V_n'W_nV_n, V_n'(W_n^2 - \operatorname{diag}(W_n^2))V_n, V_n'(X_n, W_nX_n^*, W_n^2X_n^*)]'$ 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 $\frac{1}{n}[V_n'V_n - n\sigma_0^2, V_n'W_nV_n, V_n'W_n^2V_n - \sigma_0^2\operatorname{tr}(W_n^2), V_n'(X_n, W_nX_n^*, W_n^2X_n^*)]'$ and conventional Moran's I test Moran*.

Table 4 presents empirical sizes of tests for $\tau_0 = 0$ in an SE model. PT_{EL} and PT_{ET} have relatively large sizes for small sample cases and have improved sizes for the larger sample size $n = 400$. As expected, PT_{GMM}^* and Moran* have large size distortions and the distortions do not improve with the larger sample size $n = 400$. Other tests have relatively small size distortions. Powers of these tests except PT_{GMM}^* and Moran* are presented in Table 5. Their powers are generally similar for different valid tests, but are higher for the homoskedastic model than those of the heteroskedastic model. R^2 does not have much impact on powers. These tests are powerful in cases with a larger τ_0 and a larger sample size in the data generating process (DGP).

Test results on $\tau_0 = 0$ in the SARAR model (1) are reported in Tables 6 and 7. Parameter restriction tests are based on moment conditions similar to those for the SE model. Overidentification tests are based on the moment vector $\frac{1}{n}[V_n'W_nV_n, V_n'(W_n^2 - \operatorname{diag}(W_n^2))V_n, V_n'(X_n, W_n(I_n - \hat{\kappa}_nW_n)^{-1}X_n\hat{\beta}_n)]'$, where $\hat{\kappa}_n$ and $\hat{\beta}_n$ are the FOGMM estimator of the SAR model as described above. To compute Moran's I tests, we use the 2SLS estimator $\hat{\phi}_n$ of $\phi = (\kappa, \beta')'$ with the IV matrix $Q_n = [X_n, W_nX_n^*, W_n^2X_n^*]$ for the SAR model. The test statistics employ the moment condition $g_n(\theta) = g_{1n}(\theta) - \frac{\partial g_{1n}(\theta)}{\partial\phi'}(\frac{\partial g_{2n}(\theta)}{\partial\phi'})^{-1}g_{2n}(\theta)$, where $g_{1n}(\theta) = \frac{1}{n}V_n'(\theta)W_nV_n(\theta)$ and $g_{2n}(\theta) = \frac{1}{n}Z_n'Q_n(Q_n'Q_n)^{-1}Q_nV_n(\theta)$ with $Z_n = [W_nY_n, X_n]$. Thus $g_{2n}(\hat{\theta}_n) = 0$, where $\hat{\theta}_n = (0, \hat{\phi}_n')'$. When $n = 144$, the size distortions of parameter restriction tests are larger than those of overidentification tests, and those of Moran's I tests are smallest; when $n = 400$, all sizes are general close to the nominal 5%. Different versions of parameter restriction tests have similar powers. So are different versions of overidentification tests and those of Moran's I tests. Parameter restriction tests are more powerful than overidentification tests, and the latter ones are generally more powerful than Moran's I tests. With larger R^2, sample sizes, and τ_0 in the DGP, all tests tend to be more powerful.

Tables 8 and 9 report empirical sizes and powers of spatial J tests for the SARAR model (1). "GMM_1" denotes the spatial J test implemented with the GMM distance difference test using the predictor $\hat{Y}_n = \hat{\kappa}_{1n} W_{1n} Y_n + X_{1n}\hat{\beta}_{1n}$, and "$\text{GMM}_2$" uses $\hat{Y}_n = (I_n - \hat{\kappa}_{1n})^{-1} X_{1n}\hat{\beta}_{1n}$. 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.

[Tables 1–9 about here.]

6 Conclusion

By exploring the martingale structure of the SARAR model, this paper considers its GEL estimation and tests. 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. A general 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. An EL further partially removes the bias from estimating the second moment matrix. We also investigate the GEL overidentification test, Moran's I test, 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 to deal with extra complexity of spatial regression 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.[21]

[21] For SAR models, a "smoothed" (instead of pointwise) asymptotic expansion based on martingales in Mykland (1993) is shown in Jin and Lee (2013).

Appendix A GMM estimation

In this section, we present identification and large sample properties of the GMM estimators for model (1) in both homoskedastic and heteroskedastic cases.

A.1 Homoskedastic case

First, we summarize parameter identification and asymptotic distributions of GMM estimators. The details are in Liu et al. (2010), which are general for high order SARAR models. The summarized results here are the first order SARAR case. As $V_n(\theta)$ is quadratic in θ, we may write

$$\mathrm{E}[Q_n' V_n(\theta)] = Q_n' d_n(\theta),$$

where $d_n(\theta) = R_n(\tau)[(\kappa_0 - \kappa)W_n S_n^{-1} X_n \beta_0 + X_n(\beta_0 - \beta)]$, and

$$\begin{aligned}
&\mathrm{E}[V_n'(\theta) P_{nl} V_n(\theta)] - \sigma^2 \operatorname{tr}(P_{nl}) \\
&= \mathrm{E}\{[d_n(\theta) + R_n(\tau)S_n(\kappa)S_n^{-1}R_n^{-1}V_n]' P_{nl}[d_n(\theta) + R_n(\tau)S_n(\kappa)S_n^{-1}R_n^{-1}V_n]\} - \sigma^2 \operatorname{tr}(P_{nl}) \\
&= d_n'(\theta)P_{nl}d_n(\theta) + \sigma_0^2 \operatorname{tr}\{[R_n(\tau)S_n(\kappa)S_n^{-1}R_n^{-1}]' P_{nl}[R_n(\tau)S_n(\kappa)S_n^{-1}R_n^{-1}]\} - \sigma^2 \operatorname{tr}(P_{nl}) \\
&= d_n'(\theta)P_{nl}d_n(\theta) + (\sigma_0^2 - \sigma^2)\operatorname{tr}(P_{nl}) + 2\sigma_0^2(\tau_0 - \tau)\operatorname{tr}(P_{nl}M_nR_n^{-1}) + \sigma_0^2(\tau_0 - \tau)^2 \operatorname{tr}(R_n'^{-1}M_n'P_{nl}M_nR_n^{-1}) \\
&\quad + 2\sigma_0^2(\kappa_0 - \kappa)\operatorname{tr}(P_{nl}R_nW_nS_n^{-1}R_n^{-1}) + \sigma_0^2(\kappa_0 - \kappa)^2 \operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'R_n'P_{nl}R_nW_nS_n^{-1}R_n^{-1}) \\
&\quad + 2\sigma_0^2(\kappa_0 - \kappa)(\tau_0 - \tau)\operatorname{tr}(R_n'^{-1}M_n'P_{nl}R_nW_nS_n^{-1}R_n^{-1} + P_{nl}M_nW_nS_n^{-1}R_n^{-1}) \\
&\quad + 2\sigma_0^2(\kappa_0 - \kappa)^2(\tau_0 - \tau)\operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'M_n'P_{nl}R_nW_nS_n^{-1}R_n^{-1}) + 2\sigma_0^2(\kappa_0 - \kappa)(\tau_0 - \tau)^2 \operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'M_n'P_{nl}M_nR_n^{-1}) \\
&\quad + \sigma_0^2(\kappa_0 - \kappa)^2(\tau_0 - \tau)^2 \operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'M_n'P_{nl}M_nW_nS_n^{-1}R_n^{-1}).
\end{aligned}$$

As X_n has full rank, we may assume that $\lim_{n\to\infty} \frac{1}{n}Q_n'R_n(\tau)X_n$ has full column rank for any τ in its parameter space. If $\lim_{n\to\infty} \frac{1}{n}Q_n'R_n(\tau)(X_n, W_nS_n^{-1}X_n\beta_0)$ has full column rank for any τ in its parameter space, then β and κ can be identified from the linear moments. As a result, only σ^2 and τ need to be identified from the quadratic moments. If $\lim_{n\to\infty} \frac{1}{n}Q_n'R_n(\tau)(X_n, W_nS_n^{-1}X_n\beta_0)$ does not have full rank for some τ, $\frac{1}{n}Q_n'R_n(\tau)W_nS_n^{-1}X_n\beta_0$ is linearly dependent on $\frac{1}{n}Q_n'R_n(\tau)X_n$ for large enough n. For such τ, let $Q_n'R_n(\tau)W_nS_n^{-1}X_n\beta_0 = Q_n'R_n(\tau)X_n\iota(\tau)$ for some linear coefficient vector $\iota(\tau)$. Then $Q_n'd_n(\theta) = Q_n'R_n(\tau)X_n[\iota(\tau)(\kappa_0 - \kappa) + \beta_0 - \beta]$. Thus, the solutions to $Q_n'd_n(\theta) = 0$ are described by the relation $\beta = \beta_0 + \iota(\tau)(\kappa_0 - \kappa)$ and β_0 is identified as long as κ_0 is identified. The identification of κ_0 can be from the quadratic moments. As $d_n(\theta) = 0$, a sufficient condition is a rank condition given in Assumption 11(*ii*). Let $\Delta_{n1} = [\operatorname{tr}(P_{n1}), \dots, \operatorname{tr}(P_{n,k_p})]'$, $\Delta_{n2} = [\operatorname{tr}(P_{n1}M_nR_n^{-1}), \dots, \operatorname{tr}(P_{n,k_p}M_nR_n^{-1})]'$,

$$\begin{aligned}
&\Delta_{n3} = [\operatorname{tr}(R_n'^{-1}M_n'P_{n1}M_nR_n^{-1}), \dots, \operatorname{tr}(R_n'^{-1}M_n'P_{n,k_p}M_nR_n^{-1})]', \\
&\Delta_{n4} = [\operatorname{tr}(P_{n1}R_nW_nS_n^{-1}R_n^{-1}), \dots, \operatorname{tr}(P_{n,k_p}R_nW_nS_n^{-1}R_n^{-1})]', \\
&\Delta_{n5} = [\operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'R_n'P_{n1}R_nW_nS_n^{-1}R_n^{-1}), \dots, \operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'R_n'P_{n,k_p}R_nW_nS_n^{-1}R_n^{-1})]', \\
&\Delta_{n6} = [\operatorname{tr}(R_n'^{-1}M_n'P_{n1}R_nW_nS_n^{-1}R_n^{-1} + P_{n1}M_nW_nS_n^{-1}R_n^{-1}), \dots, \operatorname{tr}(R_n'^{-1}M_n'P_{n,k_p}R_nW_nS_n^{-1}R_n^{-1} + P_{n,k_p}M_nW_nS_n^{-1}R_n^{-1})]',
\end{aligned}$$

$\Delta_{n7} = [\operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'M_n'P_{n1}R_nW_nS_n^{-1}R_n^{-1}), \ldots, \operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'M_n'P_{n,k_p}R_nW_nS_n^{-1}R_n^{-1})]'$,

$\Delta_{n8} = [\operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'M_n'P_{n1}M_nR_n^{-1}), \ldots, \operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'M_n'P_{n,k_p}M_nR_n^{-1})]'$,

and $\Delta_{n9} = [\operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'M_n'P_{n1}M_nW_nS_n^{-1}R_n^{-1}), \ldots, \operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'M_n'P_{n,k_p}M_nW_nS_n^{-1}R_n^{-1})]'$.

Assumption 11. *(i)* $\lim_{n\to\infty}\frac{1}{n}Q_n'R_n(\tau)(X_n, W_nS_n^{-1}X_n\beta_0)$ *has full column rank for any* τ *in its parameter space, and* $\lim_{n\to\infty}\frac{1}{n}[\Delta_{n1}, \Delta_{n2}, \Delta_{n3}](c_1, 2c_2, c_2^2) \neq 0$ *for any* $(c_1, c_2) \neq 0$*; or (ii)* $\lim_{n\to\infty}\frac{1}{n}Q_n'R_n(\tau)X_n$ *has full column rank for any* τ *in its parameter space, and* $\lim_{n\to\infty}\frac{1}{n}[\Delta_{n1}, \ldots, \Delta_{n9}](c_1, 2c_2, c_2^2, 2c_3, c_3^2, 2c_2c_3, 2c_2c_3^2, 2c_2^2c_3, c_2^2c_3^2)' \neq 0$ *for any* $(c_1, c_2, c_3) \neq 0$.

In the case that $M_n = W_n$, Assumption 11(*ii*) cannot hold, because when $\kappa = \tau_0$ and $\tau = \kappa_0$, $\mathrm{E}[V_n'(\theta)P_{nl}V_n(\theta)] - \sigma^2\operatorname{tr}(P_{nl}) = d_n'(\theta)P_{nl}d_n(\theta) + (\sigma_0^2 - \sigma^2)\operatorname{tr}(P_{nl})$, which implies that $\Delta_{n1}, \ldots, \Delta_{n9}$ are linearly dependent. Identification for the case with $M_n = W_n$ will rely on Assumption 11(*i*).

Next we summarize the asymptotic distributions of GMM estimators. Let $\Upsilon_n = [\operatorname{vec}(P_{n1}), \ldots, \operatorname{vec}(P_{n,k_p})]$ and $\Xi_n = [\operatorname{vec}_{\mathrm{D}}(P_{n1}), \ldots, \operatorname{vec}_{\mathrm{D}}(P_{n,k_p})]$, where $\operatorname{vec}_{\mathrm{D}}(A)$ with a square matrix A denotes a column vector consisting of the diagonal elements of A. Then,

$$\bar{G}_n = -\frac{1}{n}\begin{pmatrix} 2\sigma_0^2\Upsilon_n'\operatorname{vec}(R_nW_nS_n^{-1}R_n^{-1}) & 2\sigma_0^2\Upsilon_n'\operatorname{vec}(M_nR_n^{-1}) & 0 & \Upsilon_n'\operatorname{vec}(I_n) \\ Q_n'R_nW_nS_n^{-1}X_n\beta_0 & 0 & Q_n'R_nX_n & 0 \end{pmatrix},$$

and

$$\bar{\Omega}_n = \operatorname{var}[\sqrt{n}g_n(\theta_0)] = \frac{1}{n}\begin{pmatrix} 2\sigma_0^4\Upsilon_n'\Upsilon_n & 0 \\ 0 & \sigma_0^2Q_n'Q_n \end{pmatrix} + \frac{1}{n}\begin{pmatrix} (\mu_4 - 3\sigma_0^4)\Xi_n'\Xi_n & \mu_3\Xi_n'Q_n \\ \mu_3Q_n'\Xi_n & 0 \end{pmatrix}.$$

Proposition A.1. *(1) Under Assumptions 1(i), 2–4, 7, 8 and 11, the initial GMM estimator* $\tilde{\theta}_n$ *is consistent, and* $\sqrt{n}(\tilde{\theta}_n - \theta_0) \xrightarrow{d} N\Big(0, \lim_{n\to\infty}(\bar{G}_n'\bar{J}_n^{-1}\bar{G}_n)^{-1}\bar{G}_n'\bar{J}_n^{-1}\bar{\Omega}_n\bar{J}_n^{-1}\bar{G}_n(\bar{G}_n'\bar{J}_n^{-1}\bar{G}_n)^{-1}\Big)$;

(2) under Assumptions 1(i), 2–4, and 6–11, the optimal GMM estimator $\hat{\theta}_n$ *is consistent, and*

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

If $\mu_4 - 3\sigma_0^4 = \mu_3 = 0$, e.g., v_{ni}'s are normal, the third and fourth moments of v_{ni} in $\bar{\Omega}_n$ disappear and $\bar{\Omega}_n$ reduces to a block diagonal matrix. Then,

$$\begin{aligned} \bar{G}_n'\bar{\Omega}_n^{-1}\bar{G}_n = {} & \frac{1}{2n\sigma_0^4}[2\sigma_0^2\operatorname{vec}(R_nW_nS_n^{-1}R_n^{-1}), 2\sigma_0^2\operatorname{vec}(M_nR_n^{-1}), 0, \operatorname{vec}(I_n)]'\Upsilon_n(\Upsilon_n'\Upsilon_n)^{-1}\Upsilon_n' \\ & \times [2\sigma_0^2\operatorname{vec}(R_nW_nS_n^{-1}R_n^{-1}), 2\sigma_0^2\operatorname{vec}(M_nR_n^{-1}), 0, \operatorname{vec}(I_n)] \\ & + \frac{1}{n\sigma_0^2}[W_nS_n^{-1}X_n\beta_0, 0, X_n, 0]'R_n'Q_n(Q_n'Q_n)^{-1}Q_n'R_n[W_nS_n^{-1}X_n\beta_0, 0, X_n, 0] \\ \leq {} & \frac{1}{2n\sigma_0^4}[2\sigma_0^2\operatorname{vec}(R_nW_nS_n^{-1}R_n^{-1}), 2\sigma_0^2\operatorname{vec}(M_nR_n^{-1}), 0, \operatorname{vec}(I_n)]' \\ & \times [2\sigma_0^2\operatorname{vec}(R_nW_nS_n^{-1}R_n^{-1}), 2\sigma_0^2\operatorname{vec}(M_nR_n^{-1}), 0, \operatorname{vec}(I_n)] \\ & + \frac{1}{n\sigma_0^2}[W_nS_n^{-1}X_n\beta_0, 0, X_n, 0]'R_n'R_n[W_nS_n^{-1}X_n\beta_0, 0, X_n, 0], \end{aligned}$$

where the inequality follows by the generalized Cauchy-Schwarz inequality, and it becomes an equality when the moment vector is $\frac{1}{n}(V_n'(\theta)V_n(\theta)-n\sigma^2, V_n'(\theta)M_nR_n^{-1}V_n(\theta)-\sigma^2\operatorname{tr}(M_nR_n^{-1}), V_n'(\theta)R_nW_nS_n^{-1}R_n^{-1}V_n(\theta)-\sigma^2\operatorname{tr}(W_nS_n^{-1}), V_n'(\theta)R_n[W_nS_n^{-1}X_n\beta_0, X_n])'$. This is the best moment vector as it yields the smallest asymptotic variance. As in Lee (2007), a feasible moment vector can be obtained by replacing $(\kappa_0,\tau_0,\beta_0')'$ in the best moment vector by a consistent estimator $(\tilde\kappa_n,\tilde\tau_n,\tilde\beta_n')'$ of $(\kappa_0,\tau_0,\beta_0')'$.

A.2 Heteroskedastic case

The identification and asymptotic distributions for GMM estimators in the heteroskedastic case can be derived similarly to that in the homoskedastic case.[22] Let Σ_n be a diagonal matrix consisting of σ_{ni}^2's,

$$\Psi_{n1}=[\operatorname{tr}(P_{n1}M_nR_n^{-1}\Sigma_n),\dots,\operatorname{tr}(P_{n,k_p}M_nR_n^{-1}\Sigma_n)]',$$

$$\Psi_{n2}=[\operatorname{tr}(R_n'^{-1}M_n'P_{n1}M_nR_n^{-1}\Sigma_n),\dots,\operatorname{tr}(R_n'^{-1}M_n'P_{n,k_p}M_nR_n^{-1}\Sigma_n)]',$$

$$\Psi_{n3}=[\operatorname{tr}(P_{n1}R_nW_nS_n^{-1}R_n^{-1}\Sigma_n),\dots,\operatorname{tr}(P_{n,k_p}R_nW_nS_n^{-1}R_n^{-1}\Sigma_n)]',$$

$$\Psi_{n4}=[\operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'R_n'P_{n1}R_nW_nS_n^{-1}R_n^{-1}\Sigma_n),\dots,\operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'R_n'P_{n,k_p}R_nW_nS_n^{-1}R_n^{-1}\Sigma_n)]',$$

$$\Psi_{n5}=[\operatorname{tr}((R_n'^{-1}M_n'P_{n1}R_n+P_{n1}M_n)W_nS_n^{-1}R_n^{-1}\Sigma_n),\dots,\operatorname{tr}((R_n'^{-1}M_n'P_{n,k_p}R_n+P_{n,k_p}M_n)W_nS_n^{-1}R_n^{-1}\Sigma_n)]',$$

$$\Psi_{n6}=[\operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'M_n'P_{n1}R_nW_nS_n^{-1}R_n^{-1}\Sigma_n),\dots,\operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'M_n'P_{n,k_p}R_nW_nS_n^{-1}R_n^{-1}\Sigma_n)]',$$

$$\Psi_{n7}=[\operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'M_n'P_{n1}M_nR_n^{-1}\Sigma_n),\dots,\operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'M_n'P_{n,k_p}M_nR_n^{-1}\Sigma_n)]',$$

and $\Psi_{n8}=[\operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'M_n'P_{n1}M_nW_nS_n^{-1}R_n^{-1}\Sigma_n),\dots,\operatorname{tr}(R_n'^{-1}S_n'^{-1}W_n'M_n'P_{n,k_p}M_nW_nS_n^{-1}R_n^{-1}\Sigma_n)]'$.

Assumption 12. *(i)* $\lim_{n\to\infty}\frac{1}{n}Q_n'R_n(\tau)(X_n, W_nS_n^{-1}X_n\beta_0)$ *has full column rank for any* τ *in its parameter space, and* $\lim_{n\to\infty}\frac{1}{n}[\Psi_{n1},\Psi_{n2}](2c,c^2)\neq 0$ *for any* $c\neq 0$*; or (ii)* $\lim_{n\to\infty}\frac{1}{n}Q_n'R_n(\tau)X_n$ *has full column rank for any* τ *in its parameter space, and* $\lim_{n\to\infty}\frac{1}{n}[\Psi_{n1},\dots,\Psi_{n8}](2c_1,c_1^2,2c_2,c_2^2,2c_1c_2,2c_1c_2^2,2c_1^2c_2,c_1^2c_2^2)'\neq 0$ *for any* $(c_1,c_2)\neq 0$.

Let $\Upsilon_n=[\operatorname{vec}(\Sigma_n^{1/2}P_{n1}\Sigma_n^{1/2}),\dots,\operatorname{vec}(\Sigma_n^{1/2}P_{n,k_p}\Sigma_n^{1/2})]$. With heteroskedastic disturbances,

$$\bar G_n=-\frac{1}{n}\begin{pmatrix}2\Upsilon_n'\operatorname{vec}(\Sigma_n^{-1/2}R_nW_nS_n^{-1}R_n^{-1}\Sigma_n^{1/2}) & 2\Upsilon_n'\operatorname{vec}(\Sigma_n^{-1/2}M_nR_n^{-1}\Sigma_n^{1/2}) & 0\\ Q_n'R_nW_nS_n^{-1}X_n\beta_0 & 0 & Q_n'R_nX_n\end{pmatrix},$$

and

$$\bar\Omega_n=\operatorname{var}[\sqrt{n}g_n(\theta_0)]=\frac{1}{n}\begin{pmatrix}2\Upsilon_n'\Upsilon_n & 0\\ 0 & Q_n'\Sigma_nQ_n\end{pmatrix}.$$

Proposition A.2. *(1) Under Assumptions 1(ii), 2–4, 7, 8 and 12, the initial GMM estimator* $\tilde\theta_n$ *is consistent, and*

$$\sqrt{n}(\tilde\theta_n-\theta_0)\xrightarrow{d}N\big(0,\lim_{n\to\infty}(\bar G_n'\bar J_n^{-1}\bar G_n)^{-1}\bar G_n'\bar J_n^{-1}\bar\Omega_n\bar J_n^{-1}\bar G_n(\bar G_n'\bar J_n^{-1}\bar G_n)^{-1}\big);$$

(2) under Assumptions 1(ii), 2–4, 6–8 and 12, the GMM estimator $\hat\theta_n$ *is consistent, and*

$$\sqrt{n}(\hat\theta_n-\theta_0)\xrightarrow{d}N\big(0,\lim_{n\to\infty}(\bar G_n'\bar\Omega_n^{-1}\bar G_n)^{-1}\big).$$

[22]The results summarized here extend those in Liu et al. (2010) and Lin and Lee (2010) to the first order SARAR case.

Note that

$$\bar{G}_n'\bar{\Omega}_n^{-1}\bar{G}_n = \frac{2}{n}[\text{vec}(\Sigma_n^{-1/2}R_nW_nS_n^{-1}R_n^{-1}\Sigma_n^{1/2}), \text{vec}(\Sigma_n^{-1/2}M_nR_n^{-1}\Sigma_n^{1/2}), 0]'\Upsilon_n(\Upsilon_n'\Upsilon_n)^{-1}\Upsilon_n'$$
$$\times [\text{vec}(\Sigma_n^{-1/2}R_nW_nS_n^{-1}R_n^{-1}\Sigma_n^{1/2}), \text{vec}(\Sigma_n^{-1/2}M_nR_n^{-1}\Sigma_n^{1/2}), 0]$$
$$+ \frac{1}{n}[W_nS_n^{-1}X_n\beta_0, 0, X_n]'R_n'Q_n(Q_n'\Sigma_nQ_n)^{-1}Q_n'R_n[W_nS_n^{-1}X_n\beta_0, 0, X_n].$$

Hence, the best moment vector is

$$\frac{1}{n}\big(V_n'(\theta)[\Sigma_n^{-1}M_nR_n^{-1} - \text{diag}(\Sigma_n^{-1}M_nR_n^{-1})]V_n(\theta), V_n'(\theta)[\Sigma_n^{-1}R_nW_nS_n^{-1}R_n^{-1} - \text{diag}(\Sigma_n^{-1}R_nW_nS_n^{-1}R_n^{-1})]V_n(\theta),$$
$$V_n'(\theta)R_n[W_nS_n^{-1}X_n\beta_0, X_n]\big)'.$$

Since the best moment vector involves Σ_n, it is infeasible unless there are structures on the heteroskedasticity so that each of the variances σ_{ni}^2's can be consistently estimated.

Appendix B Identification conditions 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 of parameter estimates of a model 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 for the J test, 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 conditions

For the estimator $\hat{\theta}_{1n}$ of an alternative model while the null model is the DGP, assume that $\{\theta_{1n}^*\}$ is a sequence of nonstochastic pseudo-true values such that the following convergence is satisfied.

Assumption 13. $\hat{\theta}_{1n} - \theta_{1n}^* = o_p(1)$.

With θ_{1n}^*, the identification condition for the GEL estimation of the augmented model (11) is similar to that for model (1), by taking into account additional terms from the predictor $\hat{Y}_n$, and possible inclusion of the generated regressor $S_{1n}^{-1}(\hat{\kappa}_{1n})X_{1n}\hat{\beta}_{1n}$ in the IV matrix Q_n, when $\hat{Y}_n = S_{1n}^{-1}(\hat{\kappa}_{1n})X_{1n}\hat{\beta}_{1n}$. If $\hat{Y}_n = S_{1n}^{-1}(\hat{\kappa}_{1n})X_{1n}\hat{\beta}_{1n}$, let $\bar{Y}_n^* = S_{1n}^{-1}(\kappa_{1n}^*)X_{1n}\beta_{1n}^*$ and $\epsilon_n^* = 0_{n\times 1}$; if $\hat{Y}_n = \hat{\kappa}_{1n}W_{1n}Y_n + X_{1n}\hat{\beta}_{1n}$, let $\bar{Y}_n^* = \kappa_{1n}^*W_{1n}S_n^{-1}X_n\beta_0 + X_{1n}\beta_{1n}^*$ and $\epsilon_n^* = \kappa_{1n}^*W_{1n}S_n^{-1}R_n^{-1}V_n$. The leading order terms for elements of $\hat{Y}_n$ are elements of $\bar{Y}_n^* + \epsilon_n^*$. For $\hat{Y}_n = \hat{\kappa}_{1n}W_{1n}Y_n + X_{1n}\hat{\beta}_{1n}$, due to the presence of the stochastic part ϵ_n^* and its correlation with vectors linear in V_n, additional terms from quadratic moments appear for identification. For $\hat{Y}_n = S_{1n}^{-1}(\hat{\kappa}_{1n})X_{1n}\hat{\beta}_{1n}$, note that $\frac{1}{n}V_n'A_nS_{1n}^{-1}(\hat{\kappa}_{1n})X_{1n}\hat{\beta}_{1n} = o_p(1)$ and $\frac{1}{n}b_n'A_nS_{1n}^{-1}(\hat{\kappa}_{1n})X_{1n}\hat{\beta}_{1n} = \frac{1}{n}b_n'A_nS_{1n}^{-1}(\kappa_{1n}^*)X_{1n}\beta_{1n}^* + o_p(1)$, where A_n is an $n\times n$ nonstochastic matrix bounded in both row and column sum norms and b_n is an $n\times 1$ vector of uniformly bounded constants. Then $S_{1n}^{-1}(\hat{\kappa}_{1n})X_{1n}\hat{\beta}_{1n}$ is asymptotically exogenous and the identification conditions involving linear moments only need to be modified to account for the randomness of Q_n. Let Q_n^* be the matrix obtained by replacing $S_{1n}^{-1}(\hat{\kappa}_{1n})X_{1n}\hat{\beta}_{1n}$ with $S_{1n}^{-1}(\kappa_{1n}^*)X_{1n}\beta_{1n}^*$ in Q_n if it is in Q_n, and $Q_n^* = Q_n$ otherwise. With Q_n^*, the identification conditions involving linear moments are similar to those for (1). Define $\Delta_{n,10} = [\mathrm{E}(\epsilon_n^{*'}R_n'P_{n1}R_n\epsilon_n^*), \ldots, \mathrm{E}(\epsilon_n^{*'}R_n'P_{n,k_p}R_n\epsilon_n^*)]'$, $\Delta_{n,11} = [\mathrm{E}(\epsilon_n^{*'}M_n'P_{n1}R_n\epsilon_n^*), \ldots, \mathrm{E}(\epsilon_n^{*'}M_n'P_{n,k_p}R_n\epsilon_n^*)]'$, $\Delta_{n,12} = [\mathrm{E}(\epsilon_n^{*'}M_n'P_{n1}M_n\epsilon_n^*), \ldots, \mathrm{E}(\epsilon_n^{*'}M_n'P_{n,k_p}M_n\epsilon_n^*)]'$,

$$\Delta_{n,13} = [\mathrm{E}(\epsilon_n^{*'}R_n'P_{n1}V_n), \ldots, \mathrm{E}(\epsilon_n^{*'}R_n'P_{n,k_p}V_n)]',$$

$$\Delta_{n,14} = [\mathrm{E}(\epsilon_n^{*'}R_n'P_{n1}M_nR_n^{-1}V_n + \epsilon_n^{*'}M_n'P_{n1}V_n), \ldots, \mathrm{E}(\epsilon_n^{*'}R_n'P_{n,k_p}M_nR_n^{-1}V_n + \epsilon_n^{*'}M_n'P_{n,k_p}V_n)]',$$

$$\Delta_{n,15} = [\mathrm{E}(\epsilon_n^{*'}M_n'P_{n1}M_nR_n^{-1}V_n), \ldots, \mathrm{E}(\epsilon_n^{*'}M_n'P_{n,k_p}M_nR_n^{-1}V_n)]',$$

$$\Delta_{n,16} = [\mathrm{E}(\epsilon_n^{*'}R_n'P_{n1}R_nW_nS_n^{-1}R_n^{-1}V_n), \ldots, \mathrm{E}(\epsilon_n^{*'}R_n'P_{n,k_p}R_nW_nS_n^{-1}R_n^{-1}V_n)]',$$

$$\Delta_{n,17} = [\mathrm{E}(\epsilon_n^{*'}(R_n'P_{n1}M_n + M_n'P_{n1}R_n)W_nS_n^{-1}R_n^{-1}V_n), \ldots, \mathrm{E}(\epsilon_n^{*'}(R_n'P_{n,k_p}M_n + M_n'P_{n,k_p}'R_n)W_nS_n^{-1}R_n^{-1}V_n)]',$$

and $\Delta_{n,18} = [\mathrm{E}(\epsilon_n^{*'}M_n'P_{n1}M_nW_nS_n^{-1}R_n^{-1}V_n), \ldots, \mathrm{E}(\epsilon_n^{*'}M_n'P_{n,k_p}M_nW_nS_n^{-1}R_n^{-1}V_n)]'$. The explicit forms for the above terms can be easily derived for both the homoskedastic and heteroskedastic cases. We omit them for simplicity.

Assumption 14. *(1) If* $\hat{Y}_n = S_{1n}^{-1}(\hat{\kappa}_{1n})X_{1n}\hat{\beta}_{1n}$,

(I) $\{S_{1n}(\kappa_{1n}^*)\}$ *are invertible and* $\{S_{1n}^{-1}(\kappa_{1n}^*)\}$ *are bounded in both row and column sum norms;*

(II) in the homoskedastic case, either (i) $\lim_{n\to\infty}\frac{1}{n}Q_n^{*'}R_n(\tau)(X_n, W_nS_n^{-1}X_n\beta_0, \bar{Y}_n^*)$ *has full column rank for any* τ *in its parameter space, and* $\lim_{n\to\infty}\frac{1}{n}[\Delta_{n1}, \Delta_{n2}, \Delta_{n3}](c_1, 2c_2, c_2^2) \neq 0$ *for any* $(c_1, c_2) \neq 0$*; or (ii)* $\lim_{n\to\infty}\frac{1}{n}Q_n^{*'}R_n(\tau)(X_n, \bar{Y}_n^*)$ *has full column rank for any* τ *in its parameter space, and*

$$\lim_{n\to\infty}\frac{1}{n}[\Delta_{n1}, \ldots, \Delta_{n9}](c_1, 2c_2, c_2^2, 2c_3, c_3^2, 2c_2c_3, 2c_2c_3^2, 2c_2^2c_3, c_2^2c_3^2)' \neq 0$$

for any $(c_1, c_2, c_3) \neq 0$*;*

(III) in the heteroskedastic case, either (i) $\lim_{n\to\infty}\frac{1}{n}Q_n^{*'}R_n(\tau)(X_n, W_nS_n^{-1}X_n\beta_0, \bar{Y}_n^*)$ *has full column rank for any* τ *in its parameter space, and* $\lim_{n\to\infty}\frac{1}{n}[\Psi_{n1}, \Psi_{n2}](2c, c^2) \neq 0$ *for any* $c \neq 0$*; or (ii)* $\lim_{n\to\infty}\frac{1}{n}Q_n^{*'}R_n(\tau)(X_n, \bar{Y}_n^*)$ *has full column rank for any* τ *in its parameter space, and*

$$\lim_{n\to\infty}\frac{1}{n}[\Psi_{n1}, \ldots, \Psi_{n8}](2c_1, c_1^2, 2c_2, c_2^2, 2c_1c_2, 2c_1c_2^2, 2c_1^2c_2, c_1^2c_2^2)' \neq 0$$

for any $(c_1, c_2) \neq 0$.

(2) If $\hat{Y}_n = \hat{\kappa}_{1n}W_{1n}Y_n + X_{1n}\hat{\beta}_{1n}$,

(I) in the homoskedastic case, either (i) $\lim_{n\to\infty}\frac{1}{n}Q_n'R_n(\tau)(X_n, W_nS_n^{-1}X_n\beta_0, \bar{Y}_n^*)$ *has full column rank for any* τ *in its parameter space, and*

$$\lim_{n\to\infty}\frac{1}{n}[\Delta_{n1}, \Delta_{n2}, \Delta_{n3}, \Delta_{n,10}, \ldots, \Delta_{n,15}](c_1, 2c_2, c_2^2, \eta^2, 2c_2\eta^2, c_2^2\eta^2, -2\eta, -2c_2\eta, -2c_2^2\eta)' \neq 0$$

for any $(c_1, c_2, \eta) \neq 0$*; or (ii)* $\lim_{n\to\infty}\frac{1}{n}Q_n'R_n(\tau)(X_n, \bar{Y}_n^*)$ *has full column rank for any* τ *in its parameter space, and* $\lim_{n\to\infty}\frac{1}{n}[\Delta_{n1}, \ldots, \Delta_{n9}, \Delta_{n,10}, \ldots, \Delta_{n,18}]c(c_1, c_2, c_3, \eta) \neq 0$*, where* $c(c_1, c_2, c_3, \eta) =$

$(c_1, 2c_2, c_2^2, 2c_3, c_3^2, 2c_2c_3, 2c_2c_3^2, 2c_2^2c_3, c_2^2c_3^2, \eta^2, 2c_2\eta^2, c_2^2\eta^2, -2\eta, -2c_2\eta, -2c_2^2\eta, -2c_3\eta, -2c_2c_3\eta, -2c_2^2c_3\eta)'$, *for any* $(c_1, c_2, c_3, \eta) \neq 0$*;*

(II) in the heteroskedastic case, either (i) $\lim_{n\to\infty} \frac{1}{n} Q_n' R_n(\tau)(X_n, W_n S_n^{-1} X_n \beta_0, \bar{Y}_n^*)$ *has full column rank for any* τ *in its parameter space, and* $\lim_{n\to\infty} \frac{1}{n}[\Psi_{n1}, \Psi_{n2}, \Delta_{n,10}, \ldots, \Delta_{n,15}](2c, c^2, \eta^2, 2c\eta^2, c^2\eta^2, -2\eta, -2c\eta, -2c^2\eta)' \neq 0$ *for any* $(c, \eta) \neq 0$*; or (ii)* $\lim_{n\to\infty} \frac{1}{n} Q_n' R_n(\tau)(X_n, \bar{Y}_n^*)$ *has full column rank for any* τ *in its parameter space, and* $\lim_{n\to\infty} \frac{1}{n}[\Psi_{n1}, \ldots, \Psi_{n8}, \Delta_{n,10}, \ldots, \Delta_{n,18}]c(c_1, c_2, \eta) \neq 0$*, where*

$$c(c_1, c_2, \eta) = (2c_1, c_1^2, 2c_2, c_2^2, 2c_1c_2, 2c_1c_2^2, 2c_1^2c_2, c_1^2c_2^2, \eta^2, 2c_1\eta^2, c_1^2\eta^2, -2\eta, -2c_1\eta, -2c_1^2\eta, -2c_2\eta, -2c_1c_2\eta, -2c_1^2c_2\eta)',$$

for any $(c_1, c_2, \eta) \neq 0$*.*

B.2 GEL estimation of the alternative model

For the alternative model (10), let $V_{1n}(\theta_1) = (I_n - \tau_1 M_{1n})[(I_n - \kappa_1 W_{1n})Y_n - X_{1n}\beta_1]$, the moment vector can be

$$g_{1n}(\theta_1) = \frac{1}{n}[V_{1n}'(\theta_1)P_{1n,1}V_{1n}(\theta_1) - \sigma_1^2 \operatorname{tr}(P_{1n,1}), \ldots, V_{1n}'(\theta_1)P_{1n,k_{p1}}V_{1n}(\theta_1) - \sigma_1^2 \operatorname{tr}(P_{1n,k_{p1}}), Q_{1n}'V_{1n}(\theta_1)],$$

if elements of V_{1n} were assumed to be i.i.d., where $P_{1n,1}, \ldots, P_{1n,k_{p1}}$ are $n \times n$ spatial weights matrices and Q_{1n} is an $n \times k_{q1}$ IV matrix. On the other hand, if elements of V_{1n} were independent but heteroskedastic, the moment vector for consistent estimation would be

$$g_{1n}(\theta_1) = \frac{1}{n}[V_{1n}'(\theta_1)P_{1n,1}V_{1n}(\theta_1), \ldots, V_{1n}'(\theta_1)P_{1n,k_{p1}}V_{1n}(\theta_1), Q_{1n}'V_{1n}(\theta_1)],$$

where $P_{1n,i}$'s now have zero diagonals. With the moment vector $g_{1n}(\theta_1)$, define $g_{1n,i}(\theta_1)$ in a way similar to $g_{ni}(\theta)$ in Section 2 with the intension to capture the martingale difference property. The GEL estimators are

$$\hat{\theta}_{1n} = \arg\min_{\theta_1 \in \Theta_1} \max_{\lambda_1 \in \Lambda_1} \sum_{i=1}^n \rho(\lambda_1' g_{1n,i}(\theta_1)), \text{ and } \hat{\lambda}_{1n} = \arg\max_{\lambda_1 \in \Lambda_1} \sum_{i=1}^n \rho(\lambda_1' g_{1n,i}(\hat{\theta}_{1n})),$$

where Θ_1 and Λ_1 are compact.[23] Suppose that there exist pseudo true values $\theta_{1n}^* \in \Theta_1$ and $\lambda_{1n}^* \in \Lambda_1$ such that

$$\mathrm{E} \sum_{i=1}^n \rho(\lambda_{1n}^{*'} g_{1n,i}(\theta_{1n}^*)) = \min_{\theta_1 \in \Theta_1} \max_{\lambda_1 \in \Lambda_1} \mathrm{E} \sum_{i=1}^n \rho(\lambda_1' g_{1n,i}(\theta_1)).$$

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

Assumption 15. *(i)* Θ_1 *and* Λ_1 *are compact, and* $\mathcal{V}$ *includes all realizations of* $\lambda_1' g_{1n,i}(\theta_1)$ *for all* $1 \leq i \leq n$, $\lambda_1 \in \Lambda_1$ *and* $\theta_1 \in \Theta_1$*; (ii)* $\sup_{\lambda_1 \in \Lambda_1, \theta_1 \in \Theta_1} \frac{1}{n} |\sum_{i=1}^n \rho(\lambda_1' g_{1n,i}(\theta_1)) - \mathrm{E} \sum_{i=1}^n \rho(\lambda_1' g_{1n,i}(\theta_1))| = o_p(1)$*; (iii)* $\frac{1}{n} \mathrm{E} \sum_{i=1}^n \rho(\lambda_1' g_{1n,i}(\theta_1))$ *is uniformly equicontinuous on* (Θ_1, Λ_1)*; (iv) for each* $\theta_1 \in \Theta_1$*, the identifiably unique maximizer* $\lambda_{1n}^*(\theta_1) \in \Lambda_1$ *of* $\arg\max_{\lambda_1 \in \Lambda_1} \mathrm{E} \sum_{i=1}^n \rho(\lambda_1' g_{1n,i}(\theta_1))$ *is equicontinuous in* θ_1*;*[24] *(v)* $\mathrm{E} \sum_{i=1}^n \rho(\lambda_{1n}^{*'}(\theta_1) g_{1n,i}(\theta_1))$ *has identifiably unique minimizer* $\theta_{1n}^* \in \Theta_1$*.*

[23] For analytical convenience, the parameter space of λ_1 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 $\rho(\cdot)$.

[24] $\lambda_{1n}^*(\theta_1)$ is identifiably unique if for all $\epsilon > 0$, $\limsup_{n\to\infty} [\max_{\lambda_1 \in B_n^c(\epsilon)} \frac{1}{n} \mathrm{E} \sum_{i=1}^n \rho(\lambda_1' g_{1n,i}(\theta_1)) - \frac{1}{n} \mathrm{E} \sum_{i=1}^n \rho(\lambda_{1n}^{*'}(\theta_1) g_{1n,i}(\theta_1))] < 0$, where $B_n^c(\epsilon)$ is the complement in Λ_1 of an open ball $B_n(\epsilon)$ centered at $\lambda_{1n}^*(\theta_1)$ with radius ϵ (White, 1994).

The above assumption gives high level conditions similar to those in Hong et al. (2003).[25] Some conditions might be relaxed, e.g., the uniform convergence condition in Assumption 15(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. With Assumption 15, it justifies the convergence of the GEL estimates for the alternative model to their pseudo true values as in Assumption 13, where the null model is the DGP.

Proposition B.1. *Under Assumption 15,* $\hat{\theta}_{1n} - \theta^*_{1n} = o_p(1)$ *and* $\hat{\lambda}_{1n} - \lambda^*_{1n} = o_p(1)$.

Appendix C Lemmas related to GMM and GEL estimation

C.1 General lemmas on martingale differences of linear and quadratic moments

Lemma C.1. *Let* $u_{nl} = [u_{nl,i}]$ *be* $n \times 1$ *random vectors,* $D_{nl}(\theta) = [d_{nl,ij}(\theta)]$ *be* $n \times n$ *nonstochastic matrices whose elements are functions of* $\theta \in \Theta$, *for* $l = 1, \ldots, s$, *and* $b_n(\theta) = [b_{ni}(\theta)]$ *and* $c_n(\theta) = [c_{ni}(\theta)]$ *be* $n \times 1$ *nonstochastic vectors whose elements are functions of* $\theta \in \Theta$ *such that* $\sup_{\theta\in\Theta} \|b_n(\theta)\|_\infty = O(1)$ *and* $\sup_{\theta\in\Theta} \|c_n(\theta)\|_\infty = O(1)$. *Then,*

(i) if $D_{n1}(\theta)$ *is bounded in either row or column sum norm uniformly on* Θ, $\sup_{\theta\in\Theta} \frac{1}{n}|b_n'(\theta)D_{n1}(\theta)c_n(\theta)| = O(1)$;

(ii) if $\sup_{1\le j\le n} \mathrm{E}\,|u_{n1,j}| = O(1)$, $\sup_{\theta\in\Theta} |\frac{1}{n} b_n'(\theta)u_{n1}| = O_p(1)$;

(iii) if $D_{nl}(\theta)$*'s are bounded in row sum norm uniformly in* $\theta \in \Theta$ *and there are* $a_l > 1$ *for* $l = 1, \ldots, s$ *such that* $\sup_{1\le l\le s} \sup_{1\le j\le n} \mathrm{E}\,|u_{nl,j}|^{a_l} = O(1)$, *then* $\sup_{\theta\in\Theta} \sup_{1\le i\le n} |\prod_{l=1}^s \sum_{j=1}^n d_{nl,ij}(\theta)u_{nl,j}| = O_p(n^{\sum_{l=1}^s \frac{1}{a_l}})$, *and* $\sup_{\theta\in\Theta} \frac{1}{n}|\sum_{i=1}^n \prod_{l=1}^s \sum_{j=1}^n d_{nl,ij}(\theta)u_{nl,j}| = O_p(1)$ *if* $\sum_{l=1}^s \frac{1}{a_l} \le 1$;

(iv) if $D_{nl}(\theta)$*'s are bounded in row sum norm uniformly in* $\theta \in \Theta$ *and there are* $a_l > 1$ *for* $l = 1, \ldots, s$ *such that* $\sup_{1\le l\le s} \sup_{1\le j\le n} \mathrm{E}\,|u_{nl,j}|^{a_l} = O(1)$, *then* $\sup_{\theta\in\Theta} \sup_{1\le i\le n} \mathrm{E}\,|\prod_{l=1}^s \sum_{j=1}^n d_{nl,ij}(\theta)u_{nl,j}|^{1/\sum_{l=1}^s \frac{1}{a_l}} = O(1)$.

Proof. (i) If $D_{n1}(\theta)$ is bounded in row sum norm uniformly in $\theta \in \Theta$, then

$$\sup_{\theta\in\Theta} \frac{1}{n}|b_n'(\theta)D_{n1}(\theta)c_n(\theta)| \le \sup_{\theta\in\Theta} \frac{1}{n}\|b_n'(\theta)\|_\infty \|D_{n1}(\theta)\|_\infty \|c_n(\theta)\|_\infty \le \sup_{\theta\in\Theta} \frac{1}{n}\|b_n'(\theta)\|_\infty \sup_{\theta\in\Theta}\|D_{n1}(\theta)\|_\infty \sup_{\theta\in\Theta}\|c_n(\theta)\|_\infty = O(1).$$

Similarly, we can show the result when $D_{n1}(\theta)$ is bounded in column sum norm uniformly in $\theta \in \Theta$.

(ii) $\mathrm{E}(\sup_{\theta\in\Theta} |\frac{1}{n} b_n'(\theta)u_n|) \le \mathrm{E}(\sup_{\theta\in\Theta} \frac{1}{n}\|b_n'(\theta)\|_1 \|u_n\|_1) = \sup_{\theta\in\Theta} \|b_n'(\theta)\|_1 (\frac{1}{n}\sum_{i=1}^n \mathrm{E}\,|u_{ni}|) = O(1)$.

(iii) There exists a finite $c_l > 0$ such that $\frac{1}{c_l} + \frac{1}{a_l} = 1$. By Hölder's inequality,

$$\Big|\sum_{j=1}^n d_{nl,ij}(\theta)u_{nl,j}\Big| \le \sum_{j=1}^n |d_{nl,ij}(\theta)|^{\frac{1}{c_l}} |d_{nl,ij}(\theta)|^{\frac{1}{a_l}} |u_{nl,j}|$$

[25]Among the regularity conditions, uniform convergence of the GEL objective function is assumed. With a misspecified model, the proof strategy of Proposition 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.

$$\leq \Big(\sum_{j=1}^{n} |d_{nl,ij}(\theta)|\Big)^{1/c_l} \Big(\sum_{j=1}^{n} |d_{nl,ij}(\theta)| \cdot |u_{nl,j}|^{a_l}\Big)^{1/a_l}$$
$$\leq c^{1/c_l} \Big(\sum_{l=1}^{s}\sum_{j=1}^{n} |d_{nl,ij}(\theta)| \cdot |u_{nl,j}|^{a_l}\Big)^{1/a_l},$$

where $c = \sup_n \sup_{1\leq l\leq s} \sup_{\theta\in\Theta} \sup_{1\leq i\leq n} \sum_{j=1}^{n} |d_{nl,ij}(\theta)| < \infty$. Then,

$$\Big|\prod_{l=1}^{s}\sum_{j=1}^{n} d_{nl,ij}(\theta)u_{nl,j}\Big| \leq c^{\sum_{l=1}^{s}\frac{1}{c_l}} \Big(\sum_{l=1}^{s}\sum_{j=1}^{n} |d_{nl,ij}(\theta)| \cdot |u_{nl,j}|^{a_l}\Big)^{\sum_{l=1}^{s}\frac{1}{a_l}}. \tag{12}$$

Thus $\sup_{1\leq i\leq n} |\Pi_{l=1}^{s} \sum_{j=1}^{n} d_{nl,ij}(\theta)u_{nl,j}| \leq c^{\sum_{l=1}^{s}\frac{1}{c_l}} (\sum_{l=1}^{s}\sum_{j=1}^{n} (\sup_{1\leq i\leq n, 1\leq j\leq n} |d_{nl,ij}(\theta)|)|u_{nl,j}|^{a_l})^{\sum_{l=1}^{s}\frac{1}{a_l}} \leq c^{\sum_{l=1}^{s}\frac{1}{c_l}} (\sum_{l=1}^{s}\sum_{j=1}^{n} |u_{nl,j}|^{a_l} \sup_{\theta\in\Theta, 1\leq l\leq s} \|D_{nl}(\theta)\|_\infty)^{\sum_{l=1}^{s}\frac{1}{a_l}}$, where $\frac{1}{n}\sum_{l=1}^{s}\sum_{j=1}^{n} |u_{nl,j}|^{a_l} = O_p(1)$ by Markov's inequality. Hence the first result holds. When $\sum_{l=1}^{s}\frac{1}{a_l} \leq 1$, by (12) and Jensen's inequality, $|\frac{1}{n}\sum_{i=1}^{n} \Pi_{l=1}^{s}\sum_{j=1}^{n} d_{nl,ij}(\theta)u_{nl,j}| \leq c^{\sum_{l=1}^{s}\frac{1}{c_l}} (\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{s}\sum_{j=1}^{n} |d_{nl,ij}(\theta)|\cdot|u_{nl,j}|^{a_l})^{\sum_{l=1}^{s}\frac{1}{a_l}} \leq c^{\sum_{l=1}^{s}\frac{1}{c_l}} (\frac{1}{n}\sum_{l=1}^{s}\sum_{j=1}^{n} |u_{nl,j}|^{a_l} \sup_{\theta\in\Theta, 1\leq l\leq s} \|D_{nl}(\theta)\|_\infty)^{\sum_{l=1}^{s}\frac{1}{a_l}}$. Thus the second result follows.

(*iv*) By (12),

$$\sup_{\theta\in\Theta}\sup_{1\leq i\leq n} \mathrm{E}\Big|\prod_{l=1}^{s}\sum_{j=1}^{n} d_{nl,ij}(\theta)u_{nl,j}\Big|^{1/\sum_{l=1}^{s}\frac{1}{a_l}} \leq c^{(\sum_{l=1}^{s}\frac{1}{c_l})/\sum_{l=1}^{s}\frac{1}{a_l}} \sup_{\theta\in\Theta}\sup_{1\leq i\leq n} \sum_{l=1}^{s}\sum_{j=1}^{n} |d_{nl,ij}(\theta)|\,\mathrm{E}\,|u_{nl,j}|^{a_l} = O(1).$$

Hence the result holds. ☐

The following lemma is useful to show orders of terms in Nagar-type expansions of GMM and GEL estimators. In particular, it is used to prove Lemma C.7.

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

(i) for $r_{ni}^{(l)} = a_{nl,ii}(v_{ni}^2 - \sigma_{ni}^2) + b_{nl,ii}v_{ni} + (c_{nl,ii} + d_{nl,ii}v_{ni})\sum_{j=1}^{i-1} e_{nl,ij}v_{nj} + \sum_{j=1}^{i-1} g_{nl,ij}v_{nj}\sum_{k=1}^{j-1} h_{nl,ik}v_{nk}$ with $l = 1$ and 2, if $\sup_n \sup_{1\leq i\leq n} \mathrm{E}(v_{ni}^4) < \infty$, 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

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

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

Proof. (*i*) We shall prove the results for the simplified $r_{ni}^{(l)} = b_{nl,ii}v_{ni} + (c_{nl,ii} + d_{nl,ii}v_{ni})\sum_{j=1}^{i-1} e_{nl,ij}v_{nj} + \sum_{j=1}^{i-1} g_{nl,ij}v_{nj}\sum_{k=1}^{j-1} h_{nl,ik}v_{nk}$ for $l = 1$ and 2, and point out the results with the original $r_{ni}^{(l)}$'s hold similarly. Since $\mathrm{E}(r_{ni}^{(1)}r_{ni}^{(2)}) = \sigma_{ni}^2 b_{n1,ii}b_{n2,ii} + (c_{n1,ii}c_{n2,ii} + \sigma_{ni}^2 d_{n1,ii}d_{n2,ii})\sum_{j=1}^{i-1} e_{n1,ij}e_{n2,ij}\sigma_{nj}^2 + \sum_{j=1}^{i-1} g_{n1,ij}g_{n2,ij}\sigma_{nj}^2\sum_{k=1}^{j-1} h_{n1,ik}h_{n2,ik}\sigma_{nk}^2$,

$$\sup_n \sup_{1\leq i\leq n} |\mathrm{E}(r_{ni}^{(1)}r_{ni}^{(2)})| \leq \sup_n \sup_{1\leq i\leq n} \Big[\sigma_{ni}^2|b_{n1,ii}b_{n2,ii}| + (|c_{n1,ii}c_{n2,ii}| + \sigma_{ni}^2|d_{n1,ii}d_{n2,ii}|)\sum_{j=1}^{i-1} |e_{n1,ij}e_{n2,ij}\sigma_{nj}^2|$$

$$+\sum_{j=1}^{i-1}|g_{n1,ij}g_{n2,ij}\sigma_{nj}^2|\sum_{k=1}^{j-1}|h_{n1,ik}h_{n2,ik}\sigma_{nk}^2|\Big]<c,$$

for some constant c. Thus, $\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(r_{ni}^{(1)}r_{ni}^{(2)})=O(1)$. To prove the convergence of $\frac{1}{n}\sum_{i=1}^{n}[r_{ni}^{(1)}r_{ni}^{(2)}-\mathrm{E}(r_{ni}^{(1)}r_{ni}^{(2)})]$, rewrite $r_{ni}^{(1)}r_{ni}^{(2)}-\mathrm{E}(r_{ni}^{(1)}r_{ni}^{(2)})=\Delta_{n1,i}+\Delta_{n2,i}$, where

$$\begin{aligned}
\Delta_{n1,i}=\;&b_{n1,ii}b_{n2,ii}(v_{ni}^2-\sigma_{ni}^2)+[b_{n1,ii}c_{n2,ii}v_{ni}+b_{n1,ii}d_{n2,ii}(v_{ni}^2-\sigma_{ni}^2)]\sum_{j=1}^{i-1}e_{n2,ij}v_{nj}\\
&+[b_{n2,ii}c_{n1,ii}v_{ni}+b_{n2,ii}d_{n1,ii}(v_{ni}^2-\sigma_{ni}^2)]\sum_{j=1}^{i-1}e_{n1,ij}v_{nj}\\
&+b_{n1,ii}v_{ni}\sum_{j=1}^{i-1}g_{n2,ij}v_{nj}\sum_{k=1}^{j-1}h_{n2,ik}v_{nk}+b_{n2,ii}v_{ni}\sum_{j=1}^{i-1}g_{n1,ij}v_{nj}\sum_{k=1}^{j-1}h_{n1,ik}v_{nk}\\
&+[(c_{n1,ii}d_{n2,ii}+c_{n2,ii}d_{n1,ii})v_{ni}+d_{n1,ii}d_{n2,ii}(v_{ni}^2-\sigma_{ni}^2)]\Big(\sum_{j=1}^{i-1}e_{n1,ij}e_{n2,ij}(v_{nj}^2-\sigma_{nj}^2)\\
&+\sum_{j=1}^{i-1}\sum_{k=1}^{j-1}(e_{n1,ij}e_{n2,ik}+e_{n2,ij}e_{n1,ik})v_{nj}v_{nk}\Big)\\
&+[(c_{n1,ii}d_{n2,ii}+c_{n2,ii}d_{n1,ii})v_{ni}+d_{n1,ii}d_{n2,ii}(v_{ni}^2-\sigma_{ni}^2)]\sum_{j=1}^{i-1}e_{n1,ij}e_{n2,ij}\sigma_{nj}^2\\
&+d_{n1,ii}v_{ni}\sum_{j=1}^{i-1}e_{n1,ij}g_{n2,ij}(v_{nj}^2-\sigma_{nj}^2)\sum_{k=1}^{j-1}h_{n2,ik}v_{nk}+d_{n2,ii}v_{ni}\sum_{j=1}^{i-1}e_{n2,ij}g_{n1,ij}(v_{nj}^2-\sigma_{nj}^2)\sum_{k=1}^{j-1}h_{n1,ik}v_{nk}\\
&+d_{n1,ii}v_{ni}\sum_{j=1}^{i-1}g_{n2,ij}v_{nj}\sum_{k=1}^{j-1}e_{n1,ik}h_{n2,ik}(v_{nk}^2-\sigma_{nk}^2)+d_{n2,ii}v_{ni}\sum_{j=1}^{i-1}g_{n1,ij}v_{nj}\sum_{k=1}^{j-1}e_{n2,ik}h_{n1,ik}(v_{nk}^2-\sigma_{nk}^2)\\
&+d_{n1,ii}v_{ni}\sum_{j=1}^{i-1}e_{n1,ij}g_{n2,ij}\sigma_{nj}^2\sum_{k=1}^{j-1}h_{n2,ik}v_{nk}+d_{n2,ii}v_{ni}\sum_{j=1}^{i-1}e_{n2,ij}g_{n1,ij}\sigma_{nj}^2\sum_{k=1}^{j-1}h_{n1,ik}v_{nk}\\
&+d_{n1,ii}v_{ni}\sum_{j=1}^{i-1}g_{n2,ij}v_{nj}\sum_{k=1}^{j-1}e_{n1,ik}h_{n2,ik}\sigma_{nk}^2+d_{n2,ii}v_{ni}\sum_{j=1}^{i-1}g_{n1,ij}v_{nj}\sum_{k=1}^{j-1}e_{n2,ik}h_{n1,ik}\sigma_{nk}^2\\
&+d_{n1,ii}v_{ni}\sum_{j=1}^{i-1}\sum_{k=1}^{j-1}\sum_{l=1}^{k-1}v_{nj}v_{nk}v_{nl}(e_{n1,ij}g_{n2,ik}h_{n2,il}+g_{n2,ij}e_{n1,ik}h_{n2,il}+g_{n2,ij}h_{n2,ik}e_{n1,il})\\
&+d_{n2,ii}v_{ni}\sum_{j=1}^{i-1}\sum_{k=1}^{j-1}\sum_{l=1}^{k-1}v_{nj}v_{nk}v_{nl}(e_{n2,ij}g_{n1,ik}h_{n1,il}+g_{n1,ij}e_{n2,ik}h_{n1,il}+g_{n1,ij}h_{n1,ik}e_{n2,il}),
\end{aligned}$$

and

$$\begin{aligned}
\Delta_{n2,i}=\;&b_{n1,ii}d_{n2,ii}\sigma_{ni}^2\sum_{j=1}^{i-1}e_{n2,ij}v_{nj}+b_{n2,ii}d_{n1,ii}\sigma_{ni}^2\sum_{j=1}^{i-1}e_{n1,ij}v_{nj}\\
&+(c_{n1,ii}c_{n2,ii}+\sigma_{ni}^2d_{n1,ii}d_{n2,ii})\Big(\sum_{j=1}^{i-1}e_{n1,ij}e_{n2,ij}(v_{nj}^2-\sigma_{nj}^2)+\sum_{j=1}^{i-1}\sum_{k=1}^{j-1}(e_{n1,ij}e_{n2,ik}+e_{n2,ij}e_{n1,ik})v_{nj}v_{nk}\Big)\\
&+c_{n1,ii}\sum_{j=1}^{i-1}e_{n1,ij}g_{n2,ij}(v_{nj}^2-\sigma_{nj}^2)\sum_{k=1}^{j-1}h_{n2,ik}v_{nk}+c_{n2,ii}\sum_{j=1}^{i-1}e_{n2,ij}g_{n1,ij}(v_{nj}^2-\sigma_{nj}^2)\sum_{k=1}^{j-1}h_{n1,ik}v_{nk}\\
&+c_{n1,ii}\sum_{j=1}^{i-1}g_{n2,ij}v_{nj}\sum_{k=1}^{j-1}e_{n1,ik}h_{n2,ik}(v_{nk}^2-\sigma_{nk}^2)+c_{n2,ii}\sum_{j=1}^{i-1}g_{n1,ij}v_{nj}\sum_{k=1}^{j-1}e_{n2,ik}h_{n1,ik}(v_{nk}^2-\sigma_{nk}^2)
\end{aligned}$$

$$+ c_{n1,ii}\sum_{j=1}^{i-1} e_{n1,ij}g_{n2,ij}\sigma_{nj}^2 \sum_{k=1}^{j-1} h_{n2,ik}v_{nk} + c_{n2,ii}\sum_{j=1}^{i-1} e_{n2,ij}g_{n1,ij}\sigma_{nj}^2 \sum_{k=1}^{j-1} h_{n1,ik}v_{nk}$$

$$+ c_{n1,ii}\sum_{j=1}^{i-1} g_{n2,ij}v_{nj} \sum_{k=1}^{j-1} e_{n1,ik}h_{n2,ik}\sigma_{nk}^2 + c_{n2,ii}\sum_{j=1}^{i-1} g_{n1,ij}v_{nj} \sum_{k=1}^{j-1} e_{n2,ik}h_{n1,ik}\sigma_{nk}^2$$

$$+ c_{n1,ii}\sum_{j=1}^{i-1}\sum_{k=1}^{j-1}\sum_{l=1}^{k-1} v_{nj}v_{nk}v_{nl}(e_{n1,ij}g_{n2,ik}h_{n2,il} + g_{n2,ij}e_{n1,ik}h_{n2,il} + g_{n2,ij}h_{n2,ik}e_{n1,il})$$

$$+ c_{n2,ii}\sum_{j=1}^{i-1}\sum_{k=1}^{j-1}\sum_{l=1}^{k-1} v_{nj}v_{nk}v_{nl}(e_{n2,ij}g_{n1,ik}h_{n1,il} + g_{n1,ij}e_{n2,ik}h_{n1,il} + g_{n1,ij}h_{n1,ik}e_{n2,il})$$

$$+ \sum_{j=1}^{i-1} g_{n1,ij}g_{n2,ij}(v_{nj}^2 - \sigma_{nj}^2)\sum_{k=1}^{j-1} h_{n1,ik}h_{n2,ik}(v_{nk}^2 - \sigma_{nk}^2)$$

$$+ \sum_{j=1}^{i-1} g_{n1,ij}g_{n2,ij}\sigma_{nj}^2\sum_{k=1}^{j-1} h_{n1,ik}h_{n2,ik}(v_{nk}^2 - \sigma_0^2) + \sum_{j=1}^{i-1} g_{n1,ij}g_{n2,ij}(v_{nj}^2 - \sigma_0^2)\sum_{k=1}^{j-1} h_{n1,ik}h_{n2,ik}\sigma_{nk}^2$$

$$+ \sum_{j=1}^{i-1} g_{n1,ij}g_{n2,ij}(v_{nj}^2 - \sigma_{nj}^2)\sum_{k=1}^{j-1}\sum_{l=1}^{k-1}(h_{n1,ik}h_{n2,il} + h_{n1,il}h_{n2,ik})v_{nk}v_{nl}$$

$$+ \sum_{j=1}^{i-1} g_{n1,ij}g_{n2,ij}\sigma_{nj}^2\sum_{k=1}^{j-1}\sum_{l=1}^{k-1}(h_{n1,ik}h_{n2,il} + h_{n1,il}h_{n2,ik})v_{nk}v_{nl}$$

$$+ \sum_{j=1}^{i-1} g_{n2,ij}v_{nj}\sum_{k=1}^{j-1} g_{n1,ik}h_{n2,ik}(v_{nk}^2 - \sigma_{nk}^2)\sum_{l=1}^{k-1} h_{n1,il}v_{nl}$$

$$+ \sum_{j=1}^{i-1} g_{n1,ij}v_{nj}\sum_{k=1}^{j-1} g_{n2,ik}h_{n1,ik}(v_{nk}^2 - \sigma_{nk}^2)\sum_{l=1}^{k-1} h_{n2,il}v_{nl}$$

$$+ \sum_{j=1}^{i-1} g_{n2,ij}v_{nj}\sum_{k=1}^{j-1} g_{n1,ik}h_{n2,ik}\sigma_{nk}^2\sum_{l=1}^{k-1} h_{n1,il}v_{nl} + \sum_{j=1}^{i-1} g_{n1,ij}v_{nj}\sum_{k=1}^{j-1} g_{n2,ik}h_{n1,ik}\sigma_{nk}^2\sum_{l=1}^{k-1} h_{n2,il}v_{nl}$$

$$+ \sum_{j=1}^{i-1}\sum_{k=1}^{j-1}(g_{n1,ij}g_{n2,ik} + g_{n2,ij}g_{n1,ik})v_{nj}v_{nk}\sum_{l=1}^{k-1} h_{n1,il}h_{n2,il}(v_{nl}^2 - \sigma_{nl}^2)$$

$$+ \sum_{j=1}^{i-1}\sum_{k=1}^{j-1}(g_{n1,ij}g_{n2,ik} + g_{n2,ij}g_{n1,ik})v_{nj}v_{nk}\sum_{l=1}^{k-1} h_{n1,il}h_{n2,il}\sigma_{nl}^2$$

$$+ \sum_{j=1}^{i-1}\sum_{k=1}^{j-1}\sum_{l=1}^{k-1}\sum_{m=1}^{l-1} v_{nj}v_{nk}v_{nl}v_{nm}(g_{n1,ij}h_{n1,ik}g_{n2,il}h_{n2,im} + g_{n1,ij}g_{n2,ik}h_{n1,il}h_{n2,im}$$

$$+ g_{n1,ij}g_{n2,ik}h_{n2,il}h_{n1,im} + g_{n2,ij}g_{n1,ik}h_{n1,il}h_{n2,im} + g_{n2,ij}g_{n1,ik}h_{n2,il}h_{n1,im} + g_{n2,ij}h_{n2,ik}g_{n1,il}h_{n1,im}).$$

Note that $\Delta_{n1,i}$'s are martingale differences, and $\Delta_{n2,i}$ only involves $v_{n1}, \ldots, v_{n,i-1}$. Each term in $\Delta_{n1,i}$ has the form $\Pi_{l=1}^{s}\sum_{j=1}^{n} p_{nl,ij}(\theta)u_{nl,j}$ in Lemma C.1(iv). Under the assumption that $\sup_n \sup_{1\le i\le n} \mathrm{E}\,|v_{ni}^4| < \infty$, by Lemma C.1($iv$), $\Delta_{n1,i}$'s are uniformly integrable. Thus, by the martingale law of large numbers in Davidson (1994, p. 299, Theorem 19.7), $\frac{1}{n}\sum_{i=1}^{n}\Delta_{n1,i} = o_p(1)$. This argument still holds for the original $r_{ni}^{(1)}$ and $r_{ni}^{(2)}$ with the assumption $\sup_n \sup_{1\le i\le n} \mathrm{E}\,|v_{ni}^4| < \infty$. For $\Delta_{n2,i}$, because each term in its expression has mean zero and is uncorrelated with the corresponding one in $\Delta_{n2,s}$ for $s \ne i$, the sample average over i of each term has a variance of order $O(n^{-1})$ under the assumption $\sup_n \sup_{1\le i\le n} \mathrm{E}(v_{ni}^4) < \infty$. For example,

$$\begin{aligned}
&\mathrm{var}\Big(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{j-1}\sum_{l=1}^{k-1}\sum_{m=1}^{l-1} v_{nj}v_{nk}v_{nl}v_{nm}g_{n1,ij}h_{n1,ik}g_{n2,il}h_{n2,im}\Big)\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{i-1} g_{n1,ij}^2\sigma_{nj}^2 \sum_{k=1}^{j-1} h_{n1,ik}^2\sigma_{nk}^2 \sum_{l=1}^{k-1} g_{n2,il}^2\sigma_{nl}^2 \sum_{m=1}^{l-1} h_{n2,im}^2\sigma_{nm}^2\\
&= O(n^{-1}).
\end{aligned}$$

Thus, $\frac{1}{n}\sum_{i=1}^n \Delta_{n2,i} = O_p(n^{-1/2})$. With the original $r_{ni}^{(1)}$ and $r_{ni}^{(2)}$, the argument still applies with the assumption $\sup_n \sup_{1\le i\le n} \mathrm{E}(v_{ni}^4) < \infty$. Hence, $\frac{1}{n}\sum_{i=1}^n [r_{ni}^{(1)}r_{ni}^{(2)} - \mathrm{E}(r_{ni}^{(1)}r_{ni}^{(2)})] = o_p(1)$.

(*ii*) We decompose $r_{ni}^{(1)}r_{ni}^{(2)} - \mathrm{E}(r_{ni}^{(1)}r_{ni}^{(2)}) = \Delta_{n1,i} + \Delta_{n2,i}$ in a way similar to that in (*i*). With the assumption $\sup_n \sup_{1\le i\le n} \mathrm{E}(v_{ni}^8) < \infty$, instead of using a martingale CLT for $\frac{1}{n}\sum_{i=1}^n \Delta_{n1,i}$, we may prove that $\frac{1}{n}\sum_{i=1}^n \Delta_{n1,i} = O_p(n^{-1/2})$ by showing that the sample average of each term in its expression has a variance of order $O(n^{-1})$. Thus, $\frac{1}{n}\sum_{i=1}^n \Delta_{n1,i} = O_p(n^{-1/2})$. Similar to (*i*), $\frac{1}{n}\sum_{i=1}^n \Delta_{n2,i} = O_p(n^{-1/2})$. Hence the result holds. □

C.2 Lemmas related to the GMM

All lemmas below accommodate both the homoskedastic and heteroskedastic cases for the SARAR 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. For a square matrix A, let $\mathrm{tril}(A)$ be the strictly lower triangle matrix formed by the elements below the diagonal of A.

Lemma C.3. *Suppose that $A_{n1}(\theta) = [a_{n1,ij}(\theta)]$ and $A_{n2}(\theta) = [a_{n2,ij}(\theta)]$ are nonstochastic symmetric square matrices of dimension n, $b_{n1}(\theta) = [b_{n1,i}(\theta)]$ and $b_{n2}(\theta) = [b_{n2,i}(\theta)]$ are nonstochastic column vectors of dimension n, and their elements are functions of $\theta \in \Theta$. Assume that each element of $A_{n1}(\theta)$, $A_{n2}(\theta)$, $b_{n1}(\theta)$ and $b_{n2}(\theta)$ is differentiable with respect to θ, the sequences $A_{n1}(\theta)$, $A_{n2}(\theta)$, $\frac{\partial A_{n1}(\theta)}{\partial \theta_j}$ and $\frac{\partial A_{n2}(\theta)}{\partial \theta_j}$ for $j = 1, \ldots, k_\theta$ are bounded in both row and column sum norms, and $b_{n1}(\theta)$, $b_{n2}(\theta)$, $\frac{\partial b_{n1}(\theta)}{\partial \theta_j}$ and $\frac{\partial b_{n2}(\theta)}{\partial \theta_j}$ for $j = 1, \ldots, k_\theta$ are bounded in row sum norm, uniformly on Θ.*

Let $\xi_{nr,i}(\theta) = a_{nr,ii}(\theta)[v_{ni}^2(\theta) - \sigma^2] + 2v_{ni}(\theta)\sum_{j=1}^{i-1} a_{nr,ij}(\theta)v_{nj}(\theta) + b_{nr,i}(\theta)v_{ni}(\theta)$ for $r = 1, 2$ if the disturbances v_{ni}'s are homoskedastic, and $\xi_{nr,i}(\theta) = 2v_{ni}(\theta)\sum_{j=1}^{i-1} a_{nr,ij}(\theta)v_{nj}(\theta) + b_{nr,i}(\theta)v_{ni}(\theta)$ for $r = 1, 2$ if the disturbances v_{ni}'s are heteroskedastic. Assume that $\hat\theta_n = \theta_0 + o_p(1)$. Then, either under Assumptions 1(i) and 2–4 in the homoskedastic case, or under Assumptions 1(ii) and 2–4 with $\mathrm{diag}(A_{n1}(\theta)) = \mathrm{diag}(A_{n2}(\theta)) = 0$ in the heteroskedastic case,

$$\frac{1}{n}\sum_{i=1}^n \xi_{n1,i}(\hat\theta_n)\xi_{n2,i}(\hat\theta_n) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}[\xi_{n1,i}(\theta_0)\xi_{n2,i}(\theta_0)] + o_p(1).$$

Proof. By the mean value theorem,

$$\frac{1}{n}\sum_{i=1}^n \xi_{n1,i}(\hat\theta_n)\xi_{n2,i}(\hat\theta_n) = \frac{1}{n}\sum_{i=1}^n \xi_{n1,i}(\theta_0)\xi_{n2,i}(\theta_0) + \sum_{l=1}^{k_\theta}\frac{1}{n}\sum_{i=1}^n \Big[\frac{\partial \xi_{n1,i}(\check\theta_n)}{\partial \theta_l}\xi_{n2,i}(\check\theta_n) + \xi_{n1,i}(\check\theta_n)\frac{\partial \xi_{n2,i}(\check\theta_n)}{\partial \theta_l}\Big](\hat\theta_{nl} - \theta_{0l}),$$

where $\check{\theta}_n$ lies between θ_0 and $\hat{\theta}_n$. We shall show that the second term on the r.h.s. of the above equation goes to zero in probability. Note that $\xi_{nr,i}(\theta) = a_{nr,ii}(\theta)[(e'_{ni}V_n(\theta))^2 - \sigma^2] + 2e'_{ni}V_n(\theta)e'_{ni}\text{tril}[A_n(\theta)]V_n(\theta) + b_{nr,i}(\theta)e'_{ni}V_n(\theta)$, where e_{ni} is the ith unit column vector of dimension n, and $\text{tril}[A_n(\theta)]$ and $\frac{\partial \text{tril}[A_n(\theta)]}{\partial\theta_l}$ for $l = 1, \ldots, p$ are bounded in both row and column sum norms uniformly on Θ. Since $Y_n = S_n^{-1}X_n\beta_0 + S_n^{-1}R_n^{-1}V_n$,

$$\begin{aligned} V_n(\theta) &= [R_n + (\tau_0 - \tau)M_n]\{[S_n + (\kappa_0 - \kappa)W_n](S_n^{-1}X_n\beta_0 + S_n^{-1}R_n^{-1}V_n) - X_n\beta\} \\ &= R_nX_n(\beta_0 - \beta) + (\kappa_0 - \kappa)R_nW_nS_n^{-1}X_n\beta_0 + M_nX_n(\beta_0 - \beta)(\tau_0 - \tau) + (\kappa_0 - \kappa)(\tau_0 - \tau)M_nW_nS_n^{-1}X_n\beta_0 \\ &\quad + (\kappa_0 - \kappa)R_nW_nS_n^{-1}R_n^{-1}V_n + (\tau_0 - \tau)M_nR_n^{-1}V_n + (\tau_0 - \tau)(\kappa_0 - \kappa)M_nW_nS_n^{-1}R_n^{-1}V_n + V_n, \end{aligned} \tag{13}$$

which is linear in V_n and quadratic in θ. In (13), terms that do not involve V_n have uniformly bounded elements, and terms that involve V_n have matrices in front of V_n bounded in both row and column sum norms. We can expand $\frac{1}{n}\sum_{i=1}^n\big[\frac{\partial\xi_{n1,i}(\theta)}{\partial\theta_l}\xi_{n2,i}(\theta) + \xi_{n1,i}(\theta)\frac{\partial\xi_{n2,i}(\theta)}{\partial\theta_l}\big]$ by using (13) such that it is a sum of terms that have the forms in Lemma C.1 with $u_{nl,i} = v_{ni}$. Then $\frac{1}{n}\sum_{i=1}^n\big[\frac{\partial\xi_{n1,i}(\theta)}{\partial\theta_l}\xi_{n2,i}(\theta) + \xi_{n1,i}(\theta)\frac{\partial\xi_{n2,i}(\theta)}{\partial\theta_l}\big] = O_p(1)$ uniformly in a neighborhood of θ_0. Since $\hat{\theta}_n = \theta_0 + o_p(1)$, $\sum_{l=1}^{k_\theta}\frac{1}{n}\sum_{i=1}^n\big[\frac{\partial\xi_{n1,i}(\check{\theta}_n)}{\partial\theta_l}\xi_{n2,i}(\check{\theta}_n) + \xi_{n1,i}(\check{\theta}_n)\frac{\partial\xi_{n2,i}(\check{\theta}_n)}{\partial\theta_l}\big] = O_p(1)$. Thus,

$$\frac{1}{n}\sum_{i=1}^n \xi_{n1,i}(\hat{\theta}_n)\xi_{n2,i}(\hat{\theta}_n) = \frac{1}{n}\sum_{i=1}^n \xi_{n1,i}(\theta_0)\xi_{n2,i}(\theta_0) + o_p(1).$$

By Lemma C.2(i), $\frac{1}{n}\sum_{i=1}^n \xi_{n1,i}(\theta_0)\xi_{n2,i}(\theta_0) = \frac{1}{n}\sum_{i=1}^n \mathrm{E}[\xi_{n1,i}(\theta_0)\xi_{n2,i}(\theta_0)] + o_p(1)$. Hence, the result in the lemma follows. □

The $z_{nl,i}(\theta)$, $g_{ni}(\theta)$ and $g_n(\theta)$ have different expressions depending on variances of disturbances. In the homoskedastic case, $g_n(\theta)$, $z_{nl,i}(\theta)$ and $g_{ni}(\theta)$ are given in, respectively, (2), (3) and (4); in the heteroskedastic case, $g_n(\theta)$, $z_{nl,i}(\theta)$ and $g_{ni}(\theta)$ are given in, respectively, (2), (3) and (4) with $p_{nl,ii} = 0$ for $i = 1, \ldots, n$ and $l = 1, \ldots, k_p$.

Lemma C.4. *Under Assumptions 1–3,* $\sup_{\theta\in\Theta}\sup_{1\le i\le n}\|g_{ni}(\theta)\| = O_p(n^{2/(4+\iota)})$.

Proof. The expression for $g_{ni}(\theta)$ is given in (4), in which $\omega_{nl,i}(\theta) = p_{nl,ii}[(e'_{ni}V_n(\theta))^2 - \sigma^2] + 2e'_{ni}V_n(\theta)e'_{ni}\text{tril}(P_{nl})V_n(\theta)$ in the homoskedastic case, and $\omega_{nl,i}(\theta) = 2e'_{ni}V_n(\theta)e'_{ni}\text{tril}(P_{nl})V_n(\theta)$, where $V_n(\theta)$ is linear in V_n and quadratic in θ in (13), and $\text{tril}(P_{nl})$ is bounded in both row and column sum norms. Using the expression of $V_n(\theta)$ in (13), each element of $g_{ni}(\theta)$ can be expanded as a polynomial of θ whose coefficients have the form $\Pi_{l=1}^s\sum_{j=1}^n d_{nl,ij}(\theta)u_{nl,j}$ with $u_{nl,j} = v_{nj}$ and $s = 1$ or 2 in Lemma C.1(iii). Thus, the result follows by Lemma C.1(iii). □

Let $G_n^{(i)}(\theta) = \frac{\partial G_n(\theta)}{\partial\theta_i}$, $G_n^{(ij)}(\theta) = \frac{\partial^2 G_n(\theta)}{\partial\theta_i\partial\theta_j}$, $G_n^{(ijk)}(\theta) = \frac{\partial^3 G_n(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}$ 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 C.5. *Under Assumptions 1–3,* $\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^n\|g_{ni}(\theta)\|^2$, $\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^n\|G_{ni}(\theta)\|$, $\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^n\|G_{ni}^{(j)}(\theta)\|$, $\sup_{\theta\in\Theta}\|g_n(\theta)\|$, $\sup_{\theta\in\Theta}\|G_n(\theta)\|$, $\sup_{1\le i\le p}\sup_{\theta\in\Theta}\|G_n^{(i)}(\theta)\|$, $\sup_{1\le i,j\le p}\sup_{\theta\in\Theta}\|G_n^{(ij)}(\theta)\|$, *and* $\sup_{1\le i,j,k\le p}\sup_{\theta\in\Theta}\|G_n^{(ijk)}(\theta)\|$ *are all of order* $O_p(1)$.

Proof. By (13) and the proof of Lemma C.4, we can expand $\frac{1}{n}\sum_{i=1}^n|\omega_{nl,i}(\theta)|^2$ and $\frac{1}{n}\sum_{i=1}^n|v_{ni}(\theta)|^2$ as polynomials of θ. Since $g_n(\theta)$ is quadratic in $V_n(\theta)$, each element of $g_n(\theta)$ can be expanded as a polynomial of θ. Each coefficient of those polynomials is $O_p(1)$ by Lemma C.1. Hence the results hold. □

Let $g_{ni}^{(k)}(\theta) = \frac{\partial g_{ni}(\theta)}{\partial \theta_k}$, $g_{ni}^{(kl)}(\theta) = \frac{\partial^2 g_{ni}(\theta)}{\partial \theta_k \partial \theta_l}$, $g_{ni}^{(klr)}(\theta) = \frac{\partial^3 g_{ni}(\theta)}{\partial \theta_k \partial \theta_l \partial \theta_r}$, and $g_{ni}^{(klrs)}(\theta) = \frac{\partial^4 g_{ni}(\theta)}{\partial \theta_k \partial \theta_l \partial \theta_r \partial \theta_s}$.

Lemma C.6. *Under Assumptions 1–3,* $\sup_{\theta\in\Theta} \|\frac{1}{n}\sum_{i=1}^n g_{ni}(\theta)g_{ni}'(\theta)\|$, $\sup_{\theta\in\Theta} \|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(k)}(\theta)g_{ni}'(\theta)\|$, $\sup_{\theta\in\Theta} \|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(kl)}(\theta)g_{ni}'(\theta)\|$, $\sup_{\theta\in\Theta} \|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(k)}(\theta)g_{ni}^{(l)'}(\theta)\|$, $\sup_{\theta\in\Theta} \|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(klr)}(\theta)g_{ni}'(\theta)\|$, $\sup_{\theta\in\Theta} \|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(kl)}(\theta)g_{ni}^{(r)'}(\theta)\|$, $\sup_{\theta\in\Theta} \|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(klrs)}(\theta)g_{ni}'(\theta)\|$, $\sup_{\theta\in\Theta} \|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(klr)}(\theta)g_{ni}^{(s)'}(\theta)\|$ *and* $\sup_{\theta\in\Theta} \|\frac{1}{n}\sum_{i=1}^n g_{ni}^{(kl)}(\theta)g_{ni}^{(rs)'}(\theta)\|$ *have order* $O_p(1)$.

Proof. As in the proof of Lemma C.5, $\frac{1}{n}\sum_{i=1}^n g_{ni}(\theta)g_{ni}'(\theta)$ can be expanded as a polynomial of θ with coefficients being $O_p(1)$ by Lemma C.1. Then the results in the lemma follow. □

Lemma C.7. *Under Assumptions 1–3, (i)* $\frac{1}{n}\sum_{i=1}^n g_{ni}(\theta_0)g_{ni}'(\theta_0) = \bar{\Omega}_n + o_p(1)$*, (ii)* $\frac{1}{n}\sum_{i=1}^n [\mathrm{E}\, g_{ni}^{(k)}(\theta_0)]g_{ni}'(\theta_0) = O_p(n^{-1/2})$*, and (iii)* $\frac{1}{n}\sum_{i=1}^n \mathrm{E}[g_{ni}^{(k)}(\theta_0)g_{ni}'(\theta_0)] = O(1)$*; under Assumptions 1–3 and 9, (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)* $\frac{1}{n}\sum_{i=1}^n \{g_{ni}^{(k)}(\theta_0)g_{ni}'(\theta_0) - \mathrm{E}[g_{ni}^{(k)}(\theta_0)g_{ni}'(\theta_0)]\} = O_p(n^{-1/2})$*, for* $k = 1, \dots, k_\theta$*.*

Proof. Since P_{ni}'s are symmetric, each element of $g_{ni}(\theta_0)$ has the linear-quadratic form $a_{n,ii}(v_{ni}^2 - \sigma_0^2) + 2v_{ni}\sum_{j=1}^{i-1} a_{n,ij}v_{nj} + b_{ni}v_{ni}$ or $2v_{ni}\sum_{j=1}^{i-1} a_{n,ij}v_{nj} + b_{ni}v_{ni}$, where $a_{n,ij}$ is the (i,j)th element of a nonstochastic matrix with bounded row and column sum norms, and b_{ni} for $i = 1, \dots, n$ are bounded uniformly in i. Then (i) and (iv) follow respectively from Lemma C.2(i) and Lemma C.2(ii). It remains to show (ii), (iii) and (v).

The lth element of $g_{ni}^{(k)}(\theta)$ for $1 \le l \le k_p$ is $\frac{\partial \omega_{nl,i}(\theta)}{\partial \theta_k} = 2p_{nl,ii}v_{ni}(\theta)\frac{\partial v_{ni}(\theta)}{\partial \theta_k} - p_{nl,ii}\frac{\partial \sigma^2}{\partial \theta_k} + 2\sum_{j=1}^{i-1} p_{nl,ij}\big[v_{ni}(\theta)\frac{\partial v_{nj}(\theta)}{\partial \theta_k} + v_{nj}(\theta)\frac{\partial v_{ni}(\theta)}{\partial \theta_k}\big]$ in the homoskedastic case, and $\frac{\partial \omega_{nl,i}(\theta)}{\partial \theta_k} = 2\sum_{j=1}^{i-1} p_{nl,ij}\big[v_{ni}(\theta)\frac{\partial v_{nj}(\theta)}{\partial \theta_k} + v_{nj}(\theta)\frac{\partial v_{ni}(\theta)}{\partial \theta_k}\big]$ in the heteroskedastic case; and the last k_q elements are $Q_{ni}\frac{\partial v_{ni}(\theta)}{\partial \theta_k}$. By (13), $\frac{\partial v_{ni}(\theta_0)}{\partial \theta_k}$ has the form $a_{nk,i} + \sum_{r=1}^n b_{nk,ir}v_{nr}$, where $a_{nk,i}$ is bounded uniformly in k and i, and $b_{nk,ir}$ is the (i,r)th element of an $n\times n$ matrix bounded in both row and column sum norms. Hence, every element of $g_{ni}^{(k)}(\theta_0)$ has the form

$$\Xi_{nk,i} = 2p_{nl,ii}v_{ni}\big(a_{nk,i} + \sum_{s=1}^n b_{nk,is}v_{ns}\big) - p_{nl,ii}\frac{\partial \sigma^2}{\partial \theta_k} + 2\sum_{j=1}^{i-1} p_{nl,ij}\big[v_{ni}\big(a_{nk,j} + \sum_{s=1}^n b_{nk,js}v_{ns}\big) + v_{nj}\big(a_{nk,i} + \sum_{s=1}^n b_{nk,is}v_{ns}\big)\big].$$

Thus, $\mathrm{E}(\Xi_{nk,i}) = 2p_{nl,ii}b_{nk,ii}\sigma_{ni}^2 - p_{nl,ii}\frac{\partial \sigma^2}{\partial \theta_k} + 2\sum_{j=1}^{i-1} p_{nl,ij}(b_{nk,ji}\sigma_{ni}^2 + b_{nk,ij}\sigma_{nj}^2)$ is bounded uniformly in i and k. Note that the variance of the term in (ii) only involves the first fourth moments of v_{nj}, then the result (ii) follows by Lemma C.2(ii).

Below we shall prove (iii) in the lemma and (vi) $\frac{1}{n}\sum_{i=1}^n \{(g_{ni}^{(k)}(\theta_0) - \mathrm{E}[g_{ni}^{(k)}(\theta_0)])g_{ni}'(\theta_0) - \mathrm{E}[g_{ni}^{(k)}(\theta_0)g_{ni}'(\theta_0)]\} = O_p(n^{-1/2})$. The result (v) in the lemma follows by (iii) and (vi). Write $\Xi_{nk,i} - \mathrm{E}(\Xi_{nk,i}) = \Xi_{1nk,i} + \Xi_{2nk,i}$, where

$$\begin{aligned}\Xi_{1nk,i} = {}& 2p_{nl,ii}a_{nk,i}v_{ni} + 2p_{nl,ii}v_{ni}\sum_{s=1}^{i-1} b_{nk,is}v_{ns} + 2p_{nl,ii}b_{nk,ii}(v_{ni}^2 - \sigma_{ni}^2) + 2v_{ni}\sum_{j=1}^{i-1} p_{nl,ij}a_{nk,j} \\ & + 2v_{ni}\sum_{j=1}^{i-1} p_{nl,ij}\sum_{s=1}^{i-1} b_{nk,js}v_{ns} + 2(v_{ni}^2 - \sigma_{ni}^2)\sum_{j=1}^{i-1} p_{nl,ij}b_{nk,ji} + 2a_{nk,i}\sum_{j=1}^{i-1} p_{nl,ij}v_{nj} + 2\sum_{j=1}^{i-1} p_{nl,ij}v_{nj}\sum_{s=1}^{j-1} b_{nk,is}v_{ns} \\ & + 2\sum_{j=1}^{i-1} p_{nl,ij}b_{nk,ij}(v_{nj}^2 - \sigma_{nj}^2) + 2\sum_{j=1}^{i-1} p_{nl,ij}v_{nj}\sum_{s=j+1}^{i-1} b_{nk,is}v_{ns} + 2b_{nk,ii}v_{ni}\sum_{j=1}^{i-1} p_{nl,ij}v_{nj},\end{aligned}$$

and

$$\Xi_{2nk,i} = 2p_{nl,ii}v_{ni}\Big(\sum_{s=i+1}^n b_{nk,is}v_{ns}\Big) + 2v_{ni}\Big(\sum_{j=1}^{i-1} p_{nl,ij}\sum_{s=i+1}^n b_{nk,js}v_{ns}\Big) + 2\sum_{j=1}^{i-1} p_{nl,ij}v_{nj}\sum_{s=i+1}^n b_{nk,is}v_{ns}$$

$$= 2p_{nl,ii}v_{ni}\Big(\sum_{s=i+1}^{n} b_{nk,is}v_{ns}\Big) + 2v_{ni}\Big(\sum_{s=i+1}^{n} v_{ns}\sum_{j=1}^{i-1} p_{nl,ij}b_{nk,js}\Big) + 2\sum_{s=i+1}^{n} b_{nk,is}v_{ns}\sum_{j=1}^{i-1} p_{nl,ij}v_{nj}.$$

Note that $\Xi_{1nk,i}$ has the form of $r_{ni}^{(1)}$ in Lemma C.2(ii) because

$$2\sum_{j=1}^{i-1} p_{nl,ij}v_{nj}\sum_{s=j+1}^{i-1} b_{nk,is}v_{ns} = 2\sum_{s=1}^{i-1} b_{nk,is}v_{ns}\sum_{j=1}^{s-1} p_{nl,ij}v_{nj}.$$

Thus $\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\Xi_{1nk,i}g_{ni}'(\theta_0)] = O(1)$ and $\frac{1}{n}\sum_{i=1}^{n}\Xi_{1nk,i}g_{ni}'(\theta_0) - \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\Xi_{1nk,i}g_{ni}'(\theta_0)] = O_p(n^{-1/2})$. For $\Xi_{2nk,i}$, we shall show that $\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\Xi_{2nk,i}g_{ni}'(\theta_0)] = 0$ and $\frac{1}{n}\sum_{i=1}^{n}\Xi_{2nk,i}g_{ni}'(\theta_0) - \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[\Xi_{2nk,i}g_{ni}'(\theta_0)] = O_p(n^{-1/2})$. Each term in $\Xi_{2nk,i}$ has the form $(\sum_{s=i+1}^{n} b_{nk,is}^{*}v_{ns})(\sum_{j=1}^{i} p_{nl,ij}^{*}v_{nj})$, where $b_{nk,is}^{*}$ is the (i,s)th element of a general $n\times n$ matrix with bounded row and column sum norms uniformly in k. The term $\sum_{j=1}^{i} p_{nl,ij}^{*}v_{nj}$ is a special form of $r_{ni}^{(1)}$ in Lemma C.2(ii). Compared with the form of $r_{ni}^{(1)}r_{ni}^{(2)}$ in Lemma C.2(ii), each element of $\Xi_{2nk,i}g_{ni}'(\theta_0)$ has the additional term $\sum_{s=i+1}^{n} b_{nk,is}^{*}v_{ns}$. In the proof of Lemma C.2(ii), If we multiply each term in $r_{ni}^{(1)}r_{ni}^{(2)} - \mathrm{E}(r_{ni}^{(1)}r_{ni}^{(2)})$ by $\sum_{s=i+1}^{n} b_{nk,is}^{*}v_{ns}$, then the obtained terms have zero expected values and the sample average of those terms over i still has the order $O_p(n^{-1/2})$, because the summation $\sum_{s=i+1}^{n} b_{nk,is}^{*}v_{ns}$ starts from $s = i+1$. Hence, $\mathrm{E}(\Xi_{2nk,i}r_{ni}^{(2)}) = 0$ and $\frac{1}{n}\sum_{i=1}^{n}[\Xi_{2nk,i}g_{ni}'(\theta_0) - \mathrm{E}(\Xi_{2nk,i}g_{ni}'(\theta_0))] = O_p(n^{-1/2})$. Then ($iii$) and ($vi$) follow. □

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

$$0 = -\begin{pmatrix} G_n'(\tilde\theta_n)\tilde\lambda_n \\ g_n(\tilde\theta_n) + \hat J_n\tilde\lambda_n \end{pmatrix}, \qquad (14)$$

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 in the following Lemma, $\bar G_n$ is the expected value of G_n, $\bar J_n$ is in Assumption 7, and $e_{k_g,j}$ is the jth column of the $k_g\times k_g$ identity matrix.

Lemma C.8. *Under Assumptions 1(i) (or (ii)), 2–4, 7, 8 and 11 (or 12),* $\sqrt{n}(\tilde\gamma_n - \gamma_0) = \tilde\xi_n + n^{-1/2}\tilde\psi_n + O_p(n^{-1})$, *where* $\tilde\xi_n = -(\bar K_n^J)^{-1}\binom{0}{\sqrt{n}g_n(\theta_0)} = O_p(1)$ *and* $\tilde\psi_n = -(\bar K_n^J)^{-1}\begin{pmatrix} 0 & \sqrt{n}(G_n' - \bar G_n') \\ \sqrt{n}(G_n - \bar G_n) & \xi_n^J \end{pmatrix}\tilde\xi_n - \frac{1}{2}(\bar K_n^J)^{-1}\sum_{i=1}^{k_\theta+k_g}\tilde\xi_{ni}\bar K_{ni}\tilde\xi_n = O_p(1)$, *where* $\bar K_n^J = \begin{pmatrix} 0 & \bar G_n' \\ \bar G_n & \bar J_n \end{pmatrix}$, $\bar K_{ni} = \begin{pmatrix} 0 & \bar G_n^{(i)'} \\ \bar G_n^{(i)} & 0 \end{pmatrix}$ *for* $1\le i\le k_\theta$, *and* $\bar K_{ni} = \begin{pmatrix} [\bar G_n^{(1)'}e_{k_g,i-k_\theta},\ldots,\bar G_n^{(k_\theta)'}e_{k_g,i-k_\theta}] & 0 \\ 0 & 0 \end{pmatrix}$ *for* $k_\theta+1\le i\le k_\theta+k_g$.

Proof. By Proposition A.1, $\tilde\theta_n = \theta_0 + O_p(n^{-1/2})$. Then by the mean value theorem and Lemma C.5, $g_n(\tilde\theta_n) = O_p(n^{-1/2})$. Thus $\tilde\lambda_n = -\hat J_n^{-1}g_n(\tilde\theta_n) = O_p(n^{-1/2})$. It follows that $\tilde\gamma_n - \gamma_0 = O_p(n^{-1/2})$. Together with Assumption 7, the first order condition (14) for the initial GMM is equal to

$$0 = -\begin{pmatrix} G_n'(\tilde\theta_n)\tilde\lambda_n \\ g_n(\tilde\theta_n) + (\bar J_n + n^{-1/2}\xi_n^J)\tilde\lambda_n \end{pmatrix} + O_p(n^{-3/2}).$$

By a second order Taylor expansion of the first vector on the right hand side at γ_0, and using Lemma C.5,

$$0 = -\begin{pmatrix} 0 \\ g_n(\theta_0) \end{pmatrix} - K_n^J(\tilde\gamma_n - \gamma_0) - \frac{1}{2}\sum_{i=1}^{k_\theta+k_g}(\tilde\gamma_{ni} - \gamma_{0i})K_{ni}(\tilde\gamma_n - \gamma_0) + O_p(n^{-3/2}),$$

where $K_n^J = \begin{pmatrix} 0 & G_n'(\theta_0) \\ G_n(\theta_0) & \bar{J}_n + n^{-1/2}\xi_n^J \end{pmatrix}$, $K_{ni} = \begin{pmatrix} 0 & G_n^{(i)'}(\theta_0) \\ G_n^{(i)}(\theta_0) & 0 \end{pmatrix}$ for $1 \leq i \leq k_\theta$, and

$K_{ni} = \begin{pmatrix} [G_n^{(1)'}(\theta_0)e_{k_g,i-k_\theta}, \ldots, G_n^{(k_\theta)'}(\theta_0)e_{k_g,i-k_\theta}] & 0 \\ 0 & 0 \end{pmatrix}$ for $k_\theta + 1 \leq i \leq k_\theta + k_g$. As $\bar{K}_n^J = \mathrm{E}(K_n^J) = O(1)$ and $\bar{K}_{ni} = \mathrm{E}(K_{ni}) = O(1)$ for all $i = 1, \ldots, k_\theta + k_g$,

$$\sqrt{n}(\tilde{\gamma}_n - \gamma_0) = -(\bar{K}_n^J)^{-1}\begin{pmatrix} 0 \\ \sqrt{n}g_n(\theta_0) \end{pmatrix} - (\bar{K}_n^J)^{-1}(K_n^J - \bar{K}_n^J)\sqrt{n}(\tilde{\gamma}_n - \gamma_0) - \frac{\sqrt{n}}{2}(\bar{K}_n^J)^{-1}\sum_{i=1}^{k_\theta+k_g}(\tilde{\gamma}_{ni} - \gamma_{0i})K_{ni}(\tilde{\gamma}_n - \gamma_0) + O_p(n^{-1}). \tag{15}$$

As every element of $g_n(\theta)$ is a linear-quadratic form of $V_n(\theta)$, which is linear in V_n by (13), so it is easily seen that $G_n(\theta_0) - \bar{G}_n = O_p(n^{-1/2})$ and $G_n^{(i)}(\theta_0) - \bar{G}_n^{(i)}(\theta_0) = O_p(n^{-1/2})$. It follows that $K_n^J - \bar{K}_n^J = O_p(n^{-1/2})$ and $K_{ni} - \bar{K}_{ni} = O_p(n^{-1/2})$. Hence, $\sqrt{n}(\tilde{\gamma}_n - \gamma_0) = \tilde{\xi}_n + O_p(n^{-1/2})$, where $\tilde{\xi}_n = -(\bar{K}_n^J)^{-1}\binom{0}{\sqrt{n}g_n(\theta_0)} = O_p(1)$. Substituting $K_{ni} - \bar{K}_{ni} = O_p(n^{-1/2})$ and $\sqrt{n}(\tilde{\gamma}_n - \gamma_0) = \tilde{\xi}_n + O_p(n^{-1/2})$ into (15) yields $\sqrt{n}(\tilde{\gamma}_n - \gamma_0) = \tilde{\xi}_n + \tilde{\psi}_n + O_p(n^{-1})$. □

Lemma C.9. *Under Assumptions 1(i) (or (ii)), 2–3, 7, 8 and 11 (or 12), $\Omega_n(\tilde{\theta}_n) = \bar{\Omega}_n + o_p(1)$; under the additional Assumption 9,*

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

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

Proof. By a first order Taylor expansion and Lemma C.6,

$$\Omega_n(\tilde{\theta}_n) = \bar{\Omega}_n + \Big(\frac{1}{n}\sum_{i=1}^n g_{ni}(\theta_0)g_{ni}'(\theta_0) - \bar{\Omega}_n\Big) + \sum_{k=1}^{k_\theta}\Big\{\frac{1}{n}\sum_{i=1}^n [g_{ni}(\theta_0)g_{ni}^{(k)'}(\theta_0) + g_{ni}^{(k)}(\theta_0)g_{ni}'(\theta_0)]\Big\}(\tilde{\theta}_{nk} - \theta_{0k}) + O_p(n^{-1}).$$

Under Assumptions 1(i) (or (ii)), 2–3, 7 and 11 (or 12), by Lemma C.7(i), the second term on the r.h.s. of the above equation is $o_p(1)$; by Lemma C.6(iii), the third term is $o_p(1)$. Thus the first result follows. The second result requires the existence of higher order moments of disturbances. Substituting the expression for $\tilde{\theta}_n - \theta_0$ in Lemma C.8 into the above equation and keeping only terms with order $O_p(n^{-1/2})$ by using Lemma C.7, we obtain the result. □

C.3 Lemmas related to the GEL

Lemma C.10. *Under Assumptions 1–3, for any ζ with $\zeta > \frac{2}{4+\iota}$ and $\Lambda_n = \{\lambda : \|\lambda\| \leq n^{-\zeta}\}$, $\sup_{\theta\in\Theta,\lambda\in\Lambda_n,1\leq i\leq n}|\lambda' g_{ni}(\theta)| \xrightarrow{p} 0$, and w.p.a.1., $\Lambda_n \subset \Lambda_n(\theta)$ for all $\theta \in \Theta$.*

Proof. Let $b_n = \sup_{1\leq i\leq n}\sup_{\theta\in\Theta}\|g_{ni}(\theta)\|$. By Lemma C.4, $b_n = O_p(n^{2/(4+\iota)})$. Then by the Cauchy-Schwarz inequality, $\sup_{\theta\in\Theta,\lambda\in\Lambda_n,1\leq i\leq n}|\lambda' g_{ni}(\theta)| \leq n^{-\zeta}b_n = O_p(n^{2/(4+\iota)-\zeta}) = o_p(1)$. Given the first conclusion, w.p.a.1. $\lambda' g_{ni}(\theta) \in \mathcal{V}$ for all $1 \leq i \leq n$, $\theta \in \Theta$ and $\|\lambda\| \leq n^{-\zeta}$. □

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

Lemma C.11. *Under Assumptions 1–3, 5 and 6, 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 $\sup_{\lambda\in\Lambda_n(\bar{\theta}_n)}\varrho_n(\bar{\theta}_n, \lambda) \leq \rho(0) + O_p(n^{-1})$.*

Proof. As in the proof of Lemma C.9, $\frac{1}{n}\sum_{i=1}^n g_{ni}(\bar\theta_n)g'_{ni}(\bar\theta_n) = \bar\Omega_n + o_p(1)$. Since $\lim_{n\to\infty}\bar\Omega_n$ is nonsingular, its smallest eigenvalue is bounded away from zero for large enough n. Let $\Lambda_n = \{\lambda : \|\lambda\| \le n^{-\zeta}\}$ for some $\frac{2}{4+\iota} < \zeta < \frac{1}{2}$. By Lemma C.10 and twice continuous differentiability of $\rho(v)$ in a neighborhood of zero, $\varrho_n(\bar\theta_n,\lambda)$ is continuously differentiable on Λ_n w.p.a.1., then $\tilde\lambda_n = \arg\max_{\lambda\in\Lambda_n}\varrho_n(\bar\theta_n,\lambda)$ exists w.p.a.1. Furthermore, for any $\dot\lambda_n$ between $\tilde\lambda_n$ and 0, by Lemma C.10 and $\rho_1 = -1, \max_{1\le i\le n}\rho_1(\dot\lambda'_n g_{ni}(\bar\theta_n)) < -\frac{1}{2}$, w.p.a.1. Then by a first order Taylor expansion of $\varrho_n(\bar\theta_n,\tilde\lambda_n)$ at $\lambda = 0$, there is $\dot\lambda_n$ between $\tilde\lambda_n$ and 0, such that

$$\begin{aligned}\rho(0) = \varrho_n(\bar\theta_n,0) \le \varrho_n(\bar\theta_n,\tilde\lambda_n) &= \rho(0) - \tilde\lambda'_n g_n(\bar\theta_n) + \frac{1}{2}\tilde\lambda'_n\Big[\frac{1}{n}\sum_{i=1}^n \rho_2(\dot\lambda'_n g_{ni}(\bar\theta_n))g_{ni}(\bar\theta_n)g'_{ni}(\bar\theta_n)\Big]\tilde\lambda_n \\ &\le \rho(0) - \tilde\lambda'_n g_n(\bar\theta_n) - \frac{1}{4}\tilde\lambda'_n\Big[\frac{1}{n}\sum_{i=1}^n g_{ni}(\bar\theta_n)g'_{ni}(\bar\theta_n)\Big]\tilde\lambda_n \\ &\le \rho(0) + \|\tilde\lambda_n\|\cdot\|g_n(\bar\theta_n)\| - c\|\tilde\lambda_n\|^2,\end{aligned} \tag{16}$$

for some positive constant c. Thus $c\|\tilde\lambda_n\| \le \|g_n(\bar\theta_n)\|$, w.p.a.1. Since $g_n(\bar\theta_n) = O_p(n^{-1/2})$, $\|\tilde\lambda_n\| = O_p(n^{-1/2}) = o_p(n^{-\zeta})$. Therefore, $\tilde\lambda_n \in \mathrm{int}(\Lambda_n)$ w.p.a.1, and $\frac{\partial\varrho_n(\bar\theta_n,\tilde\lambda_n)}{\partial\lambda} = 0$, the first order condition for an interior maximum. By Lemma C.10, $\tilde\lambda_n \in \Lambda_n(\bar\theta_n)$ w.p.a.1. Then by concavity of $\varrho_n(\bar\theta_n,\lambda)$ and convexity of $\Lambda_n(\bar\theta_n)$, it follows that $\varrho_n(\bar\theta_n,\tilde\lambda_n) = \sup_{\lambda\in\Lambda_n(\bar\theta_n)}\varrho_n(\bar\theta_n,\lambda)$, which gives the first and second conclusions with $\bar\lambda_n = \tilde\lambda_n$. Finally, by $\|g_n(\bar\theta_n)\| = O_p(n^{-1/2})$, $\|\tilde\lambda_n\| = O_p(n^{-1/2})$ and the last inequality of (16), we obtain

$$\varrho_n(\bar\theta_n,\bar\lambda_n) \le \rho(0) + \|\bar\lambda_n\|\cdot\|g_n(\bar\theta_n)\| - c\|\bar\lambda_n\|^2 = \rho(0) + O_p(n^{-1}).$$ □

Lemma C.12. *Under Assumptions 1–3, 5 and 6, $\|g_n(\hat\theta_{n,\mathrm{GEL}})\| = O_p(n^{-1/2})$, where $\hat\theta_{n,\mathrm{GEL}}$ is the GEL estimator.*

Proof. Consider any $\tilde\lambda_n \in \Lambda_n = \{\lambda : \|\lambda\| \le n^{-\zeta}\}$. By Lemma C.10, $\max_{1\le i\le n}|\dot\lambda'_n g_{ni}(\hat\theta_{n,\mathrm{GEL}})| \xrightarrow{p} 0$ for any $\dot\lambda_n$ between $\tilde\lambda_n$ and 0. Thus w.p.a.1. $\rho_2(\dot\lambda'_n g_{ni}(\hat\theta_{n,\mathrm{GEL}})) \ge -c$, for some positive constant c and for all $i = 1,\ldots,n$. Because all eigenvalues of $g_{ni}(\theta)g'_{ni}(\theta)$ are nonnegative and their sum is $\mathrm{tr}(g_{ni}(\theta)g'_{ni}(\theta)) = g'_{ni}(\theta)g_n(\theta)$, $g_{ni}(\theta)g'_{ni}(\theta) \le g'_{ni}(\theta)g_{ni}(\theta)I_n$ for $i = 1,\ldots,n$. Thus, $\frac{1}{n}\sum_{i=1}^n g_{ni}(\hat\theta_{n,\mathrm{GEL}})g'_{ni}(\hat\theta_{n,\mathrm{GEL}}) \le (\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^n\|g_{ni}(\theta)\|^2)I_n$, which is $O_p(1)$ by Lemma C.5. It follows that the largest eigenvalue of $\frac{1}{n}\sum_{i=1}^n g_{ni}(\hat\theta_{n,\mathrm{GEL}})g'_{ni}(\hat\theta_{n,\mathrm{GEL}})$ is bounded above w.p.a.1. A first order Taylor expansion then gives

$$\begin{aligned}\varrho_n(\hat\theta_{n,\mathrm{GEL}},\tilde\lambda_n) &= \rho(0) - \tilde\lambda'_n g_n(\hat\theta_{n,\mathrm{GEL}}) + \frac{1}{2}\tilde\lambda'_n\big[\frac{1}{n}\sum_{i=1}^n \rho_2(\dot\lambda'_n g_{ni}(\hat\theta_{n,\mathrm{GEL}}))g_{ni}(\hat\theta_{n,\mathrm{GEL}})g'_{ni}(\hat\theta_{n,\mathrm{GEL}})\big]\tilde\lambda_n \\ &\ge \rho(0) - \tilde\lambda'_n g_n(\hat\theta_{n,\mathrm{GEL}}) - \frac{c}{2}\tilde\lambda'_n\big[\frac{1}{n}\sum_{i=1}^n g_{ni}(\hat\theta_{n,\mathrm{GEL}})g'_{ni}(\hat\theta_{n,\mathrm{GEL}})\big]\tilde\lambda_n \\ &\ge \rho(0) - \tilde\lambda'_n g_n(\hat\theta_{n,\mathrm{GEL}}) - c_1\tilde\lambda'_n\tilde\lambda_n,\end{aligned}$$

w.p.a.1. for some $c_1 > 0$. Note that Lemma C.11 holds for $\bar\theta_n = \theta_0$, then since $(\hat\theta_{n,\mathrm{GEL}},\hat\lambda_{n,\mathrm{GEL}})$ is a saddle point,

$$\rho(0) - \tilde\lambda'_n g_n(\hat\theta_{n,\mathrm{GEL}}) - c_1\tilde\lambda'_n\tilde\lambda_n \le \varrho_n(\hat\theta_{n,\mathrm{GEL}},\tilde\lambda_n) \le \varrho_n(\hat\theta_{n,\mathrm{GEL}},\hat\lambda_{n,\mathrm{GEL}}) \le \sup_{\lambda\in\Lambda_n(\theta_0)}\varrho_n(\theta_0,\lambda) \le \rho(0) + O_p(n^{-1}). \tag{17}$$

In (17), we may let $\tilde\lambda_n = -n^{-\zeta}g_n(\hat\theta_{n,\mathrm{GEL}})\|g_n(\hat\theta_{n,\mathrm{GEL}})\|^{-1}$, as the Euclidean norm of this $\tilde\lambda_n$ is $n^{-\zeta}$. Then (17) implies that $\|g_n(\hat\theta_{n,\mathrm{GEL}})\| \le c_1 n^{-\zeta} + O_p(n^{\zeta-1}) = O_p(n^{-\zeta})$, as $\zeta < 1/2$. With this order of $g_n(\hat\theta_{n,\mathrm{GEL}})$, we may let

$\tilde{\lambda}_n = -U_n g_n(\hat{\theta}_{n,\mathrm{GEL}})$ in (17), where $U_n \to 0$, since this $\tilde{\lambda}_n \in \Lambda_n$ w.p.a.1. Then w.p.a.1., $\rho(0) - \tilde{\lambda}_n' g_n(\hat{\theta}_{n,\mathrm{GEL}}) - c_1 \tilde{\lambda}_n' \tilde{\lambda}_n = \rho(0) + U_n \|g_n(\hat{\theta}_{n,\mathrm{GEL}})\|^2 - c_1 U_n^2 \|g_n(\hat{\theta}_{n,\mathrm{GEL}})\|^2 \leq \rho(0) + O_p(n^{-1})$. Since $1 - c_1 U_n$ is bounded away from zero for large enough n, $U_n \|g_n(\hat{\theta}_{n,\mathrm{GEL}})\|^2 = O_p(n^{-1})$. The conclusion then follows from the standard result that if $U_n J_n = O_p(n^{-1})$ for all $U_n \to 0$, then $J_n = O_p(n^{-1})$. Hence, $\|g_n(\hat{\theta}_{n,\mathrm{GEL}})\| = O_p(n^{-1/2})$. □

Appendix D Proofs

Proof of Proposition 3.1. By (13), each element of $g_n(\theta)$ can be expanded as a linear-quadratic form of V_n and is a polynomial of θ. Thus $\sup_{\theta \in \Theta} \|g_n(\theta) - \mathrm{E}[g_n(\theta)]\| \xrightarrow{p} 0$. By Lemma C.12, $g_n(\hat{\theta}_{n,\mathrm{GEL}}) = O_p(n^{-1/2})$. Let $\bar{g}_n(\theta) = \mathrm{E}[g_n(\theta)]$, then $\|\bar{g}_n(\hat{\theta}_{n,\mathrm{GEL}})\| = \|\bar{g}_n(\hat{\theta}_{n,\mathrm{GEL}}) - g_n(\hat{\theta}_{n,\mathrm{GEL}}) + g_n(\hat{\theta}_{n,\mathrm{GEL}})\| \leq \|\bar{g}_n(\hat{\theta}_{n,\mathrm{GEL}}) - g_n(\hat{\theta}_{n,\mathrm{GEL}})\| + \|g_n(\hat{\theta}_{n,\mathrm{GEL}})\| = o_p(1)$. Since $\lim_{n\to\infty} \bar{g}_n(\theta)$ is uniquely zero at θ_0 under Assumption 11 or Assumption 12 as discussed in Appendix A, $\|\bar{g}_n(\theta)\|$ must be bounded away from zero outside of any neighborhood of θ_0. Therefore $\hat{\theta}_{n,\mathrm{GEL}}$ must be inside any neighborhood of θ_0 w.p.a.1, i.e., $\hat{\theta}_{n,\mathrm{GEL}} \xrightarrow{p} \theta_0$. As $g_n(\hat{\theta}_{n,\mathrm{GEL}}) = O_p(n^{-1/2})$, Lemma C.11 holds for $\bar{\theta}_n = \hat{\theta}_{n,\mathrm{GEL}}$. Hence, $\hat{\lambda}_{n,\mathrm{GEL}} = \arg\max_{\lambda \in \Lambda_n(\hat{\theta}_{n,\mathrm{GEL}})} \frac{1}{n} \sum_{i=1}^n \rho(\lambda' g_{ni}(\hat{\theta}_{n,\mathrm{GEL}}))$ exists w.p.a.1, and $\hat{\lambda}_{n,\mathrm{GEL}} = O_p(n^{-1/2})$. □

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

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

is satisfied w.p.a.1. By the implicit function theorem, there is a neighborhood of $\hat{\theta}_{n,\mathrm{GEL}}$ 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

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

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

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

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 $\hat{\gamma}_{n,\mathrm{GEL}}$ and γ_0 elementwise. As $\max_{1\leq i\leq n} |\hat{\lambda}_{n,\mathrm{GEL}}' g_{ni}(\hat{\theta}_{n,\mathrm{GEL}})| \xrightarrow{p} 0$, by the twice continuous differentiability of $\rho(v)$, $\max_{1\leq i\leq n} |\rho_l(\bar{\lambda}_n' g_{ni}(\bar{\theta}_n)) + 1| = o_p(1)$ for $l = 1$ and 2. Then by Lemma C.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 C.6 and C.7,

$$\frac{1}{n}\sum_{i=1}^{n}\rho_2(\bar{\lambda}_n' g_{ni}(\bar{\theta}_n))g_{ni}(\bar{\theta}_n)g_{ni}'(\bar{\theta}_n)$$

$$= \frac{1}{n}\sum_{i=1}^{n}[\rho_2(\bar{\lambda}_n' g_{ni}(\bar{\theta}_n))+1]g_{ni}(\bar{\theta}_n)g_{ni}'(\bar{\theta}_n) - \frac{1}{n}\sum_{i=1}^{n} g_{ni}(\theta_0)g_{ni}'(\theta_0) - \frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{k_\theta}[g_{ni}^{(l)}(\check{\theta}_n)g_{ni}'(\check{\theta}_n) + g_{ni}(\check{\theta}_n)g_{ni}^{(l)'}(\check{\theta}_n)](\bar{\theta}_{nl}-\theta_{0l})$$

$$= -\bar{\Omega}_n + o_p(1),$$

$\frac{1}{n}\sum_{i=1}^{n}\rho_2(\bar{\lambda}_n' g_{ni}(\bar{\theta}_n))G_{ni}'(\bar{\theta}_n)\bar{\lambda}_n\bar{\lambda}_n' G_{ni}(\bar{\theta}_n) = -\frac{1}{n}\sum_{i=1}^{n} G_{ni}'(\bar{\theta}_n)\bar{\lambda}_n\bar{\lambda}_n' G_{ni}(\bar{\theta}_n) + o_p(1) = o_p(1)$,

$$\frac{1}{n}\sum_{i=1}^{n}\rho_1(\bar{\lambda}_n' g_{ni}(\bar{\theta}_n))[G_{ni}^{(1)'}(\bar{\theta}_n)\bar{\lambda}_n,\dots,G_{ni}^{(k_\theta)'}(\bar{\theta}_n)\bar{\lambda}_n] = -\frac{1}{n}\sum_{i=1}^{n}[G_{ni}^{(1)'}(\bar{\theta}_n)\bar{\lambda}_n,\dots,G_{ni}^{(k_\theta)'}(\bar{\theta}_n)\bar{\lambda}_n] + o_p(1) = o_p(1),$$

and $\frac{1}{n}\sum_{i=1}^{n}\rho_2(\bar{\lambda}_n' g_{ni}(\bar{\theta}_n))g_{ni}(\bar{\theta}_n)\bar{\lambda}_n' G_{ni}(\bar{\theta}_n) = -\frac{1}{n}\sum_{i=1}^{n} g_{ni}(\bar{\theta}_n)\bar{\lambda}_n' G_{ni}(\bar{\theta}_n) + o_p(1) = o_p(1)$. 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,

$$\sqrt{n}(\hat{\gamma}_{n,\mathrm{GEL}} - \gamma_0) = -\bar{K}_n^{-1}\begin{pmatrix} 0 \\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_{ni}(\theta_0)\end{pmatrix} + o_p(1).$$

Since $\bar{K}_n^{-1} = \begin{pmatrix} -\bar{\Sigma}_n & \bar{H}_n \\ \bar{H}_n' & \bar{D}_n \end{pmatrix}$,

$$\sqrt{n}(\hat{\gamma}_{n,\mathrm{GEL}} - \gamma_0) = -\begin{pmatrix} \bar{H}_n \\ \bar{D}_n \end{pmatrix}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_{ni}(\theta_0) + o_p(1). \tag{18}$$

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

Proof of Proposition 3.3. Since $\hat{\lambda}_{n,\mathrm{GMM}} = -\Omega_n^{-1}(\tilde{\theta}_n)g_n(\hat{\theta}_{n,\mathrm{GMM}}) = O_p(n^{-1/2})$, by Lemma C.9, the first order condition (9) can be written as

$$0 = -\begin{pmatrix} G_n'(\hat{\theta}_{n,\mathrm{GMM}})\hat{\lambda}_{n,\mathrm{GMM}} \\ g_n(\hat{\theta}_{n,\mathrm{GMM}}) + (\bar{\Omega}_n + n^{-1/2}\xi_n^{\Omega})\hat{\lambda}_{n,\mathrm{GMM}} \end{pmatrix} + O_p(n^{-3/2}). \tag{19}$$

By a second order Taylor expansion and Lemma C.5,

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

where $K_n^{\Omega}(\theta) = \begin{pmatrix} [G_n^{(1)'}(\theta)\lambda,\dots,G_n^{(k_\theta)'}(\theta)\lambda] & G_n'(\theta) \\ G_n(\theta) & \bar{\Omega}_n + n^{-1/2}\xi_n^{\Omega} \end{pmatrix}$, and K_{ni}'s are given in the proof of Lemma C.8. Let $\bar{K}_n = \begin{pmatrix} 0 & \bar{G}_n' \\ \bar{G}_n & \bar{\Omega}_n \end{pmatrix}$. Then,

$$\begin{aligned}\sqrt{n}(\hat{\gamma}_{n,\mathrm{GMM}} - \gamma_0) &= -\bar{K}_n^{-1}\begin{pmatrix} 0 \\ \sqrt{n}g_n(\theta_0)\end{pmatrix} - \bar{K}_n^{-1}[K_n^{\Omega}(\theta_0) - \bar{K}_n]\sqrt{n}(\hat{\gamma}_{n,\mathrm{GMM}} - \gamma_0) \\ &\quad - \frac{\sqrt{n}}{2}\sum_{i=1}^{k_\theta+k_g}\bar{K}_n^{-1}K_{ni}(\hat{\gamma}_{n,\mathrm{GMM}} - \gamma_0)(\hat{\gamma}_{ni,\mathrm{GMM}} - \gamma_{0i}) + O_p(n^{-1}).\end{aligned} \tag{20}$$

By (20), we have

$$\sqrt{n}(\hat{\gamma}_{n,\mathrm{GMM}} - \gamma_0) = \xi_n + O_p(n^{-1/2}), \tag{21}$$

where $\xi_n = -\bar{K}_n^{-1}\binom{0}{\sqrt{n}g_n(\theta_0)} = O_p(1)$. Substituting (21) into the the second and third terms of (20) yields $\sqrt{n}(\hat{\gamma}_{n,\mathrm{GMM}} - \gamma_0) = \xi_n + n^{-1/2}\psi_n + O_p(n^{-1})$, where

$$\psi_n = -\bar{K}_n^{-1}\begin{pmatrix} 0 & \sqrt{n}(G_n' - \bar{G}_n') \\ \sqrt{n}(G_n - \bar{G}_n) & \xi_n^{\Omega} \end{pmatrix}\xi_n - \frac{1}{2}\sum_{i=1}^{k_\theta+k_g}\bar{K}_n^{-1}\bar{K}_{ni}\xi_n\xi_{ni} = O_p(1). \tag{22}$$

□

Proof of Proposition 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:

$$\frac{1}{n}\sum_{i=1}^{n} m_{ni}(\hat{\gamma}_{n,\mathrm{GEL}}) = 0. \tag{23}$$

Let g_{nit} be the tth element of g_{ni}, and $g_{nit}^{(k)}$ be the tth element of $g_{ni}^{(k)}$. Then,

$$\frac{\partial h_{ni}(\gamma)}{\partial\gamma'} = \begin{pmatrix} [G_{ni}^{(1)'}(\theta)\lambda, \ldots, G_{ni}^{(k_\theta)'}(\theta)\lambda] & G_{ni}'(\theta) \\ G_{ni}(\theta) & 0 \end{pmatrix},$$

$$\frac{\partial^2 h_{ni}(\gamma)}{\partial\gamma_j\partial\gamma'} = \begin{pmatrix} [G_{ni}^{(1j)'}(\theta)\lambda, \ldots, G_{ni}^{(k_\theta,j)'}(\theta)\lambda] & G_{ni}^{(j)'}(\theta) \\ G_{ni}^{(j)}(\theta) & 0 \end{pmatrix} \text{ for } 1 \le j \le k_\theta,$$

$$\frac{\partial^2 h_{ni}(\gamma)}{\partial\gamma_j\partial\gamma'} = \begin{pmatrix} [G_{ni}^{(1)'}(\theta)e_{k_g,j-k_\theta}, \ldots, G_{ni}^{(k_\theta)'}(\theta)e_{k_g,j-k_\theta}] & 0 \\ 0 & 0 \end{pmatrix} \text{ for } k_\theta+1 \le j \le k_\theta+k_g,$$

$$\frac{\partial^3 h_{ni}(\gamma)}{\partial\gamma_k\partial\gamma_j\partial\gamma'} = \begin{pmatrix} [G_{ni}^{(1jk)'}(\theta)\lambda, \ldots, G_{ni}^{(k_\theta,jk)'}(\theta)\lambda] & G_{ni}^{(jk)'}(\theta) \\ G_{ni}^{(jk)}(\theta) & 0 \end{pmatrix} \text{ for } 1 \le j \le k_\theta, \text{ and } 1 \le k \le k_\theta,$$

$$\frac{\partial^3 h_{ni}(\gamma)}{\partial\gamma_k\partial\gamma_j\partial\gamma'} = \begin{pmatrix} [G_{ni}^{(1j)'}(\theta)e_{k_g,k-k_\theta}, \ldots, G_{ni}^{(k_\theta,j)'}(\theta)e_{k_g,k-k_\theta}] & 0 \\ 0 & 0 \end{pmatrix} \text{ for } 1 \le j \le k_\theta \text{ and } k_\theta+1 \le k \le k_\theta+k_g,$$

$$\frac{\partial^3 h_{ni}(\gamma)}{\partial\gamma_k\partial\gamma_j\partial\gamma'} = \begin{pmatrix} [G_{ni}^{(1k)'}(\theta)e_{k_g,j-k_\theta}, \ldots, G_{ni}^{(k_\theta,k)'}(\theta)e_{k_g,j-k_\theta}] & 0 \\ 0 & 0 \end{pmatrix} \text{ for } k_\theta+1 \le j \le k_\theta+k_g \text{ and } 1 \le k \le k_\theta,$$

and $\frac{\partial^3 h_{ni}(\gamma)}{\partial\gamma_k\partial\gamma_j\partial\gamma'} = 0$ for $k_\theta+1 \le j \le k_\theta+k_g$ and $k_\theta+1 \le k \le k_\theta+k_g$. Hence, by the chain rule of differentiation,

$$\frac{\partial m_{ni}(\gamma_0)}{\partial\gamma'} = -\begin{pmatrix} 0 & G_{ni}'(\theta_0) \\ G_{ni}(\theta_0) & g_{ni}(\theta_0)g_{ni}'(\theta_0) \end{pmatrix},$$

$$\frac{\partial^2 m_{ni}(\gamma_0)}{\partial\gamma_j\partial\gamma'} = -\begin{pmatrix} 0 & G_{ni}^{(j)'} \\ G_{ni}^{(j)} & g_{ni}^{(j)}g_{ni}' + g_{ni}g_{ni}^{(j)'} \end{pmatrix} \text{ for } 1 \le j \le k_\theta,$$

$$\frac{\partial^2 m_{ni}(\gamma_0)}{\partial\gamma_j\partial\gamma'} = -\begin{pmatrix} [G_{ni}^{(1)'}e_{k_g,s}, \ldots, G_{ni}^{(k_\theta)'}e_{k_g,s}] & G_{ni}'e_{k_g,s}g_{ni}' + g_{nis}G_{ni}' \\ g_{ni}e_{k_g,s}'G_{ni} + g_{nis}G_{ni} & -\rho_3 g_{nis}g_{ni}g_{ni}' \end{pmatrix} \text{ for } k_\theta+1 \le j \le k_\theta+k_g, \text{ and } s = j-k_\theta,$$

$$\frac{\partial^3 m_{ni}(\gamma_0)}{\partial\gamma_k\partial\gamma_j\partial\gamma'} = -\begin{pmatrix} 0 & G_{ni}^{(jk)'} \\ G_{ni}^{(jk)} & g_{ni}^{(jk)}g_{ni}' + g_{ni}g_{ni}^{(jk)'} + g_{ni}^{(j)}g_{ni}^{(k)'} + g_{ni}^{(k)}g_{ni}^{(j)'} \end{pmatrix} \text{ for } 1 \le j \le k_\theta \text{ and } 1 \le k \le k_\theta,$$

$$\frac{\partial^3 m_{ni}(\gamma_0)}{\partial\gamma_k\partial\gamma_j\partial\gamma'} = -\begin{pmatrix} [G_{ni}^{(1j)'}e_{k_g,k-k_\theta},\ldots,G_{ni}^{(k_\theta,j)'}e_{k_g,k-k_\theta}] & G_{ni}^{(j)'}e_{k_g,t}g_{ni}' + G_{ni}'e_{k_g,t}g_{ni}^{(j)'} + g_{nit}^{(j)}G_{ni}' + g_{nit}'G_{ni}^{(j)'} \\ g_{ni}e_{k_g,t}'G_{ni}^{(j)} + g_{ni}^{(j)}e_{k_g,t}'G_{ni} + g_{nit}^{(j)}G_{ni} + g_{nit}G_{ni}^{(j)} & -\rho_3 g_{nit}^{(j)}g_{ni}g_{ni}' - \rho_3 g_{nit}(g_{ni}^{(j)}g_{ni}' + g_{ni}g_{ni}^{(j)'}) \end{pmatrix}$$

for $1 \le j \le k_\theta$, $k_\theta + 1 \le k \le k_\theta + k_g$, and $t = k - k_\theta$,

$$\frac{\partial^3 m_{ni}(\gamma_0)}{\partial\gamma_k\partial\gamma_j\partial\gamma'} = -\begin{pmatrix} [G_{ni}^{(1k)'}e_{k_g,j-k_\theta},\ldots,G_{ni}^{(k_\theta,k)'}e_{k_g,j-k_\theta}] & G_{ni}^{(k)'}e_{k_g,t}g_{ni}' + G_{ni}'e_{k_g,t}g_{ni}^{(k)'} + g_{nit}^{(k)}G_{ni}' + g_{nit}'G_{ni}^{(k)'} \\ g_{ni}e_{k_g,t}'G_{ni}^{(k)} + g_{ni}^{(k)}e_{k_g,t}'G_{ni} + g_{nit}^{(k)}G_{ni} + g_{nit}G_{ni}^{(k)} & -\rho_3 g_{nit}^{(k)}g_{ni}g_{ni}' - \rho_3 g_{nit}(g_{ni}^{(k)}g_{ni}' + g_{ni}g_{ni}^{(k)'}) \end{pmatrix}$$

for $k_\theta + 1 \le j \le k_\theta + k_g$, $1 \le k \le k_\theta$ and $t = j - k_\theta$, and

$$\frac{\partial^3 m_{ni}(\gamma_0)}{\partial\gamma_k\partial\gamma_j\partial\gamma'} = \begin{pmatrix} -G_{ni}'e_{k_g,s}e_{k_g,t}'G_{ni} - G_{ni}'e_{k_g,t}e_{k_g,s}'G_{ni} & \rho_3 g_{nis}G_{ni}'e_{k_g,t}g_{ni}' + \rho_3 g_{nit}G_{ni}'e_{k_g,s}g_{ni}' + \rho_3 g_{nis}g_{nit}G_{ni}' \\ \rho_3 g_{nis}g_{ni}e_{k_g,t}'G_{ni} + \rho_3 g_{nit}g_{ni}e_{k_g,s}'G_{ni} + \rho_3 g_{nis}g_{nit}G_{ni} & \rho_4 g_{nis}g_{nit}g_{ni}g_{ni}' \end{pmatrix}$$
$$- \begin{pmatrix} g_{nis}[G_{ni}^{(1)'}e_{k_g,t},\ldots,G_{ni}^{(k_\theta)'}e_{k_g,t}] + g_{nit}[G_{ni}^{(1)'}e_{k_g,s},\ldots,G_{ni}^{(k_\theta)'}e_{k_g,s}] & 0 \\ 0 & 0 \end{pmatrix},$$

for $k_\theta + 1 \le j \le k_\theta + k_g$, $k_\theta + 1 \le k \le k_\theta + k_g$, $s = j - k_\theta$ and $t = k - k_\theta$. By a second order Taylor expansion of (23),

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

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

$$\begin{aligned} &\sqrt{n}(\hat\gamma_{n,\mathrm{GEL}} - \gamma_0) \\ &= -\Big[\frac{1}{n}\sum_{i=1}^n \mathrm{E}\Big(\frac{\partial m_{ni}(\gamma_0)}{\partial\gamma'}\Big)\Big]^{-1}\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^n m_{ni}(\gamma_0) + \frac{1}{\sqrt{n}}\Big[\sum_{i=1}^n \frac{\partial m_{ni}(\gamma_0)}{\partial\gamma'} - \mathrm{E}\Big(\sum_{i=1}^n \frac{\partial m_{ni}(\gamma_0)}{\partial\gamma'}\Big)\Big](\hat\gamma_{n,\mathrm{GEL}} - \gamma_0) \\ &\quad + \frac{1}{2}\frac{1}{\sqrt{n}}\sum_{i=1}^n\sum_{j=1}^{k_\theta+k_g}(\hat\gamma_{nj,\mathrm{GEL}} - \gamma_{0j})\frac{\partial^2 m_{ni}(\gamma_0)}{\partial\gamma_j\partial\gamma'}(\hat\gamma_{n,\mathrm{GEL}} - \gamma_0)\Big\} + O_p(\sqrt{n}\|\hat\gamma_{n,\mathrm{GEL}} - \gamma_0\|^3). \end{aligned} \tag{24}$$

Thus,

$$\sqrt{n}(\hat\gamma_{n,\mathrm{GEL}} - \gamma_0) = \xi_n + O_p(n^{-1/2}), \tag{25}$$

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 (25) into the second and third terms of (24) yields $\sqrt{n}(\hat\gamma_{n,\mathrm{GEL}} - \gamma_0) = \xi_n + n^{-1/2}\psi_n + O_p(n^{-1})$, where

$$\begin{aligned} \psi_n &= -\Big[\frac{1}{n}\sum_{i=1}^n \mathrm{E}\Big(\frac{\partial m_{ni}(\gamma_0)}{\partial\gamma'}\Big)\Big]^{-1}\Big\{\frac{1}{\sqrt{n}}\Big[\sum_{i=1}^n \frac{\partial m_{ni}(\gamma_0)}{\partial\gamma'} - \mathrm{E}\Big(\sum_{i=1}^n \frac{\partial m_{ni}(\gamma_0)}{\partial\gamma'}\Big)\Big]\xi_n + \frac{1}{2n}\sum_{i=1}^n\sum_{j=1}^{k_\theta+k_g}\xi_{nj}\Big[\mathrm{E}\Big(\frac{\partial^2 m_{ni}(\gamma_0)}{\partial\gamma_j\partial\gamma'}\Big)\Big]\xi_n\Big\} \\ &= -\sqrt{n}\bar K_n^{-1}\begin{pmatrix} 0 & G_n' - \bar G_n' \\ G_n - \bar G_n & \Omega_n - \bar\Omega_n \end{pmatrix}\xi_n + \frac{1}{2n}\bar K_n^{-1}\sum_{i=1}^n\sum_{j=1}^{k_\theta+k_g}\xi_{nj}\Big[\mathrm{E}\Big(\frac{\partial^2 m_{ni}(\gamma_0)}{\partial\gamma_j\partial\gamma'}\Big)\Big]\xi_n \\ &= -\sqrt{n}\bar K_n^{-1}\begin{pmatrix} 0 & G_n' - \bar G_n' \\ G_n - \bar G_n & \Omega_n - \bar\Omega_n \end{pmatrix}\xi_n - \frac{1}{2n}\bar K_n^{-1}\sum_{i=1}^n\sum_{j=1}^{k_\theta}\xi_{nj}\,\mathrm{E}\begin{pmatrix} 0 & G_{ni}^{(j)'} \\ G_{ni}^{(j)} & g_{ni}^{(j)}g_{ni}' + g_{ni}g_{ni}^{(j)'} \end{pmatrix}\xi_n \\ &\quad - \frac{1}{2n}\bar K_n^{-1}\sum_{i=1}^n\sum_{s=1}^{k_g}\xi_{n,k_\theta+s}\,\mathrm{E}\begin{pmatrix} [G_{ni}^{(1)'}e_{k_g,s},\ldots,G_{ni}^{(k_\theta)'}e_{k_g,s}] & G_{ni}'e_{k_g,s}g_{ni}' + g_{nis}G_{ni}' \\ g_{ni}e_{k_g,s}'G_{ni} + g_{nis}G_{ni} & -\rho_3 g_{nis}g_{ni}g_{ni}' \end{pmatrix}\xi_n. \end{aligned} \tag{26}$$

$\square$

Proof of Proposition 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 Proposition 3.3, with ψ_n in (22),

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

Since $\xi_n^{\Omega} = \sqrt{n}(\Omega_n - \bar{\Omega}_n) + \sum_{j=1}^{k_\theta}\frac{1}{n}\sum_{i=1}^{n}[\mathrm{E}(g_{ni}g_{ni}^{(j)'} + g_{ni}^{(j)}g_{ni}')]\tilde{\xi}_{nj}$ by Lemma C.9, where $\tilde{\xi}_n = -\binom{\bar{H}_n^J}{\bar{D}_n^J}\sqrt{n}g_n(\theta_0)$ by Lemma C.8,

$$\mathrm{E}(n^{-1/2}\xi_n^{\Omega}\bar{D}_n g_n) = \mathrm{E}(\Omega_n\bar{D}_n g_n) - \sum_{j=1}^{k_\theta}\frac{1}{n^2}\sum_{i=1}^{n}[\mathrm{E}(g_{ni}g_{ni}^{(j)'} + g_{ni}^{(j)}g_{ni}')]\bar{D}_n\bar{\Omega}_n\bar{H}_n^{J'}e_{k_\theta,j}.$$

Since $\bar{K}_n^{-1} = \begin{pmatrix}-\bar{\Sigma}_n & \bar{H}_n \\ \bar{H}_n' & \bar{D}_n\end{pmatrix}$ and $\bar{D}_n\bar{\Omega}_n\bar{H}_n' = 0$, 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

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

$\square$

Proof of Proposition 3.6. As in the proof of Proposition 3.5, $\mathrm{E}(\xi_n\xi_n') = \mathrm{diag}(\bar{\Sigma}_n, \bar{D}_n)$. Then by Proposition 3.4, with ψ_n in (26),

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

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

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

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

$$
\begin{aligned}
&-\bar{\Sigma}_n \,\mathrm{E}(G_n'\bar{D}_n g_n) + \frac{1}{n^2}\bar{\Sigma}_n\sum_{i=1}^{n}\mathrm{E}(G_{ni}'\bar{D}_n g_{ni}) + \bar{H}_n \,\mathrm{E}(G_n\bar{H}_n g_n) + \bar{H}_n \,\mathrm{E}(\Omega_n\bar{D}_n g_n) - \frac{1}{2n}\sum_{j=1}^{k_\theta}\bar{H}_n\bar{G}_n^{(j)}\bar{\Sigma}_n e_{k_\theta,j} \\
&+\frac{\rho_3}{2n^2}\sum_{i=1}^{n}\bar{H}_n \,\mathrm{E}(g_{ni}g_{ni}'\bar{D}_n g_{ni}).
\end{aligned}
$$
□

Proof of Proposition 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 C.5 and (18) successively, we have

$$
\begin{aligned}
-2n\Big[\rho(0) - \frac{1}{n}\sum_{i=1}^{n}\rho(\hat{\lambda}_n' g_{ni}(\hat{\theta}_n))\Big] &= -\sum_{i=1}^{n}\rho_2(\check{\lambda}_n' g_{ni}(\hat{\theta}_n))\hat{\lambda}_n' g_{ni}(\hat{\theta}_n)g_{ni}'(\hat{\theta}_n)\hat{\lambda}_n \\
&= \sum_{i=1}^{n}\hat{\lambda}_n' g_{ni}(\hat{\theta}_n)g_{ni}'(\hat{\theta}_n)\hat{\lambda}_n + o_p(1) \\
&= n\hat{\lambda}_n'\bar{\Omega}_n\hat{\lambda}_n + o_p(1) \\
&= [\bar{\Omega}_n^{-1/2}\sqrt{n}g_n(\theta_0)]'\bar{\Omega}_n^{1/2}\bar{D}_n\bar{\Omega}_n^{1/2}[\bar{\Omega}_n^{-1/2}\sqrt{n}g_n(\theta_0)] + o_p(1),
\end{aligned}
\tag{27}
$$

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 Proposition 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 Proposition 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 (27) above:

$$
-2n\Big[\rho(0) - \frac{1}{n}\sum_{i=1}^{n}\rho(\dot{\lambda}_n' g_{ni}(\dot{\theta}_n))\Big] = [\bar{\Omega}_n^{-1/2}\sqrt{n}g_n(\theta_0)]'\bar{\Omega}_n^{1/2}\bar{D}_{n\phi}\bar{\Omega}_n^{1/2}[\bar{\Omega}_n^{-1/2}\sqrt{n}g_n(\theta_0)] + o_p(1). \tag{28}
$$

Combining (27) and (28) yields

$$2\Big[\sum_{i=1}^{n}\rho(\dot{\lambda}_n' g_{ni}(\dot{\theta}_n)) - \sum_{i=1}^{n}\rho(\hat{\lambda}_n' g_{ni}(\hat{\theta}_n))\Big] = [\bar{\Omega}_n^{-1/2}\sqrt{n}g_n(\theta_0)]'\bar{\Omega}_n^{1/2}(\bar{D}_{n\phi} - \bar{D}_n)\bar{\Omega}_n^{1/2}[\bar{\Omega}_n^{-1/2}\sqrt{n}g_n(\theta_0)] + o_p(1). \quad (29)$$

Since $\bar{G}_{n\phi}$ is a submatrix of $\bar{G}_n$ and $\text{plim}_{n\to\infty}\frac{1}{n}\bar{G}_n$ has full rank, $\bar{\Omega}_n^{1/2}(\bar{D}_{n\phi}-\bar{D}_n)\bar{\Omega}_n^{1/2} = \bar{\Omega}_n^{-1/2}\bar{G}_n(\bar{G}_n'\bar{\Omega}_n^{-1}\bar{G}_n)^{-1}\bar{G}_n'\bar{\Omega}_n^{-1/2} - \bar{\Omega}_n^{-1/2}\bar{G}_{n\phi}(\bar{G}_{n\phi}'\bar{\Omega}_n^{-1}\bar{G}_{n\phi})^{-1}\bar{G}_{n\phi}'\bar{\Omega}_n^{-1/2} = \mathcal{M}_n\bar{\Omega}_n^{-1/2}\bar{G}_{n\alpha}(\bar{G}_{n\alpha}'\bar{\Omega}_n^{-1/2}\mathcal{M}_n\bar{\Omega}_n^{-1/2}\bar{G}_{n\alpha})^{-1}\bar{G}_{n\alpha}'\bar{\Omega}_n^{-1/2}\mathcal{M}_n$ is a projection matrix with rank k_α, where $\mathcal{M}_n = I_{k_g} - \bar{\Omega}_n^{-1/2}\bar{G}_{n\phi}(\bar{G}_{n\phi}'\bar{\Omega}_n^{-1}\bar{G}_{n\phi})^{-1}\bar{G}_{n\phi}'\bar{\Omega}_n^{-1/2}$ (Ruud, 2000, p. 60, (3.13)). Hence the proposition follows. □

Proof of Proposition 4.2. With the Pitman drift in the proposition, 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 $V_n(\theta)$ can be expanded in a form similar to that in (13), where θ_0, S_n and R_n are replaced by, respectively, $\theta_n = (\alpha_n', \phi_0)'$, $S_n(\kappa_n)$ and $R_n(\tau_n)$. Then under the Pitman drift, Lemmas C.3–C.7 and C.10–C.12 all hold by similar arguments. Hence, as in the proof of Proposition 4.1, we have (29). 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, $\sqrt{n}g_n(\theta_n) = \frac{1}{\sqrt{n}}[V_n'P_{n1}V_n - \text{E}(V_n'P_{n1}V_n), \ldots, V_n'P_{n,k_p}V_n - \text{E}(V_n'P_{n,k_p}V_n), V_n'Q_n] \xrightarrow{d} N(0, \lim_{n\to\infty}\bar{\Omega}_n)$. 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 proposition holds by (29). □

Proof of Proposition 4.3. The asymptotic distribution follows by (27) in the proof of Proposition 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 Proposition 4.4. Explicitly, $\hat{g}_n = \frac{1}{n}\sum_{i=1}^n \hat{g}_{ni} = \frac{1}{n}\hat{V}_n'W_n\hat{V}_n = \frac{1}{n}V_n'W_nV_n - \frac{1}{n}V_n'(W_n+W_n')\mathbb{P}_nV_n + \frac{1}{n}V_n'\mathbb{P}_nW_n\mathbb{P}_nV_n$, where $\mathbb{P}_n = X_n(X_n'X_n)^{-1}X_n'$. Note that $\frac{1}{n}V_n'(W_n + W_n')\mathbb{P}_nV_n = \frac{1}{n\sqrt{n}}V_n'(W_n + W_n')X_n(\frac{1}{n}X_n'X_n)^{-1}\frac{1}{\sqrt{n}}X_n'V_n = O_p(n^{-1})$. Similarly, $\frac{1}{n}V_n'\mathbb{P}_nW_n\mathbb{P}_nV_n = O_p(n^{-1})$. As $\frac{1}{n}V_n'W_nV_n$ has mean zero under both homoskedasticity and unknown heteroskedasticity, $\hat{g}_n = \frac{1}{n}V_n'W_nV_n + O_p(n^{-1}) = O_p(n^{-1/2})$. Then by Lemma C.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 for $\hat{\lambda}_n$ 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 $0 = \sum_{i=1}^n \rho_1(0)\hat{g}_{ni} + \sum_{i=1}^n \rho_2(\check{\lambda}_n\hat{g}_{ni})\hat{g}_{ni}^2\hat{\lambda}_n$, where $\check{\lambda}_n$ lies between 0 and $\hat{\lambda}_n$. Then, because $\sqrt{n}\hat{g}_n = \sqrt{n}g_n + O_p(n^{-1/2})$, where $g_n = \frac{1}{n}V_n'W_nV_n$,

$$\begin{aligned}\sqrt{n}\hat{\lambda}_n &= \Big[\frac{1}{n}\sum_{i=1}^n \rho_2(\check{\lambda}_n\hat{g}_{ni})\hat{g}_{ni}^2\Big]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_{ni} = -\Big[\frac{1}{n}\sum_{i=1}^n \text{E}(g_{ni}^2)\Big]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n g_{ni} + o_p(1)\\ &\xrightarrow{d} N\Big(0, \lim_{n\to\infty}\Big[\frac{1}{n}\sum_{i=1}^n \text{E}(g_{ni}^2)\Big]^{-1}\Big),\end{aligned}$$

where $g_{ni} = v_{ni}\sum_{j=1}^{i-1}(w_{n,ij}+w_{n,ji})v_{nj}$, and the second equality holds because $\frac{1}{n}\sum_{i=1}^n \rho_1(\check{\lambda}_n\hat{g}_{ni})\hat{g}_{ni}^2 = -\frac{1}{n}\sum_{i=1}^n \hat{g}_{ni}^2 + o_p(1)$ as in the proof of Proposition 3.2 and $\frac{1}{n}\sum_{i=1}^n \hat{g}_{ni}^2 = \frac{1}{n}\sum_{i=1}^n \text{E}(g_{ni}^2) + o_p(1)$ by Lemma C.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 $\hat{\lambda}_n$, $\rho(0) = \frac{1}{n}\sum_{i=1}^n \rho(\hat{\lambda}_n\hat{g}_{ni}) + \frac{1}{2n}\sum_{i=1}^n \rho_2(\check{\lambda}_n\hat{g}_{ni})\hat{g}_{ni}^2\hat{\lambda}_n^2$, where $\check{\lambda}_n$ lies between 0 and $\hat{\lambda}_n$. Hence,

$$2n\Big[\frac{1}{n}\sum_{i=1}^n \rho(\hat{\lambda}_n\hat{g}_{ni}) - \rho(0)\Big] = -(\sqrt{n}\hat{\lambda}_n)^2\frac{1}{n}\sum_{i=1}^n \rho_2(\check{\lambda}_n\hat{g}_{ni})\hat{g}_{ni}^2 = (\sqrt{n}\hat{\lambda}_n)^2\frac{1}{n}\sum_{i=1}^n \text{E}(g_{ni}^2) + o_p(1) \xrightarrow{d} \chi^2(1). \quad (30)$$

□

Proof of Proposition 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 Proposition 4.4,

$$\sqrt{n}\hat{\lambda}_n = \Big[\frac{1}{n}\sum_{i=1}^n \rho_1(\check{\lambda}_n \hat{g}_{ni})\hat{g}_{ni}^2\Big]^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{g}_{ni} = -\Big[\frac{1}{n}\sum_{i=1}^n \mathrm{E}(g_{ni}^2)\Big]^{-1} \frac{1}{\sqrt{n}} V_n'(\theta_0) W_n V_n(\theta_0) + o_p(1).$$

By the mean value theorem,

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

where $\check{\theta}_n$ lies between θ_0 and θ_n elementwise. Since $\frac{1}{\sqrt{n}} V_n'(\theta_n) W_n V_n(\theta_n) = \frac{1}{n} V_n' W_n V_n$, and $\frac{1}{n}\frac{\partial [V_n'(\check{\theta}_n) W_n V_n(\check{\theta}_n)]}{\partial \tau} = -\frac{1}{n}\mathrm{E}[V_n'(W_n + W_n') W_n V_n] + o_p(1)$, the result in the proposition follows by the expansion in (30). □

Proof of Proposition 4.6. By the mean value theorem and Lemma C.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. 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 Proposition 4.4. □

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

First, for $\bar{Y}_n^*$ and ϵ_n^* defined in Appendix B, we have the following results: (i) $\frac{1}{n} C_n'(\hat{Y}_n - \bar{Y}_n^*) = o_p(1)$, (ii) $\frac{1}{n} V_n' A_n(\hat{Y}_n - \bar{Y}_n^*) = \frac{1}{n}\mathrm{E}(V_n' A_n \epsilon_n^*) + o_p(1)$, and (iii) $\frac{1}{n}(\hat{Y}_n - \bar{Y}_n^*)' A_n (\hat{Y}_n - \bar{Y}_n^*) = \frac{1}{n}\mathrm{E}(\epsilon_n^{*'} A_n \epsilon_n^*) + o_p(1)$, 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. For $\hat{Y}_n = S_{1n}^{-1}(\hat{\kappa}_{1n}) X_{1n}\hat{\beta}_{1n}$, the results follow by the mean value theorem, Proposition B.1 and the fact that $S_{1n}^{-1}(\kappa_1)$ is bounded in both row and column sum norms in a neighborhood of κ_{1n}^*. For $\hat{Y}_n = \hat{\kappa}_{1n} W_{1n} Y_n + X_{1n}\hat{\beta}_{1n}$, note that $\hat{Y}_n - \bar{Y}_n^* = (\hat{\kappa}_{1n} - \kappa_{1n}^*) W_{1n} S_n^{-1} X_n \beta_0 + X_{1n}(\hat{\beta}_{1n} - \beta_{1n}^*) + \hat{\kappa}_{1n} W_{1n} S_n^{-1} R_n^{-1} V_n$. Substituting this expression into the terms in (i)–(iii), we can easily see that the results hold.

With the above results, we first prove the uniform convergence $\sup_{\vartheta\in\Theta}|g_n(\vartheta) - \bar{g}_n(\vartheta)| = o_p(1)$ if the IV matrix Q_n does not contain the generated regressor $S_{1n}^{-1}(\hat{\kappa}_{1n}) X_{1n}\hat{\beta}_{1n}$. Since $\frac{1}{n} Q_n' V_n(\vartheta) = \frac{1}{n} Q_n' V_n(\theta) - \frac{1}{n}\eta Q_n' R_n(\tau)\hat{Y}_n = \frac{1}{n} Q_n' V_n(\theta) - \frac{1}{n}\eta Q_n'[R_n + (\tau_0 - \tau) M_n]\hat{Y}_n$ and $V_n(\theta)$ in (13) is linear in V_n and quadratic in θ, $\sup_{\vartheta\in\Theta}\|\frac{1}{n} Q_n' V_n(\vartheta) - \frac{1}{n} Q_n' d_n(\vartheta)\| = o_p(1)$, where $d_n(\vartheta) = d_n(\theta) - \eta R_n(\tau)\bar{Y}_n^*$ and $d_n(\theta)$ is given in Appendix A. Rewrite $V_n(\vartheta) = d_n(\vartheta) + \tilde{V}_n(\theta) - \eta R_n(\tau)(\hat{Y}_n - \bar{Y}_n^*)$, where $\tilde{V}_n(\theta) = V_n(\theta) - d_n(\theta)$. As $V_n(\vartheta)$ is quadratic in ϑ, $\sup_{\vartheta\in\Theta}|\frac{1}{n} V_n'(\vartheta) P_{ni} V_n(\vartheta) - \frac{1}{n} d_n'(\vartheta) P_{ni} d_n(\vartheta) - \frac{1}{n}\mathrm{E}[\tilde{V}_n'(\theta) P_{ni}\tilde{V}_n(\theta)] - \frac{1}{n}\eta^2\mathrm{E}[\epsilon_n^{*'} R_n'(\tau) P_{ni} R_n(\tau)\epsilon_n^*] + \frac{2}{n}\mathrm{E}[\tilde{V}_n'(\theta) P_{ni} R_n(\tau)\epsilon_n^*]| = o_p(1)$. Hence, $\sup_{\vartheta\in\Theta}|g_n(\vartheta) - \bar{g}_n(\vartheta)| = o_p(1)$, where

$$\bar{g}_n(\vartheta) = \frac{1}{n}\{\mathrm{E}[\tilde{V}_n'(\theta) P_{n1}\tilde{V}_n(\theta)] + \eta^2\mathrm{E}[\epsilon_n^{*'} R_n'(\tau) P_{n1} R_n(\tau)\epsilon_n^*] - 2\mathrm{E}[\tilde{V}_n'(\theta) P_{n1} R_n(\tau)\epsilon_n^*], \ldots,$$

$$\mathrm{E}[\tilde{V}_n'(\theta)P_{n,k_p}\tilde{V}_n(\theta)] + \eta^2\,\mathrm{E}[\epsilon_n^{*'}R_n'(\tau)P_{n,k_p}R_n(\tau)\epsilon_n^*] - 2\,\mathrm{E}[\tilde{V}_n'(\theta)P_{n,k_p}R_n(\tau)\epsilon_n^*], d_n'(\vartheta)Q_n\}'.$$

In the case that Q_n contains $S_{1n}^{-1}(\hat{\kappa}_{1n})X_{1n}\hat{\beta}_{1n}$, as explained in Section B.1, $\frac{1}{n}V_n'A_nS_{1n}^{-1}(\hat{\kappa}_{1n})X_{1n}\hat{\beta}_{1n} = o_p(1)$ and $\frac{1}{n}b_n'A_nS_{1n}^{-1}(\hat{\kappa}_{1n})X_{1n}\hat{\beta}_{1n} = \frac{1}{n}b_n'A_nS_{1n}^{-1}(\kappa_{1n}^*)X_{1n}\beta_{1n}^* + o_p(1)$ by the mean value theorem, where A_n is an $n\times n$ nonstochastic matrix that is bounded in both row and column sum norms and b_n is an $n\times 1$ vector of uniformly bounded constants. Then $S_{1n}^{-1}(\hat{\kappa}_{1n})X_{1n}\hat{\beta}_{1n}$ is asymptotically exogenous. For Q_n^* obtained by replacing $S_{1n}^{-1}(\hat{\kappa}_{1n})X_{1n}\hat{\beta}_{1n}$ with $S_{1n}^{-1}(\kappa_{1n}^*)X_{1n}\beta_{1n}^*$ in Q_n, the above argument for $\sup_{\vartheta\in\Theta}|g_n(\vartheta)-\bar{g}_n(\vartheta)| = o_p(1)$ still holds if Q_n in $\bar{g}_n(\vartheta)$ is replaced by Q_n^*. Since each element of $\bar{g}_n(\vartheta)$ can be expressed as a polynomial of $(\vartheta-\vartheta_0)$, as in Appendix A, the identification condition in Assumption 14 guarantees that $\lim_{n\to\infty}\frac{1}{n}\bar{g}_n(\vartheta)$ is uniquely zero at ϑ_0.

Next we can show that Lemmas C.4 and C.10–C.12 hold if $g_{ni}(\theta)$ and $g_n(\theta)$ are replaced by, respectively, $g_{ni}(\vartheta)$ and $g_n(\vartheta)$. By these lemmas, the consistency of $\hat{\vartheta}_n$ follows as in the proof of Proposition 3.1. For Lemma C.4, note that $V_n(\theta)-V_n(\vartheta) = \eta R_n(\tau)\hat{Y}_n$. If $\hat{Y}_n = S_{1n}^{-1}(\hat{\lambda}_{1n})X_{1n}\hat{\beta}_{1n}$, by the mean value theorem, $\eta R_n(\tau)\hat{Y}_n = \eta R_n(\tau)\bar{Y}_n^* + \eta R_n(\tau)S_{1n}^{-1}(\check{\lambda}_{1n})W_{1n}S_{1n}^{-1}(\check{\lambda}_{1n})X_{1n}\check{\beta}_{1n}(\hat{\lambda}_{1n} - \bar{\lambda}_{1n}^*) + \eta R_n(\tau)S_{1n}^{-1}(\check{\lambda}_{1n})X_{1n}(\hat{\beta}_{1n}-\beta_{1n}^*)$, where $\check{\theta}_{1n}$ lies between $\hat{\theta}_{1n}$ and θ_{1n}^* elementwise. Since elements of $R_n(\tau)\bar{Y}_n^*$ are uniformly bounded, and elements of $R_n(\tau)S_{1n}^{-1}(\check{\lambda}_{1n})X_{1n}$ and $R_n(\tau)S_{1n}^{-1}(\check{\lambda}_{1n})X_{1n}$ are uniformly bounded in probability, elements of $\eta R_n(\tau)\hat{Y}_n$ are uniformly bounded in probability. By the original Lemma C.4 for $g_{ni}(\theta)$, the lemma still holds if $g_{ni}(\theta)$ is replaced by $g_{ni}(\vartheta)$ when $\hat{Y}_n = S_{1n}^{-1}(\hat{\lambda}_{1n})X_{1n}\hat{\beta}_{1n}$. If $\hat{Y}_n = \hat{\lambda}_{1n}W_{1n}Y_n + X_{1n}\hat{\beta}_{1n}$, then $\eta R_n(\tau)\hat{Y}_n = \eta R_n(\tau)\bar{Y}_n^* + \eta R_n(\tau)W_{1n}S_n^{-1}X_n\beta_0(\hat{\lambda}_{1n} - \lambda_{1n}^*) + \eta R_n(\tau)X_{1n}(\hat{\beta}_{1n}-\beta_{1n}^*) + \eta R_n(\tau)W_{1n}S_n^{-1}R_n^{-1}V_n\hat{\lambda}_{1n}$. Since elements of $\eta R_n(\tau)\bar{Y}_n^*$, $\eta R_n(\tau)W_{1n}S_n^{-1}X_n\beta_0$ and $\eta R_n(\tau)X_{1n}$ are uniformly bounded, and $\eta R_n(\tau)W_{1n}S_n^{-1}R_n^{-1}V_n$ is linear in V_n, Lemma C.4 still holds for $g_{ni}(\vartheta)$. Then Lemmas C.10–C.12 also hold.

With the consistency that $\hat{\vartheta}_n = \vartheta_0 + o_p(1)$, similar to the proof of Proposition 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 Proposition 4.1. □

Proof of Proposition B.1. Let $\hat{\lambda}_{1n}(\theta_1) = \arg\max_{\lambda_1\in\Lambda_1}\sum_{i=1}^n \rho(\lambda_1'g_{1n,i}(\theta_1))$. Under Assumption 13, as in the proof of Lemma 1 in Hong et al. (2003), $\sup_{\theta_1\in\Theta_1}\|\hat{\lambda}_{1n}(\theta_1)-\lambda_{1n}^*(\theta_1)\| = o_p(1)$. Then $\sup_{\theta_1\in\Theta_1}\frac{1}{n}|\,\mathrm{E}\sum_{i=1}^n\rho(\hat{\lambda}_{1n}'(\theta_1)g_{1n,i}(\theta_1)) - \mathrm{E}\sum_{i=1}^n\rho(\lambda_{1n}^{*'}(\theta_1)g_{1n,i}(\theta_1))| = o_p(1)$ under Assumption 13(*iii*). By Assumption 13(*ii*),

$$\sup_{\theta_1\in\Theta_1}\frac{1}{n}\Big|\sum_{i=1}^n\rho\big(\hat{\lambda}_{1n}'(\theta_1)g_{1n,i}(\theta_1)\big) - \mathrm{E}\sum_{i=1}^n\rho\big(\hat{\lambda}_{1n}'(\theta_1)g_{1n,i}(\theta_1)\big)\Big| = o_p(1).$$

Thus, $\sup_{\theta_1\in\Theta_1}\frac{1}{n}|\sum_{i=1}^n\rho(\hat{\lambda}_{1n}'(\theta_1)g_{1n,i}(\theta_1)) - \mathrm{E}\sum_{i=1}^n\rho((\lambda_{1n}^{*'}(\theta_1)g_{1n,i}(\theta_1))| = o_p(1)$. Hence, $\hat{\theta}_{1n}-\theta_{1n}^* = o_p(1)$ (White, 1994, Theorem 3.4 on p. 28). If follows that $\hat{\lambda}_{1n} = \hat{\lambda}_{1n}(\hat{\theta}_{1n}) = \lambda_{1n}^*(\hat{\theta}_{1n}) + o_p(1) = \lambda_{1n}^*(\theta_{1n}^*) + o_p(1) = \lambda_{1n}^* + o_p(1)$. □

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 asymptotc 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.

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.

Davidson, J., 1994. Stochastic limit theory: An introduction for econometricians. Oxford University Press, USA.

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

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. the 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.

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. The 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., 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., 2016. Outer-product-of-gradients tests for spatial autoregressive models. Working paper.

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.

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.

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.

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

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.

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.

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.

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.

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

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

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

R^2, κ_0, τ_0		κ	τ	β_1	β_2	β_3	σ^2
	$n = 144$						
0.8, 0.2, 0.2	GMM	-0.003[0.080]0.080	-0.010[0.131]0.131	0.009[0.157]0.157	-0.003[0.032]0.032	-0.001[0.031]0.031	-0.011[0.015]0.019
	EL	-0.003[0.079]0.079	-0.013[0.127]0.127	0.010[0.154]0.154	-0.003[0.032]0.032	-0.001[0.031]0.031	-0.007[0.015]0.016
	ET	-0.003[0.078]0.078	-0.013[0.126]0.127	0.009[0.153]0.154	-0.003[0.032]0.032	-0.001[0.031]0.031	-0.009[0.015]0.018
0.8, 0.2, 0.4	GMM	-0.002[0.081]0.082	-0.010[0.119]0.120	0.008[0.169]0.170	-0.002[0.031]0.031	-0.002[0.031]0.031	-0.012[0.015]0.019
	EL	0.000[0.079]0.079	-0.019[0.115]0.117	0.003[0.165]0.165	-0.001[0.031]0.031	-0.002[0.031]0.031	-0.007[0.015]0.016
	ET	0.000[0.079]0.079	-0.018[0.115]0.117	0.003[0.165]0.165	-0.001[0.031]0.031	-0.002[0.031]0.031	-0.009[0.015]0.018
0.8, 0.4, 0.2	GMM	-0.005[0.069]0.069	-0.009[0.127]0.127	0.013[0.172]0.173	-0.001[0.031]0.031	-0.000[0.030]0.030	-0.012[0.015]0.019
	EL	-0.006[0.067]0.068	-0.011[0.122]0.122	0.016[0.169]0.169	-0.001[0.031]0.031	-0.001[0.030]0.030	-0.007[0.015]0.017
	ET	-0.006[0.067]0.067	-0.010[0.122]0.122	0.016[0.168]0.169	-0.001[0.031]0.031	-0.001[0.030]0.030	-0.010[0.015]0.018
0.8, 0.4, 0.4	GMM	-0.006[0.076]0.076	-0.011[0.121]0.122	0.016[0.198]0.199	-0.002[0.030]0.030	0.000[0.032]0.032	-0.011[0.015]0.019
	EL	-0.005[0.075]0.075	-0.017[0.118]0.119	0.013[0.195]0.196	-0.002[0.030]0.030	0.000[0.032]0.032	-0.007[0.015]0.016
	ET	-0.005[0.075]0.075	-0.016[0.118]0.119	0.014[0.195]0.196	-0.002[0.030]0.030	0.000[0.032]0.032	-0.009[0.015]0.018
0.4, 0.2, 0.2	GMM	-0.005[0.177]0.177	-0.021[0.211]0.212	0.018[0.357]0.358	-0.006[0.077]0.078	-0.004[0.076]0.076	-0.081[0.089]0.120
	EL	-0.002[0.160]0.160	-0.025[0.191]0.193	0.008[0.327]0.327	-0.004[0.077]0.077	-0.002[0.076]0.076	-0.050[0.089]0.102
	ET	-0.002[0.160]0.160	-0.024[0.191]0.192	0.008[0.328]0.328	-0.004[0.078]0.078	-0.001[0.076]0.076	-0.066[0.088]0.110
0.4, 0.2, 0.4	GMM	0.000[0.185]0.185	-0.026[0.200]0.201	0.015[0.394]0.394	-0.001[0.079]0.079	-0.005[0.077]0.077	-0.081[0.093]0.123
	EL	0.008[0.169]0.169	-0.035[0.184]0.187	-0.004[0.365]0.365	0.001[0.079]0.079	-0.003[0.077]0.077	-0.050[0.092]0.105
	ET	0.009[0.169]0.169	-0.035[0.183]0.187	-0.004[0.366]0.366	0.002[0.079]0.079	-0.003[0.077]0.077	-0.066[0.092]0.113
0.4, 0.4, 0.2	GMM	-0.017[0.175]0.176	-0.019[0.208]0.209	0.048[0.446]0.448	-0.005[0.081]0.081	-0.003[0.078]0.078	-0.079[0.094]0.123
	EL	-0.023[0.164]0.165	-0.012[0.193]0.193	0.058[0.419]0.423	-0.003[0.080]0.080	-0.001[0.077]0.077	-0.049[0.094]0.106
	ET	-0.023[0.161]0.163	-0.011[0.191]0.191	0.057[0.415]0.419	-0.003[0.080]0.080	-0.001[0.077]0.077	-0.065[0.093]0.114
0.4, 0.4, 0.4	GMM	-0.021[0.184]0.186	-0.021[0.197]0.198	0.054[0.478]0.481	-0.005[0.076]0.076	-0.007[0.076]0.077	-0.082[0.093]0.124
	EL	-0.019[0.169]0.170	-0.023[0.180]0.181	0.048[0.443]0.446	-0.003[0.076]0.076	-0.005[0.077]0.077	-0.052[0.093]0.107
	ET	-0.019[0.169]0.170	-0.023[0.179]0.181	0.047[0.442]0.444	-0.003[0.076]0.076	-0.005[0.076]0.077	-0.068[0.092]0.114
	$n = 400$						
0.8, 0.2, 0.2	GMM	-0.000[0.043]0.043	-0.004[0.069]0.069	0.001[0.086]0.086	-0.000[0.017]0.017	-0.001[0.018]0.018	-0.004[0.009]0.010
	EL	-0.000[0.043]0.043	-0.004[0.068]0.069	0.001[0.086]0.086	-0.000[0.017]0.017	-0.001[0.018]0.018	-0.002[0.009]0.009
	ET	-0.000[0.043]0.043	-0.004[0.068]0.068	0.001[0.086]0.086	-0.000[0.017]0.017	-0.001[0.018]0.018	-0.003[0.009]0.010
0.8, 0.2, 0.4	GMM	-0.002[0.048]0.048	-0.002[0.065]0.065	0.004[0.098]0.098	-0.001[0.019]0.019	-0.001[0.018]0.018	-0.004[0.009]0.010
	EL	-0.001[0.048]0.048	-0.004[0.065]0.065	0.003[0.097]0.098	-0.001[0.019]0.019	-0.000[0.018]0.018	-0.002[0.009]0.009
	ET	-0.001[0.048]0.048	-0.004[0.065]0.065	0.003[0.097]0.097	-0.001[0.019]0.019	-0.000[0.018]0.018	-0.003[0.009]0.010
0.8, 0.4, 0.2	GMM	-0.002[0.040]0.040	-0.003[0.073]0.073	0.006[0.099]0.099	-0.000[0.019]0.019	-0.001[0.018]0.018	-0.004[0.009]0.010
	EL	-0.003[0.039]0.040	-0.003[0.072]0.072	0.007[0.098]0.098	-0.000[0.019]0.019	-0.001[0.018]0.018	-0.002[0.009]0.009
	ET	-0.003[0.039]0.039	-0.003[0.072]0.072	0.007[0.098]0.098	-0.000[0.019]0.019	-0.001[0.018]0.018	-0.003[0.009]0.009
0.8, 0.4, 0.4	GMM	-0.003[0.045]0.045	-0.005[0.070]0.070	0.008[0.116]0.117	-0.000[0.017]0.017	-0.001[0.018]0.018	-0.004[0.009]0.010
	EL	-0.003[0.045]0.045	-0.006[0.069]0.070	0.008[0.116]0.116	-0.000[0.017]0.017	-0.001[0.018]0.018	-0.002[0.009]0.009
	ET	-0.003[0.045]0.045	-0.006[0.069]0.070	0.008[0.116]0.116	-0.000[0.017]0.017	-0.001[0.018]0.018	-0.003[0.009]0.009
0.4, 0.2, 0.2	GMM	-0.001[0.106]0.106	-0.008[0.118]0.118	0.006[0.209]0.209	-0.002[0.043]0.044	-0.002[0.047]0.047	-0.031[0.054]0.063
	EL	-0.000[0.101]0.101	-0.009[0.113]0.113	0.005[0.202]0.202	-0.001[0.043]0.043	-0.001[0.047]0.047	-0.017[0.054]0.057
	ET	-0.000[0.101]0.101	-0.009[0.113]0.113	0.005[0.202]0.202	-0.001[0.043]0.043	-0.001[0.046]0.046	-0.024[0.054]0.059
0.4, 0.2, 0.4	GMM	-0.004[0.108]0.108	-0.007[0.110]0.110	0.011[0.232]0.233	-0.002[0.044]0.044	-0.002[0.043]0.043	-0.032[0.055]0.064
	EL	-0.000[0.105]0.105	-0.012[0.107]0.108	0.004[0.227]0.227	-0.001[0.044]0.044	-0.002[0.043]0.043	-0.017[0.055]0.058
	ET	0.000[0.104]0.104	-0.012[0.106]0.107	0.003[0.226]0.226	-0.001[0.044]0.044	-0.002[0.043]0.043	-0.024[0.055]0.060
0.4, 0.4, 0.2	GMM	-0.009[0.094]0.094	-0.001[0.117]0.117	0.016[0.234]0.234	0.001[0.045]0.045	0.001[0.045]0.045	-0.029[0.054]0.062
	EL	-0.011[0.091]0.092	0.002[0.113]0.113	0.020[0.228]0.229	0.001[0.045]0.045	0.001[0.045]0.045	-0.015[0.054]0.056
	ET	-0.011[0.091]0.091	0.002[0.113]0.113	0.020[0.227]0.228	0.001[0.045]0.045	0.001[0.045]0.045	-0.022[0.054]0.058
0.4, 0.4, 0.4	GMM	-0.003[0.106]0.106	-0.011[0.121]0.121	0.013[0.277]0.278	-0.002[0.043]0.043	-0.003[0.045]0.045	-0.033[0.055]0.064
	EL	-0.003[0.102]0.102	-0.012[0.116]0.117	0.013[0.267]0.268	-0.001[0.043]0.043	-0.003[0.045]0.045	-0.019[0.055]0.058
	ET	-0.003[0.102]0.102	-0.012[0.116]0.116	0.012[0.267]0.267	-0.001[0.043]0.043	-0.002[0.045]0.045	-0.026[0.054]0.060

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

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

R^2, κ_0, τ_0		κ	τ	β_1	β_2	β_3
	$n = 144$					
0.8, 0.2, 0.2	GMM	-0.001[0.048]0.048	0.001[0.229]0.229	0.005[0.108]0.108	-0.001[0.022]0.022	-0.001[0.022]0.022
	EL	-0.001[0.048]0.048	-0.025[0.235]0.237	0.006[0.110]0.110	-0.002[0.023]0.023	-0.001[0.022]0.022
	ET	-0.001[0.046]0.046	-0.022[0.229]0.230	0.005[0.106]0.106	-0.001[0.022]0.022	-0.001[0.021]0.021
0.8, 0.2, 0.4	GMM	-0.001[0.051]0.051	0.013[0.232]0.232	0.004[0.131]0.131	-0.000[0.023]0.023	-0.001[0.022]0.022
	EL	0.001[0.051]0.051	-0.021[0.258]0.259	0.001[0.127]0.127	-0.000[0.023]0.023	-0.000[0.023]0.023
	ET	0.000[0.050]0.050	-0.018[0.245]0.246	0.001[0.123]0.123	-0.000[0.022]0.022	-0.000[0.022]0.022
0.8, 0.4, 0.2	GMM	-0.001[0.047]0.047	0.000[0.240]0.240	0.005[0.130]0.130	-0.001[0.022]0.022	-0.001[0.021]0.021
	EL	-0.001[0.041]0.041	-0.032[0.234]0.236	0.005[0.116]0.116	-0.001[0.021]0.021	-0.000[0.021]0.021
	ET	-0.001[0.041]0.041	-0.029[0.229]0.230	0.004[0.114]0.114	-0.001[0.020]0.020	-0.000[0.020]0.020
0.8, 0.4, 0.4	GMM	-0.003[0.050]0.050	-0.004[0.230]0.230	-0.001[0.270]0.270	-0.001[0.021]0.021	-0.001[0.021]0.021
	EL	-0.003[0.049]0.049	-0.043[0.238]0.242	0.008[0.146]0.146	-0.001[0.021]0.021	-0.001[0.021]0.021
	ET	-0.003[0.047]0.047	-0.036[0.224]0.227	0.010[0.178]0.179	-0.001[0.020]0.020	-0.001[0.020]0.020
0.4, 0.2, 0.2	GMM	-0.001[0.121]0.121	-0.002[0.277]0.277	0.009[0.370]0.370	-0.005[0.054]0.054	-0.005[0.055]0.055
	EL	0.001[0.114]0.114	-0.035[0.261]0.263	0.010[0.263]0.263	-0.004[0.053]0.053	-0.004[0.053]0.053
	ET	0.001[0.109]0.109	-0.030[0.255]0.256	0.010[0.253]0.253	-0.004[0.051]0.051	-0.004[0.051]0.051
0.4, 0.2, 0.4	GMM	-0.005[0.134]0.134	0.003[0.269]0.269	0.031[0.702]0.703	-0.005[0.059]0.059	-0.005[0.058]0.059
	EL	0.000[0.124]0.124	-0.041[0.248]0.251	0.006[0.297]0.297	-0.003[0.057]0.057	-0.003[0.057]0.057
	ET	0.001[0.121]0.121	-0.033[0.258]0.260	0.005[0.288]0.288	-0.003[0.056]0.056	-0.003[0.055]0.055
0.4, 0.4, 0.2	GMM	-0.011[0.122]0.122	0.017[0.279]0.279	0.034[0.357]0.358	-0.004[0.055]0.055	-0.003[0.056]0.056
	EL	-0.011[0.111]0.112	-0.012[0.263]0.263	0.035[0.333]0.335	-0.002[0.054]0.054	-0.002[0.055]0.055
	ET	-0.010[0.106]0.106	-0.009[0.254]0.254	0.031[0.310]0.312	-0.002[0.051]0.051	-0.001[0.053]0.053
0.4, 0.4, 0.4	GMM	-0.009[0.132]0.133	-0.001[0.255]0.255	0.028[0.460]0.461	-0.006[0.056]0.056	-0.004[0.054]0.054
	EL	-0.003[0.119]0.119	-0.040[0.244]0.247	0.012[0.351]0.351	-0.004[0.054]0.054	-0.002[0.052]0.052
	ET	-0.003[0.116]0.116	-0.035[0.238]0.240	0.011[0.344]0.345	-0.004[0.053]0.053	-0.002[0.051]0.051
	$n = 400$					
0.8, 0.2, 0.2	GMM	-0.000[0.025]0.025	0.002[0.127]0.127	0.001[0.060]0.060	-0.000[0.012]0.012	-0.000[0.012]0.012
	EL	-0.000[0.025]0.025	-0.012[0.127]0.128	0.002[0.060]0.060	-0.000[0.012]0.012	-0.000[0.012]0.012
	ET	-0.000[0.025]0.025	-0.011[0.125]0.126	0.001[0.059]0.059	-0.000[0.012]0.012	-0.000[0.012]0.012
0.8, 0.2, 0.4	GMM	0.002[0.029]0.029	-0.004[0.114]0.114	-0.004[0.071]0.072	0.001[0.013]0.013	0.000[0.013]0.013
	EL	0.002[0.029]0.029	-0.018[0.115]0.116	-0.005[0.072]0.072	0.001[0.013]0.013	0.000[0.013]0.013
	ET	0.002[0.029]0.029	-0.017[0.113]0.114	-0.005[0.071]0.071	0.001[0.013]0.013	0.000[0.013]0.013
0.8, 0.4, 0.2	GMM	-0.001[0.024]0.024	0.004[0.120]0.120	0.005[0.066]0.066	-0.001[0.012]0.012	-0.001[0.011]0.011
	EL	-0.001[0.024]0.024	-0.009[0.119]0.119	0.005[0.067]0.067	-0.001[0.012]0.012	-0.001[0.011]0.011
	ET	-0.001[0.024]0.024	-0.008[0.117]0.118	0.005[0.065]0.065	-0.001[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[0.115]0.115	0.003[0.082]0.082	0.000[0.012]0.012	-0.000[0.012]0.012
	EL	-0.001[0.027]0.027	-0.015[0.116]0.116	0.004[0.080]0.080	-0.000[0.012]0.012	-0.000[0.012]0.012
	ET	-0.001[0.026]0.026	-0.014[0.113]0.114	0.004[0.078]0.078	-0.000[0.012]0.012	-0.000[0.012]0.012
0.4, 0.2, 0.2	GMM	0.002[0.062]0.062	-0.002[0.140]0.140	-0.003[0.145]0.145	-0.001[0.029]0.029	-0.002[0.029]0.029
	EL	0.002[0.063]0.063	-0.015[0.141]0.142	-0.003[0.147]0.147	-0.001[0.029]0.029	-0.002[0.029]0.029
	ET	0.003[0.062]0.062	-0.014[0.139]0.139	-0.003[0.144]0.144	-0.001[0.029]0.029	-0.001[0.029]0.029
0.4, 0.2, 0.4	GMM	-0.002[0.069]0.069	0.001[0.129]0.129	0.007[0.174]0.174	-0.001[0.031]0.031	-0.002[0.030]0.030
	EL	-0.001[0.069]0.069	-0.015[0.123]0.124	0.004[0.175]0.175	-0.000[0.031]0.031	-0.001[0.031]0.031
	ET	-0.001[0.068]0.068	-0.014[0.121]0.122	0.004[0.172]0.172	-0.000[0.031]0.031	-0.001[0.030]0.030
0.4, 0.4, 0.2	GMM	-0.005[0.060]0.060	0.014[0.138]0.139	0.015[0.175]0.176	-0.001[0.028]0.029	-0.002[0.029]0.029
	EL	-0.005[0.059]0.059	0.001[0.136]0.136	0.016[0.166]0.167	-0.001[0.028]0.028	-0.002[0.029]0.029
	ET	-0.005[0.058]0.058	0.002[0.134]0.134	0.015[0.162]0.163	-0.001[0.028]0.028	-0.002[0.028]0.028
0.4, 0.4, 0.4	GMM	-0.000[0.067]0.067	-0.003[0.129]0.129	0.006[0.200]0.200	-0.000[0.030]0.030	-0.001[0.029]0.029
	EL	-0.001[0.068]0.068	-0.017[0.129]0.130	0.007[0.202]0.202	-0.000[0.030]0.030	-0.001[0.030]0.030
	ET	-0.001[0.067]0.067	-0.016[0.127]0.128	0.006[0.197]0.197	-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 model (1)

R^2, κ_0, τ_0		Homoskedastic case						Heteroskedastic case				
		κ	τ	β_1	β_2	β_3	σ^2	κ	τ	β_1	β_2	β_3
	$n = 144$											
0.8, 0.2, 0.2	GMM	0.893	0.893	0.902	0.910	0.926	0.769	0.921	0.880	0.921	0.906	0.924
	EL	0.912	0.917	0.915	0.926	0.935	0.873	0.929	0.922	0.929	0.923	0.940
	ET	0.923	0.925	0.927	0.935	0.947	0.828	0.933	0.931	0.935	0.939	0.949
0.8, 0.2, 0.4	GMM	0.920	0.917	0.925	0.917	0.920	0.787	0.933	0.856	0.932	0.926	0.920
	EL	0.932	0.941	0.928	0.921	0.932	0.886	0.938	0.907	0.930	0.937	0.932
	ET	0.938	0.952	0.939	0.936	0.938	0.848	0.947	0.919	0.940	0.949	0.938
0.8, 0.4, 0.2	GMM	0.908	0.907	0.899	0.926	0.923	0.781	0.925	0.873	0.927	0.932	0.937
	EL	0.922	0.920	0.922	0.938	0.935	0.882	0.941	0.904	0.946	0.946	0.957
	ET	0.932	0.930	0.926	0.946	0.945	0.838	0.947	0.921	0.947	0.954	0.963
0.8, 0.4, 0.4	GMM	0.916	0.893	0.923	0.919	0.914	0.793	0.934	0.848	0.934	0.951	0.931
	EL	0.930	0.920	0.934	0.932	0.930	0.885	0.935	0.911	0.935	0.958	0.938
	ET	0.937	0.929	0.941	0.946	0.936	0.847	0.949	0.924	0.945	0.964	0.949
0.4, 0.2, 0.2	GMM	0.848	0.833	0.861	0.917	0.923	0.750	0.919	0.872	0.902	0.925	0.927
	EL	0.878	0.880	0.896	0.928	0.939	0.866	0.934	0.907	0.920	0.945	0.940
	ET	0.885	0.889	0.907	0.935	0.944	0.817	0.948	0.919	0.929	0.950	0.949
0.4, 0.2, 0.4	GMM	0.841	0.840	0.856	0.896	0.919	0.751	0.911	0.841	0.929	0.916	0.927
	EL	0.875	0.878	0.893	0.915	0.934	0.854	0.928	0.897	0.931	0.935	0.943
	ET	0.880	0.892	0.902	0.929	0.939	0.825	0.936	0.905	0.944	0.947	0.953
0.4, 0.4, 0.2	GMM	0.864	0.850	0.870	0.912	0.912	0.776	0.909	0.851	0.914	0.913	0.913
	EL	0.897	0.867	0.891	0.932	0.925	0.890	0.928	0.901	0.920	0.946	0.941
	ET	0.911	0.887	0.910	0.940	0.930	0.849	0.936	0.912	0.929	0.952	0.950
0.4, 0.4, 0.4	GMM	0.836	0.838	0.851	0.916	0.918	0.765	0.912	0.869	0.922	0.902	0.913
	EL	0.872	0.866	0.874	0.929	0.928	0.856	0.925	0.906	0.924	0.922	0.927
	ET	0.881	0.880	0.887	0.937	0.937	0.821	0.937	0.923	0.938	0.935	0.938
	$n = 400$											
0.8, 0.2, 0.2	GMM	0.945	0.932	0.941	0.920	0.951	0.886	0.942	0.916	0.924	0.937	0.948
	EL	0.943	0.934	0.944	0.921	0.955	0.932	0.937	0.922	0.929	0.934	0.946
	ET	0.948	0.942	0.948	0.927	0.958	0.917	0.941	0.929	0.934	0.940	0.950
0.8, 0.2, 0.4	GMM	0.928	0.945	0.934	0.958	0.936	0.887	0.943	0.929	0.940	0.950	0.942
	EL	0.933	0.947	0.936	0.957	0.941	0.938	0.931	0.938	0.941	0.949	0.938
	ET	0.934	0.949	0.941	0.964	0.944	0.911	0.939	0.946	0.944	0.954	0.942
0.8, 0.4, 0.2	GMM	0.948	0.942	0.950	0.940	0.947	0.868	0.942	0.914	0.957	0.949	0.951
	EL	0.953	0.955	0.946	0.940	0.950	0.921	0.942	0.933	0.954	0.949	0.952
	ET	0.955	0.956	0.949	0.946	0.956	0.893	0.943	0.936	0.957	0.951	0.958
0.8, 0.4, 0.4	GMM	0.937	0.920	0.940	0.934	0.950	0.888	0.946	0.931	0.939	0.955	0.956
	EL	0.943	0.929	0.939	0.938	0.951	0.937	0.939	0.942	0.928	0.952	0.950
	ET	0.943	0.933	0.943	0.941	0.953	0.916	0.945	0.945	0.935	0.958	0.957
0.4, 0.2, 0.2	GMM	0.922	0.928	0.922	0.932	0.937	0.880	0.951	0.923	0.943	0.937	0.952
	EL	0.931	0.934	0.929	0.934	0.942	0.924	0.945	0.932	0.937	0.940	0.950
	ET	0.933	0.936	0.934	0.939	0.941	0.909	0.951	0.940	0.944	0.942	0.956
0.4, 0.2, 0.4	GMM	0.911	0.908	0.904	0.937	0.941	0.886	0.948	0.929	0.946	0.935	0.947
	EL	0.922	0.920	0.923	0.938	0.944	0.923	0.947	0.941	0.945	0.933	0.944
	ET	0.925	0.922	0.927	0.944	0.948	0.904	0.952	0.945	0.948	0.935	0.947
0.4, 0.4, 0.2	GMM	0.917	0.909	0.934	0.933	0.939	0.876	0.950	0.934	0.958	0.940	0.961
	EL	0.926	0.917	0.942	0.936	0.942	0.927	0.942	0.944	0.950	0.939	0.966
	ET	0.930	0.920	0.945	0.937	0.945	0.902	0.944	0.949	0.960	0.940	0.967
0.4, 0.4, 0.4	GMM	0.906	0.898	0.907	0.932	0.944	0.868	0.940	0.938	0.943	0.944	0.951
	EL	0.910	0.917	0.915	0.936	0.947	0.917	0.934	0.943	0.935	0.946	0.945
	ET	0.914	0.922	0.920	0.940	0.950	0.897	0.946	0.948	0.941	0.952	0.949

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 Proposition 3.2.

Table 4: Empirical sizes of tests for $\tau_0 = 0$ in an SE model

	Homoskedastic case				Heteroskedastic case			
	$n = 144$		$n = 400$		$n = 144$		$n = 400$	
	$R^2 = 0.8$	$R^2 = 0.4$	$R^2 = 0.8$	$R^2 = 0.4$	$R^2 = 0.8$	$R^2 = 0.4$	$R^2 = 0.8$	$R^2 = 0.4$
PT_{GMM}	0.034	0.061	0.051	0.050	0.057	0.068	0.053	0.058
PT_{EL}	0.072	0.107	0.061	0.061	0.113	0.144	0.082	0.078
PT_{ET}	0.065	0.092	0.063	0.060	0.094	0.122	0.080	0.074
OT_{GMM}	0.043	0.050	0.053	0.042	0.040	0.050	0.052	0.050
OT_{EL}	0.049	0.063	0.054	0.044	0.058	0.073	0.057	0.059
OT_{ET}	0.050	0.059	0.054	0.044	0.050	0.067	0.058	0.057
Moran	0.043	0.050	0.054	0.043	0.041	0.052	0.055	0.050
Moran_{EL}	0.049	0.066	0.054	0.044	0.055	0.063	0.058	0.055
Moran_{ET}	0.052	0.060	0.054	0.044	0.047	0.064	0.058	0.058
PT^*_{GMM}					0.105	0.107	0.156	0.175
Moran^*					0.011	0.013	0.020	0.018

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

Table 5: Powers of tests for $\tau_0 = 0$ in an SE model

		$n = 144$			$n = 400$		
		$\tau_0 = 0.2$	$\tau_0 = 0.4$	$\tau_0 = 0.6$	$\tau_0 = 0.2$	$\tau_0 = 0.4$	$\tau_0 = 0.6$
Homoskedastic case							
$R^2 = 0.8$	PT_{GMM}	0.441	0.952	1.000	0.901	1.000	1.000
	PT_{EL}	0.551	0.982	1.000	0.915	1.000	1.000
	PT_{ET}	0.537	0.980	1.000	0.910	1.000	1.000
	OT_{GMM}	0.474	0.977	1.000	0.913	1.000	1.000
	OT_{EL}	0.500	0.984	1.000	0.913	1.000	1.000
	OT_{ET}	0.503	0.985	1.000	0.915	1.000	1.000
	Moran	0.471	0.978	1.000	0.912	1.000	1.000
	Moran_{EL}	0.497	0.982	1.000	0.914	1.000	1.000
	Moran_{ET}	0.505	0.983	1.000	0.915	1.000	1.000
$R^2 = 0.4$	PT_{GMM}	0.429	0.959	0.999	0.911	1.000	1.000
	PT_{EL}	0.546	0.985	1.000	0.921	1.000	1.000
	PT_{ET}	0.517	0.982	1.000	0.926	1.000	1.000
	OT_{GMM}	0.461	0.973	1.000	0.918	1.000	1.000
	OT_{EL}	0.493	0.977	1.000	0.922	1.000	1.000
	OT_{ET}	0.488	0.975	1.000	0.922	1.000	1.000
	Moran	0.461	0.970	1.000	0.918	1.000	1.000
	Moran_{EL}	0.492	0.977	1.000	0.922	1.000	1.000
	Moran_{ET}	0.488	0.976	1.000	0.923	1.000	1.000
Heteroskedastic case							
$R^2 = 0.8$	PT_{GMM}	0.196	0.550	0.915	0.348	0.932	1.000
	PT_{EL}	0.270	0.631	0.927	0.384	0.932	1.000
	PT_{ET}	0.245	0.625	0.933	0.385	0.939	1.000
	OT_{GMM}	0.127	0.480	0.881	0.296	0.904	1.000
	OT_{EL}	0.133	0.477	0.874	0.294	0.887	1.000
	OT_{ET}	0.133	0.498	0.886	0.304	0.904	1.000
	Moran	0.122	0.457	0.873	0.285	0.898	1.000
	Moran_{EL}	0.131	0.469	0.872	0.287	0.887	1.000
	Moran_{ET}	0.136	0.489	0.885	0.302	0.901	1.000
$R^2 = 0.4$	PT_{GMM}	0.186	0.586	0.930	0.371	0.916	1.000
	PT_{EL}	0.267	0.654	0.944	0.394	0.926	1.000
	PT_{ET}	0.249	0.651	0.952	0.399	0.927	1.000
	OT_{GMM}	0.119	0.496	0.903	0.294	0.869	0.999
	OT_{EL}	0.134	0.489	0.887	0.297	0.864	0.999
	OT_{ET}	0.134	0.512	0.901	0.300	0.873	0.999
	Moran	0.114	0.476	0.900	0.289	0.865	0.999
	Moran_{EL}	0.131	0.479	0.885	0.295	0.861	0.999
	Moran_{ET}	0.131	0.502	0.902	0.299	0.871	0.999

"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; and "Moran", "Moran_{EL} " and "Moran_{EL} " denote, respectively, the robust, EL and ET Moran's I tests.

Table 6: Empirical sizes of tests for $\tau_0 = 0$ in the SARAR model (1)

	$n = 144$				$n = 400$			
	$R^2 = 0.8$		$R^2 = 0.4$		$R^2 = 0.8$		$R^2 = 0.4$	
	$\kappa_0 = 0.2$	$\kappa_0 = 0.4$	$\kappa_0 = 0.2$	$\kappa_0 = 0.4$	$\kappa_0 = 0.2$	$\kappa_0 = 0.4$	$\kappa_0 = 0.2$	$\kappa_0 = 0.4$
Homoskedastic case								
PT_{GMM}	0.049	0.049	0.043	0.048	0.055	0.047	0.048	0.045
PT_{EL}	0.083	0.089	0.067	0.066	0.065	0.050	0.058	0.053
PT_{ET}	0.072	0.079	0.057	0.064	0.068	0.054	0.057	0.056
OT_{GMM}	0.047	0.048	0.048	0.052	0.052	0.050	0.050	0.053
OT_{EL}	0.071	0.070	0.066	0.075	0.065	0.055	0.056	0.061
OT_{ET}	0.068	0.070	0.063	0.066	0.061	0.055	0.057	0.061
Moran	0.054	0.047	0.041	0.052	0.051	0.048	0.059	0.043
$Moran_{EL}$	0.060	0.059	0.054	0.060	0.056	0.052	0.061	0.046
$Moran_{ET}$	0.059	0.059	0.053	0.060	0.056	0.053	0.062	0.046
Heteroskedastic case								
PT_{GMM}	0.054	0.069	0.092	0.067	0.052	0.042	0.063	0.051
PT_{EL}	0.105	0.123	0.127	0.125	0.068	0.060	0.076	0.073
PT_{ET}	0.093	0.103	0.117	0.104	0.062	0.056	0.071	0.069
OT_{GMM}	0.032	0.049	0.040	0.035	0.056	0.045	0.055	0.049
OT_{EL}	0.075	0.093	0.073	0.086	0.065	0.055	0.078	0.064
OT_{ET}	0.064	0.081	0.070	0.075	0.064	0.056	0.072	0.066
Moran	0.043	0.058	0.040	0.048	0.047	0.037	0.048	0.053
$Moran_{EL}$	0.061	0.079	0.063	0.060	0.056	0.044	0.061	0.061
$Moran_{ET}$	0.056	0.071	0.056	0.058	0.055	0.041	0.059	0.062

"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; and "Moran", "$Moran_{EL}$ " and "$Moran_{EL}$ " denote, respectively, the robust, EL and ET Moran's I tests. The nominal size is 5%.

Table 7: Powers of tests for $\tau_0 = 0$ in the SARAR model (1)

		$n = 144$, $\kappa_0 = 0.2$			$n = 144$, $\kappa_0 = 0.4$			$n = 400$, $\kappa_0 = 0.2$			$n = 400$, $\kappa_0 = 0.4$		
		$\tau_0 = 0.2$	$\tau_0 = 0.4$	$\tau_0 = 0.6$	$\tau_0 = 0.2$	$\tau_0 = 0.4$	$\tau_0 = 0.6$	$\tau_0 = 0.2$	$\tau_0 = 0.4$	$\tau_0 = 0.6$	$\tau_0 = 0.2$	$\tau_0 = 0.4$	$\tau_0 = 0.6$
Homoskedastic case													
$R^2 = 0.8$	PT_{GMM}	0.301	0.825	0.986	0.297	0.819	0.985	0.758	0.998	1.000	0.735	1.000	1.000
	PT_{EL}	0.399	0.895	0.997	0.390	0.881	0.998	0.790	0.999	1.000	0.777	1.000	1.000
	PT_{ET}	0.392	0.883	0.996	0.381	0.874	0.998	0.791	0.999	1.000	0.778	1.000	1.000
	OT_{GMM}	0.207	0.762	0.979	0.238	0.755	0.989	0.668	0.998	1.000	0.681	0.999	1.000
	OT_{EL}	0.276	0.811	0.986	0.299	0.809	0.991	0.682	0.999	1.000	0.702	0.999	1.000
	OT_{ET}	0.272	0.807	0.986	0.287	0.805	0.991	0.682	0.998	1.000	0.705	0.999	1.000
	Moran	0.282	0.828	0.988	0.253	0.808	0.992	0.746	0.998	1.000	0.737	1.000	1.000
	$Moran_{EL}$	0.308	0.841	0.991	0.280	0.818	0.990	0.753	0.998	1.000	0.744	1.000	1.000
	$Moran_{ET}$	0.315	0.846	0.993	0.279	0.829	0.994	0.752	0.998	1.000	0.745	1.000	1.000
$R^2 = 0.4$	PT_{GMM}	0.147	0.488	0.833	0.176	0.506	0.820	0.362	0.912	0.997	0.432	0.928	1.000
	PT_{EL}	0.207	0.572	0.868	0.236	0.574	0.860	0.382	0.916	0.996	0.438	0.936	1.000
	PT_{ET}	0.188	0.552	0.861	0.232	0.561	0.858	0.394	0.921	0.997	0.442	0.939	1.000
	OT_{GMM}	0.105	0.350	0.679	0.145	0.380	0.672	0.265	0.840	0.991	0.328	0.880	0.999
	OT_{EL}	0.137	0.412	0.728	0.173	0.422	0.731	0.292	0.852	0.990	0.339	0.881	0.999
	OT_{ET}	0.132	0.404	0.718	0.174	0.417	0.725	0.287	0.853	0.991	0.342	0.887	0.999
	Moran	0.048	0.125	0.097	0.025	0.033	0.045	0.259	0.828	0.978	0.214	0.761	0.920
	$Moran_{EL}$	0.060	0.159	0.138	0.040	0.046	0.056	0.271	0.831	0.979	0.227	0.778	0.929
	$Moran_{ET}$	0.061	0.154	0.137	0.038	0.048	0.056	0.269	0.834	0.982	0.228	0.778	0.927
Heteroskedastic case													
$R^2 = 0.8$	PT_{GMM}	0.202	0.545	0.919	0.188	0.553	0.926	0.369	0.908	1.000	0.366	0.914	0.999
	PT_{EL}	0.278	0.590	0.921	0.249	0.617	0.938	0.390	0.910	1.000	0.378	0.911	0.999
	PT_{ET}	0.255	0.595	0.926	0.239	0.610	0.946	0.403	0.919	1.000	0.387	0.922	0.999
	OT_{GMM}	0.136	0.439	0.862	0.137	0.451	0.872	0.307	0.857	0.999	0.273	0.863	0.999
	OT_{EL}	0.174	0.478	0.856	0.172	0.468	0.875	0.310	0.852	0.999	0.277	0.856	0.999
	OT_{ET}	0.171	0.490	0.872	0.166	0.477	0.890	0.317	0.864	0.999	0.288	0.865	0.999
	Moran	0.116	0.382	0.842	0.087	0.393	0.838	0.283	0.846	0.999	0.250	0.843	0.998
	$Moran_{EL}$	0.129	0.400	0.822	0.105	0.404	0.820	0.286	0.837	0.999	0.246	0.834	0.997
	$Moran_{ET}$	0.133	0.414	0.849	0.104	0.422	0.848	0.296	0.850	0.999	0.257	0.847	0.998
$R^2 = 0.4$	PT_{GMM}	0.202	0.516	0.871	0.186	0.495	0.845	0.339	0.891	1.000	0.308	0.892	0.999
	PT_{EL}	0.242	0.515	0.868	0.224	0.500	0.839	0.322	0.878	0.996	0.291	0.872	0.995
	PT_{ET}	0.224	0.509	0.878	0.210	0.501	0.849	0.336	0.889	0.998	0.302	0.882	0.997
	OT_{GMM}	0.116	0.341	0.695	0.111	0.323	0.633	0.244	0.789	0.987	0.221	0.739	0.983
	OT_{EL}	0.145	0.362	0.709	0.141	0.351	0.656	0.226	0.763	0.982	0.222	0.729	0.976
	OT_{ET}	0.142	0.377	0.723	0.129	0.356	0.666	0.233	0.783	0.985	0.223	0.749	0.982
	Moran	0.052	0.232	0.589	0.037	0.165	0.411	0.190	0.753	0.989	0.158	0.690	0.981
	$Moran_{EL}$	0.073	0.276	0.604	0.052	0.192	0.432	0.204	0.747	0.981	0.167	0.687	0.974
	$Moran_{ET}$	0.070	0.276	0.621	0.051	0.196	0.448	0.206	0.759	0.988	0.171	0.699	0.982

"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; and "Moran", "$Moran_{EL}$ " and "$Moran_{EL}$ " denote, respectively, the robust, EL and ET Moran's I tests.

Table 8: Empirical sizes of spatial J tests for the SARAR model (1)

	$n = 144$						$n = 400$					
	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 $\hat{Y}_n = \hat{\kappa}_{1n} W_{1n} Y_n + X_{1n}\hat{\beta}_{1n}$, and "$GMM_2$" uses $\hat{Y}_n = (I_n - \hat{\kappa}_{1n})^{-1} X_{1n}\hat{\beta}_{1n}$. 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 9: Powers of spatial J tests for the SARAR model (1)

	$n=144$						$n=400$					
	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 $\hat{Y}_n = \hat{\kappa}_{1n} W_{1n} Y_n + X_{1n}\hat{\beta}_{1n}$, and "$GMM_2$" uses $\hat{Y}_n = (I_n - \hat{\kappa}_{1n})^{-1} X_{1n}\hat{\beta}_{1n}$. 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.