GMM inference in spatial autoregressive stochastic frontier analysis*

Ming-Yu Deng† and Mingxi Wang

School of International Trade and Economics, University of International Business and Economics, Beijing 100029, China

April 30, 2023

Abstract

This paper mainly considers GMM inference for a spatial autoregressive stochastic frontier (SARSF) model, whose inefficiency term follows a half-normal distribution. Such a model is partially distribution-free, since the distributional assumption regarding the random disturbance is relaxed from normality to symmetry. To investigate the large sample properties, we first centralize the composed error term with nonzero expectation, which aims to establish the cubic and modified linear-quadratic moment functions. Accordingly, we generalize the central limit theorem (Kelejian and Prucha, 2001) designed for linear-quadratic processes to one for cubic forms. As technical inefficiency exists, GMM estimators (GMMEs) of all parameters have the $\sqrt{n}$-rate of convergence and are asymptotically normal under regularity conditions. As technical inefficiency does not exist, the best GMME (BGMME) for the variance of inefficiency has a slower convergence rate than $\sqrt{n}$, and the rest still have the usual $\sqrt{n}$-asymptotics. We also derive a test statistic based on BGMME for whether the inefficiency exists. Monte Carlo simulations suggest that to estimate the SARSF model, BGMME outperforms the corrected 2-stage least square and optimum GMM estimators. Moreover, our proposed BGMME-based test is robust against non-normal disturbance for testing the existence of inefficiency.

Keywords: Stochastic frontier, Spatial autoregression, Inefficiency, GMM estimation, Hypothesis testing

JEL Classification: C12, C13, C21, C51, R32

---

*This work is supported by the National Natural Science Foundation of China (No. 71971062), the Natural Science Foundation of Beijing (No. 9202015), and the Excellent Young Scholar Foundation of UIBE (No. 21JQ01).

†Corresponding author: my_deng@stu.ecnu.edu.cn

1 Introduction

Stochastic frontier (SF) analysis, established almost simultaneously by Aigner et al. (1977) and Meeusen and van den Broeck (1977), has been widely applied to measure productivity. See Fried et al. (2008) for a survey of the literature. In practice, firms usually tend to concentrate in clusters to take advantage of geographical proximity, thereby facilitating imitation and improvement in production (Baptista, 2000; Galli, 2023). Therefore, the spatial lag term of the dependent variable (Cliff and Ord, 1973, 1981, SAR term), which captures spatial correlation across production units, should be introduced in analyzing efficiencies. Otherwise, the omission of the SAR term leads to inconsistent parameter estimators in a cross-sectional or panel data model (Baltagi, 2011, 2013), and accordingly, efficiencies will be misestimated (Glass et al., 2016). Given this, SAR SF (SARSF) analysis, an emerging spatial regression technique in the last decades, has been developed to measure inefficiency/efficiency spillovers in production.[1]

In the early work of Glass et al. (2013, 2014), a distribution-free panel SARSF model is exemplified to calculate time-variant efficiencies. However, Kutlu (2012, 2017) argues that when outliers appear in the data, the distribution-free model would have some robustness issues regarding efficiency estimates. This problem can be solved by modeling the inefficiency as a nonnegative random term (Kutlu, 2022). Therefore, in the distribution-based SARSF model, a specific distribution is assumed as what the inefficiency component follows, which is pioneered by Adetutu et al. (2015) and amplified by Glass et al. (2016).

So far, the distribution-based SARSF model has gradually attracted more attention (Kutlu, 2018; Jin and Lee, 2020; Kutlu et al., 2020; Lai and Tran, 2022; Tran et al., 2023; Kutlu, 2023). Typically, the one-sided (nonnegative) efficiency and two-sided disturbance terms have been carried out following the framework of Aigner et al. (1977). That is, the former intending to capture the effects of purely random statistical noise is supposed to follow a normal distribution, and the latter intending to capture the effects of technical inefficiency is specified to follow a half-normal distribution. SARSF analysis aims to predict efficiency spillovers, which involves the parameter estimates in the two distributions, thereby obtaining the consistent estimators of model parameters is a prerequisite for identifying efficiency spillovers correctly. However, as stated by Glass et al. (2016) and Kutlu (2018), the study on the SARSF models is relatively sparse, especially in terms of econometric theory. Glass et al. (2016) early list the system of score equations for the panel SARSF model, without researching consistency and asymptotic distribution for maximum likelihood (ML) estimation. Next, they do not solve the system but provide a so-called combination of ML strategy. Unfortunately, their strategy does not necessarily solve the score equations, because the intercept estimator (ibid, Eq. (9)) is inconsistent (see footnote 3 later). Jin and Lee (2020) are the first to consider large sample properties of the SARSF model. Based on rigorously theoretical analysis, they establish in spirit the ML framework for estimating such a model, and provide a closed-form of the corrected two-stage ordinary least square (C2SLS) approach.

As the asymptotically efficient estimation, the ML estimator (MLE) is better than the C2SLS estimator (C2SLSE). While in practice, the ML estimation for the SARSF model may suffer from some problems. Firstly, in the generalized setting where the spatial weight matrix does not obsess some special properties like sparseness and symmetry, computation complexities of the MLE for the SARSF model can be overwhelming in the large sample size (Kelejian and Prucha, 1999). Secondly, parameters appear in score equations for the SARSF model with a highly nonlinear form, which further burdens the computation. Finally and notably, in the iteration process of maximizing the related log-likelihood function (or solving such score equations),[2] it is troublesome to solve the ML estimate. This is because the log-likelihood function (or score equations) has at least three stationary points (or solutions, proved in the next section) in most cases. Particularly, iterative methods

[1]Indeed, considering spatial effects in the presence of SF analysis started with Druska and Horrace (2004), where a spatial panel error SF model is developed with fixed effects and time-invariant efficiency. Since then, a large number of theoretical/empirical studies also employ the SF model with spatial dependence in error terms, such as Schmidt et al. (2009), Pavlyuk (2011), Areal et al. (2012), Fusco and Vidoli (2013), Tsionas and Michaelides (2016), Vidoli et al. (2016) and Carvalho (2018), among others.

[2]It is a consensus that solving nonlinear systems (optimal problems) without the explicit expression for its solution depends on iterative methods.

generally rely on selecting the initial iteration value, which indicates that with an improper starting point, the methods converge to some locally optimal solution but the ML estimate.

It is seen from the above discussion that the MLE as efficient is solved cumbersomely, and the C2SLSE as easy to compute is not efficient. This motivates us to find a trade-off estimator, which is not only easier to obtain than the MLE computationally, but also more efficient than the C2SLSE statistically.

In this paper, it is of interest to estimate the SARSF model within the generalized method of moment (GMM) framework, which aims to complement the growing theoretical literature on SARSF analysis. In the context of our GMM, the distributional assumption regarding the disturbance term is relaxed to symmetry around zero, which can avoid the severity of the distributional specification giving rise to the composed error (Schmidt, 1986; Kopp and Mullahy, 1990), and guarantee the variance of the inefficiency term to be identifiable. Therefore, our setting is general and nests the model considered in Adetutu et al. (2015), Glass et al. (2016), Jin and Lee (2020), and Lai and Tran (2022) as a special case with the noise term following a normal distribution. The corresponding methodology is, therefore, partially distribution-free.

Besides such relaxation, our research contributes to the literature regarding GMM's asymptotic theory. First, consisting of efficiency and disturbance terms, the composed error term in the SARSF model, as the main salient feature that differs from the conventional SAR counterpart, possesses a nonzero expectation. This violates the assumption of zero-mean in the GMM setting and leads to an inconsistent GMM estimator (GMME). Thus we modify the composed error term to the central one, whereby a cubic moment is proposed, and the (corrected) linear-quadratic moments (Kelejian and Prucha, 1999; Lee, 2007a) are re-established. Thus, laws of large numbers (LLN) and central limit theorems (CLT) designed for linear-quadratic processes (Kelejian and Prucha, 2001) are not enough. In terms of consistency, the uniform convergence of the objective function accordingly consists of a linear-quadratic part and a cubic counterpart, where the analysis of the former follows Lee (2007a). As for the latter, the proof is provided with the help of the near-epoch dependence (NED) concept in spatial processes and the corresponding LLN in Jenisha and Prucha (2012).

In terms of the asymptotic distribution of GMME, one simultaneously relies on the limiting distribution of moment functions at parameters' true values, and whether the expectation of the gradient matrix of the functions has full rank. Given that such functions refer to cubic forms of the centered error, we generalize the CLT in Kelejian and Prucha (2001) from the linear-quadratic form to the cubic form for deriving the limiting distribution. Notably, it is generally reasonable to assume that the asymptotic variance matrix of moment functions at true parameters is nonsingular. But for the expectation of the gradient matrix, whether it has full rank is up to whether technical inefficiency exists. For the general case where technical inefficiency exists, the asymptotic distribution of the GMME can be derived by our proposed cubic CLT, and the GMME is $\sqrt{n}$-consistent and is distributed as asymptotically normal under some regular assumptions. Ditto for the feasible optimum and best GMMEs. For the irregular case without technical inefficiency, by reparameterizing the model so that the aforementioned gradient matrix has full rank, the asymptotic distribution of BGMME is derived. The BGMME for the variance in the half-normal distribution has a slower rate of convergence than $\sqrt{n}$, and the remaining BGMMEs are still $\sqrt{n}$-consistent. Moreover, a test statistic based on the best GMME is derived for the existence of inefficiency. The inefficiency parameter is nonnegative, so our proposed BGMM-based test is left-sided.

Finally, Monte Carlo experiments investigate finite sample properties of our GMMEs and test statistic. For estimating the SARSF model, the BGMME performs the best by generating a more standard empirical distribution for model parameters than the optimum GMME and C2SLSE. For testing the existence of inefficiency, our BGMM-based test is more robust against non-normal symmetric disturbance than the score test derived by Jin and Lee (2020) and therefore applies to more scenarios.

The remainder of this paper is organized as follows. Section 2 introduces the SARSF model and analyzes theoretically stationary points of the related log-likelihood function. Section 3 proposes the GMM estimation strategy and discusses its consistency. Section 4 proves the asymptotic normality of GMMEs by considering whether technical inefficiency exists. Section 5 derives a test statistic for such existence. Section 6 designs Monte Carlo simulations to investigate the finite sample performance of GMMEs and the test statistic, and Section 7 concludes. Appendix A lists some useful lemmas. Main proofs are collected in Appendix B. Appendix C reports experimental results.

Natation. Throughout this article, $\xrightarrow{p}$ and $\xrightarrow{d}$ signify convergence in probability and distribution, respectively. In addition, $o_p(1)$ and $O_p(1)$ represent successively converging in probability to zero and a bounded constant.

2 SARSF model and stationary points of its log-likelihood function

In this section, we briefly introduce the distribution-based SARSF model, and discuss stationary points of its log-likelihood function in detail. Without loss of generality, consider the half-normal and normal SARSF model for production as shown below (Adetutu et al., 2015; Glass et al., 2016; Jin and Lee, 2020; Lai and Tran, 2022),

$$Y_n = \lambda W_n Y_n + X_n \beta + V_n - U_n, \tag{2.1}$$

where n is the number of producers, $Y_n \in \mathbb{R}^n$ is the vector of observations on the objective variable, λ is a spatial dependence parameter, $W_n \in \mathbb{R}^{n\times n}$ is the row-normalized spatial weight matrix given previously, $X_n \in \mathbb{R}^{n\times k_x}$ (including intercept) is the matrix of observations on regressors with coefficient $\beta \in \mathbb{R}^{k_x}$, $V_n = [v_{n1}, \cdots, v_{nn}]^\top$ ($\top$ denotes the transpose of matrix/vector) is the classical disturbance term, and $U_n = [u_{n1}, \cdots, u_{nn}]^\top$ is the nonnegative inefficiency term.

SARSF analysis aims to measure the efficiency/inefficiency spillovers, which refers to formulating the expectation of u_{ni} conditional on $\epsilon_{ni} = v_{ni} - u_{ni}$, i.e., $\mathrm{E}[u_{ni}|\epsilon_{ni}]$. Therefore, distributional specifications regarding u_{ni} and v_{ni} are dispensable and specified as in Aigner et al. (1977). For each i, v_{ni} follows the normal distribution with mean zero and variances $\sigma^2_{v,o}$, i.e, $v_{ni} \sim \mathcal{N}(0, \sigma^2_{v,o})$, and u_{ni} follows the half-normal distribution with mean zero and variance $\sigma_{u,o}$, i.e., $u_{ni} \sim \mathcal{N}^+(0, \sigma^2_{u,o})$. u_{ni} and v_{ni} are supposed to be independent of each other, and of $\mathbf{x}_{i,n} \in \mathbb{R}^{k_x}$ that is the i^{th} observations on regressors. For each i, $\mathbf{x}^\top_{i,n}$ and $\mathbf{x}_{n,i}$ represent successively the i^{th} row and column vectors of X_n.

For estimating the SARSF model within the ML framework, Jin and Lee (2020) are the first to provide a theoretical foundation. By applying the parametrization in Aigner et al. (1977) that $\sigma^2 = \sigma^2_u + \sigma^2_v, \delta = \sigma_u/\sigma_v$, the log-likelihood function of θ for model (2.1) is derived as

$$\ln L_n(\theta) = n \ln 2 - \frac{n}{2}\ln\left(2\pi\sigma^2\right) + \ln|S_n(\lambda)| - \frac{1}{2\sigma^2}\sum_{i=1}^n \epsilon^2_{ni}(\lambda,\beta) + \sum_{i=1}^n \ln\Phi\left(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\right), \tag{2.2}$$

where $\theta = [\lambda, \beta^\top, \sigma^2, \delta]^\top$, $S_n(\lambda) = I_n - \lambda W_n$ (I_n represents the n-dimensional), $\epsilon_{ni}(\lambda,\beta)$ is the i^{th} component of $\epsilon_n(\lambda,\beta) = S_n(\lambda)Y_n - X_n\beta$, and $\Phi(\cdot)$ is the distribution function of $\mathcal{N}(0,1)$. Meanwhile the first order derivatives of (2.2) with respect to θ are

$$\begin{aligned}
\frac{\partial \ln L_n(\theta)}{\partial\lambda} &= -\mathrm{tr}\left[G_n(\lambda)\right] + \frac{1}{\sigma^2}\sum_{i=1}^n \mathbf{w}_{i,n}Y_n\epsilon_{ni}(\lambda,\beta) + \frac{\delta}{\sigma}\sum_{i=1}^n \mathbf{w}_{i,n}Y_n f\left(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\right),\\
\frac{\partial \ln L_n(\theta)}{\partial\beta} &= \frac{1}{\sigma^2}\sum_{i=1}^n \mathbf{x}_{i,n}\epsilon_{ni}(\lambda,\beta) + \frac{\delta}{\sigma}\sum_{i=1}^n \mathbf{x}_{i,n} f\left(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\right),\\
\frac{\partial \ln L_n(\theta)}{\partial\sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n \epsilon^2_{ni}(\lambda,\beta) + \frac{\delta}{2\sigma^3}\sum_{i=1}^n f\left(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\right)\epsilon_{ni}(\lambda,\beta),\\
\frac{\partial \ln L_n(\theta)}{\partial\delta} &= -\frac{1}{\sigma}\sum_{i=1}^n f\left(-\frac{\delta}{\sigma}\epsilon_{ni}(\lambda,\beta)\right)\epsilon_{ni}(\lambda,\beta),
\end{aligned} \tag{2.3}$$

where $\mathbf{w}_{i,n}$ is the i^{th} row of W_n, $G_n(\lambda) = W_n S_n^{-1}(\lambda)$, and $f(\cdot) = \phi(\cdot)/\Phi(\cdot)$ with $\phi(\cdot)$ being the density function of $\mathcal{N}(0,1)$. Let $\theta_o = [\lambda_o, \beta_o^\top, \delta_o^2, \sigma_o^2]$ be the true value of θ, and denote $S_n = S_n(\lambda_o)$ and $G_n = G_n(\lambda_o)$. Jin and Lee (2020, Proposition 2.1) have proved that the MLE $\hat{\theta}_{m,n}$ that solves the score system $\partial \ln L_n(\theta)/\partial\theta = 0$ is consistent under some reasonable assumptions. However, in practice, these score equations are very cumbersome to solve, because parameters appear in (2.3) with a highly nonlinear form so that neither analytical expression nor related toolbox is available for the MLE. Besides, we find more than one stationary point exists in (2.2), i.e., more than one root can solve the nonlinear score equations. We now prove that this conclusion holds as the technical inefficiency always exists, i.e., $\sigma_{u,o} > 0$ strictly holds.

Firstly and apparently, the global maximum point of (2.2) is $\hat{\theta}_{m,n}$, where $\hat{\delta}_{m,n}$ should be positive because of the existence of inefficiency.

Secondly, consider the MLE of the SAR model $Y_n = \lambda W_n Y_n + X_n\beta + V_n$, say $\hat{\lambda}_{m-sar,n}$, $\hat{\beta}_{m-sar,n}$, and $\hat{\sigma}^2_{m-sar,n}$. In conjunction with $\hat{\delta}_n = 0$, $\theta_{1,*} = [\hat{\lambda}_{m-sar,n}, \hat{\beta}^\top_{m-sar,n}, \hat{\delta}_n, \hat{\sigma}^2_{m-sar,n}]^\top$ can also set scores in (2.3) to zero. That the score vector with respect to δ in (2.3) is equal to zero follows from $\partial \ln L_n(\theta_{1,*})/\partial\delta = -\sqrt{\frac{2}{\pi}}\hat{\sigma}_{m-sar,n}[\partial \ln L_n(\theta_{1,*})/\partial\beta_1] = 0$. Indeed, the SARSF model degenerates to the classical SAR model as $\sigma_u = \delta = 0$. If the model specification in (2.1) is right, then $\hat{\lambda}_{m-sar,n}$, $\hat{\beta}_{m-sar,n}$, and $\hat{\sigma}^2_{m-sar,n}$ are the quasi-MLE of the modified SARSF model $Y_n = \lambda W_n Y_n + (\beta_1 - \sqrt{2/\pi}\sigma_u)\mathbf{1}_n + X_{-1,n}\beta_{-1} + (V_n - U_n + \sqrt{2/\pi}\sigma_u\mathbf{1}_n)$, where β_{-1} and $X_{-1,n}$ are the remaining parts of β losing β_1 and X_n losing $\mathbf{x}_{n,1} = \mathbf{1}_n$, respectively. So $\hat{\beta}_{1,m-ssf,n}$ is consistent for $\beta_{1,o} - \sqrt{2/\pi}\sigma_{u,o}$ rather than $\beta_{1,o}$,[3] and $\hat{\sigma}^2_{m-sar,n}$ is consistent for $\mathrm{Var}[\epsilon_{ni} + \sqrt{2/\pi}\sigma_{u,o}] = \sigma^2_{u,o} + (1-2/\pi)\sigma^2_{v,o}$ rather than $\sigma^2_{u,o} + \sigma^2_{v,o}$. Thus, $\theta_{1,*}$ is different from $\hat{\theta}_{m,n}$.

To be more precise, $\theta_{1,*}$ is not only a stationary point of (2.2), but also is a saddle point. That is to say, change on $\theta_{1,*}$ following some direction $\mathbf{d}$ will leads to a continuing increscent of (2.2), and we now prove that such direction $\mathbf{d}$ is the eigenvector of the empirical Hessian matrix of (2.2) at $\theta_{1,*}$, say $H_n(\theta_{1,*})$, with the corresponding eigenvalue being 0. By Lemma A.1, we know $\mathbf{d} = [0, \sqrt{2/\pi}\hat{s}_n, \mathbf{0}^\top_{k_x-1}, 1, 0]^\top$, where $\hat{s}_n = (\frac{1}{n}\mathbf{e}^\top_{1,*}\mathbf{e}_{1,*})^{1/2}$ as the sample standard deviation of the residuals $\mathbf{e}_{1,*} = S_n(\hat{\lambda}_{m-sar,n})Y_n - X_n\hat{\beta}^\top_{m-sar,n}$, and $\mathbf{0}_{k_x-1}$ is the (k_x-1)-dimensional zero-vector. We are interested then, in the sign of $\ln L_n(\theta_{1,*}+\alpha\mathbf{d}) - \ln L_n(\theta_{1,*})$ for arbitrarily small, positive α. By the third-order Taylor expansion of multi-variable function,

$$\begin{aligned}\ln L_n(\theta_{1,*}+\alpha\mathbf{d}) - \ln L_n(\theta_{1,*}) &= \frac{\partial \ln L_n(\theta_{1,*})}{\partial\theta} + H_n(\theta_{1,*})\mathbf{d} + \frac{\alpha^3}{6}\left[\frac{\partial^3 \ln L_n(\theta_{1,*})}{\partial\theta^3}\right]\mathbf{d}^3 + o(\|\alpha\mathbf{d}\|^3) \\ &= \frac{\alpha^3}{6}\left[\frac{\partial^3 \ln L_n(\theta_{1,*})}{\partial\theta^3}\right]\mathbf{d}^3 + o(\|\alpha\mathbf{d}\|^3)\end{aligned} \tag{2.4}$$

since $\partial \ln L_n(\theta_{1,*})/\partial\theta = 0$ and $H_n(\theta_{1,*})\mathbf{d} = 0$ by Lemma A.1.[4] By considering whether elements of $\mathbf{d}$ are equal to zero, we only need to calculate $\partial^3 \ln L_n(\theta_{1,*})/\partial\beta_1^3$, $\partial^3 \ln L_n(\theta_{1,*})/\partial\delta^3$, $\partial^3 \ln L_n(\theta_{1,*})/\partial\delta\partial\beta_1^2$ and $\partial^3 \ln L_n(\theta_{1,*})/\partial\delta^2\partial\beta_1$. By direct calculation,

$$\frac{\partial^3 \ln L_n(\theta_{1,*})}{\partial\beta_1^3} = \frac{1}{\sigma^3}\sqrt{\frac{2}{\pi}}\left(1-\frac{4}{\pi}\right)\sum_{i=1}^n e^3_{ni} \text{ and } \frac{\partial^3 \ln L_n(\theta_{1,*})}{\partial\delta^3} = \frac{\partial^3 \ln L_n(\theta_{1,*})}{\partial\delta\partial\beta_1^2} = \frac{\partial^3 \ln L_n(\theta_{1,*})}{\partial\delta^2\partial\beta_1} = 0,$$

where e_{ni}'s are elements of the residual vector $\mathbf{e}_{1,*}$. Generally speaking, there should have $\sum_{i=1}^n e^3_{ni} < 0$.[5] Then (2.4) simplifies to

$$\ln L_n(\theta_{1,*}+\alpha\mathbf{d}) - \ln L_n(\theta_{1,*}) = \frac{\alpha^3}{6\sigma^3}\sqrt{\frac{2}{\pi}}\left(1-\frac{4}{\pi}\right)\sum_{i=1}^n e^3_{ni} + o(\|\alpha\mathbf{d}\|^3), \tag{2.5}$$

where the first term on the right-hand side of (2.5) is the leading term which decides the sign of (2.4), and the

[3] This is why the intercept estimator Glass et al. (2016, Eq. (9)) is inconsistent.

[4] $\partial^3 \ln L_n(\theta_{1,*})/\partial\theta\partial\theta^\top$ is an order-3 (k_x+3)-dimensional tensor. For any compatible (k_x+3)-dimensional tensor $\mathcal{A}$ of order-3, and $\mathbf{x} \in \mathbb{R}^{k_x+3}$, $\mathcal{A}\mathbf{x}^3 = \sum_{i,j,l=1}^{k_x+3}\mathcal{A}_{ijl}x_i x_k x_l$.

[5] $\sum_{i=1}^n e^3_{ni}$ may be positive in practice for a finite (small) sample size, even if $\sigma_{u,o} > 0$. This is the "wrong skew" problem (Olson et al., 1980; Waldman, 1982). As the sample size tends to infinity, the problem fades away gradually.

remainder $o(\|\alpha\mathbf{d}\|^3)$ is not enough to influence such sign. The sign should be positive as $\sum_{i=1}^n e_{ni}^3 < 0$. This indicates the likelihood (2.2) will be increased by moving away from $\theta_{1,*}$ in the direction $\mathbf{d}$. By the compactness assumption for the parameter space, there exists the third point, say $\theta_{2,*}$, such that the scores (2.3) can be set to zero. We can see from $\theta_{2,*} = \theta_{1,*} + \alpha\mathbf{d}$ that the last component of $\theta_{2,*}$ is consistent for $\sigma_{u,o}^2 + (1-2/\pi)\sigma_{v,o}^2$, and that of $\hat{\theta}_{m,n}$ is consistent for $\sigma_{u,o}^2 + \sigma_{v,o}^2$. So, $\theta_{2,*}$ and $\hat{\theta}_{m,n}$ are different.

In summary, besides the computational burden it entails, the log-likelihood function (score equations) has at least 3 local maximum points (zero points) if the technical inefficiency always exists.[6] Thus, the MLE is asymptotically efficient, but how to get its value is still troublesome. In view of this, we consider the GMM estimation for model (2.1).

3 GMM estimation and consistency

The SARSF model is supposed to be an equilibrium model, and the structural equation (2.1) implies the corresponding reduced-form $Y_n = S_n^{-1}X_n\beta_o - S_n^{-1}U_n + S_n^{-1}V_n$. It follows that $W_nY_n = G_nX_n\beta_o - G_nU_n + G_nV_n$ and the SAR term W_nY_n is correlated with (U_n, V_n), because in general $\mathrm{E}[(W_nY_n)^\top V_n] = \sigma_{v,o}^2\mathrm{tr}(G_n) \neq 0$ and so does $\mathrm{E}[(W_nY_n)^\top U_n]$. This indicates that the SAR term is endogenous.

3.1 GMM estimation

Let $Q_n \in \mathbb{R}^{n\times k_q}$ be the IV matrix. Denote $\mathcal{P}_{1n}$ and $\mathcal{P}_{2n}$ as the family of constant $n\times n$ matrices with trace being zero and with diagonal elements being zero, respectively. By Lemma A.3, $\mathrm{E}[\epsilon_{ni}] = -\sqrt{2/\pi}\sigma_{u,o} \neq 0$ unless $\sigma_{u,o} = 0$, which implies the classical linear moment function $Q_n^\top \epsilon_n(\lambda,\beta)$ is no longer applicable. Similarly, since $\epsilon_n^\top P_{nj}\epsilon_n \neq \mathrm{Var}[\epsilon_{ni}]\mathrm{tr}(P_{nj}) = 0$ for $j = 1,\cdots,k_p$, the quadratic moment functions $\epsilon_n^\top(\lambda,\beta)P_{nj}\epsilon_n(\lambda,\beta)$'s suggested by Lee (2001, 2007a) also should not be applicable, where the matrices P_{nj}'s are selected from $\mathcal{P}_{1n}$ or $\mathcal{P}_{2n}$. Both problems are caused by the violation of the mean-0 assumption regarding the composed error term ϵ_n.

To overcome such problems, we modify the composed error ϵ_{ni} to the centered error $\epsilon_{ni}^\dagger = \epsilon_{ni} + \sqrt{2/\pi}\sigma_{u,o}$ so that $\mathrm{E}[\epsilon_{ni}^\dagger] = 0$. To streamline the following discussion, denote $\mu_s^\dagger = \mathrm{E}[\epsilon_{ni}^{\dagger s}]$ for $s = 1, 2\cdots, 6$. Here we list the first six moments of $\epsilon_{ni}^\dagger$ in Lemma A.3. Accordingly, the modified reduced form of vector Y_n is determined by the system as $Y_n = S_n^{-1}(\beta_{1,o} - \sqrt{2/\pi}\sigma_{u,o})\mathbf{1}_n + S_n^{-1}X_{-1}\beta_{-1,o} + S_n^{-1}\epsilon_n^\dagger$, where $\beta_{1,o}$ and $\beta_{-1,o}$ are the trues of β_1 and β_{-1}. Note that the parameters $\beta_{1,o}$ and $\sigma_{u,o}$ aries with the form $\beta_{1,o} - \sqrt{2/\pi}\sigma_{u,o}$, so we employ the parameterization, $\beta_1^\dagger = \beta_1 + \sqrt{2/\pi}\sigma_u$ and $\sigma_u = \sigma_u$, to make analysis more efficient. Here such parameterization is a one-to-one transformation from (β_1, σ_u) to $(\beta_1^\dagger, \sigma_u)$. Accordingly, the true value is $\beta_{1,o}^\dagger = \beta_{1,o} + \sqrt{2/\pi}\sigma_{u,o}$. Denote $\beta^\dagger = [\beta_1^\dagger, \beta_{-1}^\top]^\top$, $\kappa = [\lambda, \beta^{\dagger\top}]^\top$ and $\theta^\dagger = [\kappa^\top, \sigma_u]^\top$, and their corresponding trues are $\beta_o^\dagger$, κ_o, and $\theta_o^\dagger$. Then, the modified equilibrium vector Y_n is also given by

$$Y_n = S_n^{-1}X_n\beta_o^\dagger + S_n^{-1}\epsilon_n^\dagger. \tag{3.1}$$

Within the GMM framework, the correct linear and quadratic moment functions are proposed as

$$g_{n1}(\kappa) = \left[\epsilon_n^{\dagger\top}(\kappa)P_{n1}\epsilon_n^\dagger(\kappa), \cdots, \epsilon_n^{\dagger\top}(\kappa)P_{nk_p}\epsilon_n^\dagger(\kappa), \epsilon_n^{\dagger\top}(\kappa)Q_n\right]^\top, \tag{3.2}$$

where $\epsilon_n^\dagger(\kappa) = S_n(\lambda)Y_n - X_n\beta^\dagger$. Note that the parameters $\beta_{1,o}$ and $\sigma_{u,o}$ aries in (3.1) with the form $\beta_{1,o}^\dagger = \beta_{1,o} - \sqrt{2/\pi}\sigma_{u,o}$. Thus, for the identification of $\sigma_{u,o}$, another scalar moment function by Lemma A.3 is considered as follows,

$$g_{n2}(\theta^\dagger) = \sum_{i=1}^n \epsilon_{ni}^{\dagger 3}(\kappa) - n\cdot\sqrt{\frac{2}{\pi}}\left(1-\frac{4}{\pi}\right)\sigma_u^3. \tag{3.3}$$

[6]This phenomenon also appears in the ML estimation of the SF model Waldman (1982).

By taking (3.2) and (3.3) together, it yields the GMM estimator (GMME) with the moment vector $g_n(\theta^\dagger) = [g_{n1}^\top(\kappa), g_{n2}(\theta^\dagger)]^\top$,

$$\hat{\theta}^\dagger_{g,n} = \arg\min_{\theta^\dagger \in \Theta} g_n^\top(\theta^\dagger) a_n^\top a_n g_n(\theta^\dagger), \tag{3.4}$$

where Θ is the (k_x+2)-dimensional parameter space of $\theta^\dagger$, and $a_n^\top a_n$ is a $(k_q+k_p+1)\times(k_q+k_p+1)$ distance (weighting) matrix. From the perspective of computation complexity caused by finding the GMME $\hat{\theta}^\dagger_{g,n}$ of the SARSF model (3.1), $\theta^\dagger$ appears in the moment functions $g_n(\theta^\dagger)$ with cubic form, so the objective function should be multinomial in $\theta^\dagger$ of order 6. The order is slightly larger than that of the objective function for the SAR model, and the latter is 4. This indicates that obtaining the GMME for the SARSF model is slightly more complicated than for the SAR model. Fortunately, finding the GMM estimate of the SARSF model will refer to solving multinomial equations (gradient) of order 5. The coefficients in such equations do not rely on $\theta^\dagger$ and therefore need to be calculated once, on which the evaluation of the objective function can be easily computed at different values of $\theta^\dagger$ based. These show that for estimating the SARSF model, the GMM estimation is easier than the cumbersome ML approach computationally.

For subsequent analysis, some basic model specifications are necessary.

Assumption 1. Individual units in the economy are located or living in a region $D_n \subset D \subset \mathbb{R}^d$, where D is a (possibly) unevenly spaced lattice, the cardinality $|D_n|$ of a finite set D_n satisfies $\lim_{n\to\infty}|D_n| = \infty$. The distance $d(i,j)$ between any two different individuals i and j is larger than or equal to a positive constant, which will be assumed to be 1 for convenience.

Assumption 2. v_{ni}'s follow an independent, identical, and symmetric distribution with mean zero and variance $\sigma^2_{v,o}$. Moreover, $\mathrm{E}[|v_{ni}|^{6+\iota}] < \infty$ for some $\iota > 0$.

Assumption 3. (i) $u_{ni} \sim \mathcal{N}^+(0, \sigma^2_{u,o})$. (ii) u_{ni} and v_{ni} are independent of each other, and of $\mathbf{x}_{i,n}$.

Assumption 4. The elements of X_n are uniformly bounded constants, X_n has full rank, and the matrix limit $\lim_{n\to\infty}\frac{1}{n}X_n^\top X_n$ exists and is nonsingular.

Assumption 5. The spatial weight matrix W_n and S_n^{-1} are uniformly bounded in both row and column sums in absolute value.

Assumption 6. (i) $c_1 \equiv \lambda_m \sup_n \|W_n\|_\infty < 1$, and $[-\lambda_m, \lambda_m]$ is the compact parameter space of λ on the real line. (ii) The parameter space of $(\beta^\dagger, \sigma_u)$ are a compact subset of $\mathbb{R}^{k_x+1}$. Here $\|\cdot\|_\infty$ denotes the row sum matrix norm.

Owning to the cubic moment function in (3.3), besides the LLN for linear-quadratic forms of disturbances in Kelejian and Prucha (2001), we also require the LLN (Jenisha and Prucha, 2012, Prop. 1) for spatial mixing and NED processes to analyze the uniform convergence of (3.3). This LLN is generalized from the time series literature (Some related definitions are given in Appendix A). Assumption 1, established by Jenisha and Prucha (2009, 2012), maintains some conditions for the LLN. For example, it allows the spaced lattice to be high-dimensional as a subset of $\mathbb{R}^d$, because the distance between any two spatial units can be a geometrical distance, an economic distance, or a mixture of both. The specification of increasing domain asymptotics in Assumption 1 is natural in a regional study, and essentially it rules out the scenario of infilled asymptotic, under which even some popular estimators like the least squares and the method of moments may not be consistent (Lahiri, 1996). Assumption 2 relaxes the distributional assumption regarding ϵ_n from normality to symmetry around zero (Kopp and Mullahy, 1990), which aims to make the calculation of the third moment $\mathrm{E}[\epsilon^\dagger_{ni}]$ irrelevant to the distributional parameter of v_{ni}. This is because the symmetry also implies $\mathrm{E}[\epsilon^3_{ni}] = 0$, and then the calculation of $\mathrm{E}[\epsilon^{\dagger 3}_{ni}] = \mathrm{E}[(u_{ni} - \mathrm{E}[u_{ni}])^3]$ does not involve the parameter in the distribution of V_n. The existence of the sixth moment of v_{ni}, in conjunction with Assumption 3(i), implies the existence of $\mu^\dagger_6$, which guarantees finite variance of the cubic moment $g_{n2}(\theta^\dagger_o)$ in asymptotic distribution.

Assumptions 2-3 still summarize the exogeneity of explanatory variables. In Assumption 4, the nonstochasticity and uniform boundedness are imposed on X_n for simplicity, as in Lee (2004, 2007a). The uniform boundedness of W_n and S_n^{-1} in Assumption 5, originated in the series of work Kelejian and Prucha (1998, 1999, 2010), is a standard condition in spatial econometric literature for restricting the spatial dependence across production units to a manageable degree. The compactness condition of parameter space in Assumption 6(i) is basic for the theory of extremum estimators. Here the specification for the parameter space of λ_o in Assumption 6(ii), originated in Xu and Lee (2015), implies the existence of the modified reduced-form of Y_n in (3.1) and the Neuman series expansion $(I_n - \lambda W_n)^{-1} = \sum_{i=0}^{\infty}(\lambda W_n)^i$ for any $\lambda \in [-\lambda_m, \lambda_m]$. If the compactness in Assumption 6 is relaxed, then the parameter space should be convex, and the objective function should be concave (Engle and McFadden, 1994, p. 2111–2245). But the objective function in (3.4), as a multinomial in $\theta^\dagger$ of order 6, is not concave fully, even via some parametric transformations.

3.2 Consistency

For the uniform convergence of (3.2) in the consistency of the GMME $\hat{\theta}^\dagger_{g,n}$, we summarize the following regularity conditions. Let $A_n^s = A_n^\top + A_n$ for any square matrix A_n.

Assumption 7. The matrices P_{nj}'s from $\mathcal{P}_{1n}$ are uniformly bounded in both row and column sums in absolute value, and elements of the IV Q_n are uniformly bounded.

Assumption 8. Either the following (i) or (ii) holds.

(i) $\lim_{n\to\infty} \frac{1}{n} Q_n^\top [G_n X_n \beta_o^\dagger, X_n]$ has the full rank $k_x + 1$.

(ii) $\lim_{n\to\infty} \frac{1}{n} Q_n^\top X_n$ has the full rank k_x, $\lim_{n\to\infty} \frac{1}{n}\mathrm{tr}(P_{jn}^s G_n) \neq 0$ for some j, and $\lim_{n\to\infty} \frac{1}{n}[\mathrm{tr}(P_{n1}^s G_n), \cdots, \mathrm{tr}(P_{nk_p}^s G_n)]$ is linearly independent of $\lim_{n\to\infty} \frac{1}{n}\left[\mathrm{tr}(G_n^\top P_{n1} G_n), \cdots, \mathrm{tr}(G_n^\top P_{nk_p} G_n)\right]$ holds.

Conditions on P_{nj}'s and Q_n in Assumption 7 are for analytic tractability. Assumption 8, as a modified version of Lee (2007a, Assumption 5),[7] plays an important role in the identification condition of model (2.1). Here Assumption 8(i) is decided by the characteristic $\mathrm{E}[\epsilon_n] \neq 0$ for the SARSF model. In the case of the 8(i), the linear moment equation $\lim_{n\to\infty} \frac{1}{n}\mathrm{E}[Q_n \top \epsilon_n^\dagger(\kappa)] = 0$ is uniquely zero at $\kappa = \kappa_o$. However, as linear correlation between $G_n X_n \beta_o^\dagger$ and X_n exists, which indicates the spatial lag parameter λ_o cannot be identified from such equation, 8(ii) also assures the identification of λ_o from the quadratic moment equation $\lim_{n\to\infty} \frac{1}{n}\mathrm{E}[\epsilon_n^{\dagger\top}(\kappa) P_{nj} \epsilon_n^\dagger(\kappa)] = 0$ for $j = 1, \cdots, k_p$. Here more details can be found in Lee (2001).

For the uniform convergence of (3.3) in the consistency of $\hat{\theta}^\dagger_{g,n}$, we introduce other basic regularity assumptions.

Assumption 9. At least one of the following two conditions (a) and (b) is satisfied:

(a) Only individuals whose distances are less than or equal to some positive constant d_o may affect each other directly, i.e., $w_{n,ij} \neq 0$ only if $d(i,j) \leq d_o$.

(b) (i) For every n, the number of columns $\mathbf{w}_{n,j}$ of W_n with $|\lambda_o| \sum_{i=1}^n |w_{n,ij}| > c_1$ is less than or equal to some fixed nonnegative integer that does not depend on n, where $c_1 \in (0,1)$ is defined in Assumption 6(i). (ii) there exists an $\alpha > d$ and a constant c_2 such that $|w_{n,ij}| \leq c_2/d(i,j)^\alpha$.

Assumption 10. *(i)* $\sup_{i,n} \|\mathbf{x}_{i,n}\|_{4+\iota} < \infty$ *for some* $\iota > 0$. *(ii)* $\{\mathbf{x}_{i,n}\}_{i=1}^n$ *is* α*-mixing random field with* α*-mixing coefficient* $\alpha(u,v,s) \leq (u+v)^{c_3}\hat{\alpha}(s)$ *for some* $c_3 > 0$*, where* $\hat{\alpha}(s)$ *satisfies* $\sum_{s=1}^\infty s^{d-1}\hat{\alpha}(s) < \infty$*.*

Assumption 9 considers two different kinds of the weights matrix (Xu and Lee, 2015).[8] Assumption 9(a) only allows direct interaction for any two units when their distance is not bigger than the specific number d_o.

[7] Lee (2007a, Assumption 5), originated in Lee (2001), has been widely considered in theoretical literature of spatial econometrics, such as Lee (2007b), Liu et al. (2010), Lin and Lee (2010), Su (2012), Jin and Wang (2021) and Lee et al. (2022), among others.

[8] This assumption is standard in spatial NED processes and has been gradually considered in research on nonlinear spatial models, such as Xu and Lee (2015), Xu and Lee (2018), Liu and Lee (2019), Jin and Lee (2020), and Jin and Wang (2023), among others.

For any different units i and j, Assumption 9(b)(ii) requires their interactions to decline geometrically fast as the interactions exist. Moreover, by imposing a constraint on the column sums of W_n in absolute value, Assumption 9(b)(i) guarantees that the number of spatial units of having large aggregated effects on other units is fixed. Assumption 10(i) is equivalent to $\sup_{1\leq k\leq k_x,i,n}|x_{ik,n}|^{4+\iota}<\infty$ for some $\iota>0$, which further strengthen regressors X_n for the NED properties. The mixing coefficient for the random field $\{\mathbf{x}_{i,n}\}$ in Assumption 10(ii) does not only depend on the distance between two separate subsets of spatial units but also their sizes (see Jenisha and Prucha (2012) for details). Actually, Assumptions 9–10 can help us establish NED properties of the dependent variable y_{ni} so that the LLN in Jenisha and Prucha (2012) can be applicable (Xu and Lee, 2015, Lemma 1). Then we can further establish the NED properties of $\mathbf{w}_{i,n}Y_n$ and $\epsilon^{\dagger}_{ni}(\kappa)=y_{ni}-\lambda\mathbf{w}_{i,n}Y_n-\mathbf{x}^{\top}_{i,n}\beta^{\dagger}$ as well as their quadratic and cubic forms.

In the GMM setting of Hansen (1982), the selection of the optimal distance matrix requires that equation $\lim_{n\to\infty}\frac{1}{n}a_n\mathrm{E}[g_n(\theta^{\dagger})]=0$ is zero uniquely at $\theta^{\dagger}_o$, which implies that a_n converges to some constant full rank matrix.

Assumption 11. $a_0=\lim_{n\to\infty}a_n$ exists and has full column rank k_q+k_p+1.

Working with Assumption 8, Assumption 11 can help us prove the existence and uniqueness of the solution to the aforementioned moment equation.

Proposition 1. *Under Assumptions 8 and 11, $\theta^{\dagger}_o$ uniquely solves the equation $a_o\lim_{n\to\infty}\frac{1}{n}\mathrm{E}[g_n(\theta^{\dagger})]=0$.*

Based on Proposition 1, we can establish the consistency of $\hat{\theta}^{\dagger}_{g,n}$ defined by (3.4) in the following proposition. Meanwhile consistent estimators for $\beta_{1,o}$ and $\sigma^2_{v,o}$ are suggested.

Proposition 2. *Under Assumption 1–11, the GMME $\hat{\theta}^{\dagger}_{g,n}$ is consistent. Especially, $\hat{\beta}^{\dagger}_{1,g,n}+\sqrt{2/\pi}\hat{\sigma}_{u,g,n}\xrightarrow{p}\beta_{1,o}$, and $\frac{1}{n}\mathbf{e}^{\top}_{g,n}\mathbf{e}_{g,n}-(1-2/\pi)\hat{\sigma}^2_{u,g,n}\xrightarrow{p}\sigma^2_{v,o}$, where $\mathbf{e}_{g,n}=S_n(\hat{\lambda}_{g,n})-X_n\hat{\beta}_{g,n}$ is the generated residual vector.*

We can see from Proposition 2 that except $\hat{\theta}^{\dagger}_{g,n}$, estimators for $\beta_{1,o}$ and $\sigma^2_{v,o}$ are indirectly obtained, rather than directly minimizing (3.4).

4 Asymptotic normality

As the usual setting in the asymptotic distribution of extreme estimators, we also consider the true value $\theta^{\dagger}_o$ as an interior point in its parameter space Θ.

Assumption 12. Θ is a compact subset of $\mathbb{R}^{k_x+2}$. (i) If $\sigma_{u,o}>0$, $\theta^{\dagger}_o\in\mathrm{Int}(\Theta)$. (ii) If $\sigma_{u,o}=0$, the true values of the remaining parameters are still in the interior of the corresponding parameter space.

As is known to all, the asymptotic distribution of the GMME $\hat{\theta}^{\dagger}_{g,n}$ simultaneously refers to that of the random vector $(1/\sqrt{n})g_n(\theta^{\dagger}_o)$, and whether the expectation of gradient matrix $\partial g_n(\theta^{\dagger}_o)/\partial\theta^{\dagger}$ has full rank. We now discuss the former. Consider

$$\frac{1}{\sqrt{n}}g_n(\theta^{\dagger}_o)=\frac{1}{\sqrt{n}}\left[\epsilon^{\dagger\top}_n P_{n1}\epsilon^{\dagger}_n,\cdots,\epsilon^{\dagger\top}_n P_{nk_p}\epsilon^{\dagger}_n,\epsilon^{\dagger\top}_n Q_n,g_{2n}(\theta^{\dagger}_o)\right]^{\top}.$$

Proof for multivariate normality needs the Cramér–Wold device. Let $\mathbf{c}_n=[c_1,c_2,\cdots,c_{k_p},\mathbf{c}^{\top}_{k_q},c_{pq}]^{\top}$ be an arbitrary (k_q+k_p+1)-dimensional constant vector with c_j, $j=1,\cdots,k_p$, and c_{pq} being scalars and $\mathbf{c}_{k_q}$ being a k_q-dimensional column subvector. Then, it is equivalent to consider the asymptotic distribution of

$$\frac{1}{\sqrt{n}}\mathbf{c}^{\top}_n g_n(\theta^{\dagger}_o)=\frac{1}{\sqrt{n}}\left\{c_{pq}\left[\sum_{i=1}^{n}\epsilon^{\dagger 3}_{ni}-n\sqrt{\frac{2}{\pi}}(1-\frac{4}{\pi})\sigma^3_{u,o}\right]+\epsilon^{\dagger\top}_n\left(\sum_{j=1}^{k_p}c_jP_{nj}\right)\epsilon^{\dagger}_n+\mathbf{c}^{\top}_{k_q}Q^{\top}_n\epsilon^{\dagger}_n\right\},$$

which is cubic functions in $\epsilon_{ni}^{\dagger}$ rather than quadratic forms so that the CLT in Kelejian and Prucha (2001, Thm. 1) is no longer applicable. An alternative is to apply the CLT in Jenisha and Prucha (2012), but this must further reinforce conditions on the α-mixing random field $\{\mathbf{x}_{i,n}\}_{i=1}^{n}$ and its α-mixing coefficient in Assumption 10, such as Xu and Lee (2015), Xu and Lee (2018), Liu and Lee (2019), Jin and Lee (2020), and Jin and Wang (2023). Here we do not follow them but generalize the CLT in Kelejian and Prucha (2001, Thm. 1) from the case of quadratic forms to the cubic case.

Theorem 1. *Define* $F_n = \sum_{i=1}^{n} c_{ni}\varepsilon_{ni}^3 + \varepsilon_n^\top A_n \varepsilon_n + \mathbf{b}_n^\top \varepsilon_n$, *where* $\varepsilon_n = [\varepsilon_{n1}, \cdots, \varepsilon_{nn}]^\top$ *is a random vector,* c_{ni}*'s are scalars for* $i = 1, \cdots, n$, A_n *is a* $n \times n$ *matrix with diagonal elements being zero, and* $\mathbf{b}_n = [b_{n1}, \cdots, b_{nn}]^\top$ *is a n-dimensional column vector. Suppose that*

(i) ε_{ni}*'s are totally i.i.d., and* $\mathrm{E}[\varepsilon_{ni}] = 0$ *for each* i. *Moreover,* $\sup_n \mathrm{E}[|\varepsilon_{ni}|^{6+\eta_1}] < \infty$ *for some* $\eta_1 > 0$.

(ii) $\sup_{j,n} \sum_{i=1}^{n} |a_{ij,n}| < \infty$.

(iii) $\sup_n \frac{1}{n} \sum_{i=1}^{n} |b_{ni}|^{2+\eta_2} < \infty$ *for some* $\eta_2 > 0$, *and* $\sup_n \frac{1}{n}\sum_{i=1}^{n} |c_{ni}|^{2+\eta_3} < \infty$ *for some* $\eta_3 > 0$.

Then, $(F_n - \mu_{F,n})/\sigma_{F,n} \xrightarrow{d} \mathcal{N}(0,1)$ *provided that* $\frac{1}{n}\sigma_{F,n}^2 \geq \nu$ *for some* $\nu > 0$, *where* $\mu_{F,n} = \mu_{\varepsilon,3}\sum_{i=1}^{n} c_{ni}$, *and*

$$\sigma_{F,n}^2 = \sum_{i=1}^{n} c_{ni}^2(\mu_{\varepsilon,6} - \mu_{\varepsilon,3}^2) + 4\sigma_\varepsilon^4 \sum_{i=1}^{n}\sum_{j=1}^{i-1} a_{ij,n}^2 + \sigma_\varepsilon^2 \sum_{i=1}^{n} b_{ni}^2 + 2\mu_{\varepsilon,4}\sum_{i=1}^{n} b_{ni}$$

with $\sigma_\varepsilon^2 = \mathrm{E}[\varepsilon_{ni}^2]$ *and* $\mu_{\varepsilon,s} = \mathrm{E}[\varepsilon_{ni}^s]$ *for* $s = 3, 4, 6$.

Remark. Kelejian and Prucha (2001, Thm. 1) allow the off-diagonal elements of A_n to be nonzero, and here we set $a_{ii,n} = 0$ for simplicity. Given the cubic form of ε_{ni} in the expression of F_n, we strengthen the moment condition on ε_{ni} that a moment of order higher than the sixth exists. Here Conditions (ii)–(iii), also imposed in such Thm. 1, are standard.

With the help of Theorem 1, we can derive the asymptotic distribution of $\hat{\theta}_{g,n}^{\dagger}$. It is then of interest to discuss the asymptotic variance Asy.var$[g_n(\theta_o^\dagger)]$, say Ω_n. Let $\mathrm{vec}_d(A_n) = [a_{11,n}, \cdots, a_{nn,n}]^\top$ denote the column vector generated by the diagonal elements of $A_n \in \mathbb{R}^{n\times n}$. Then, by Lemma A.4

$$\Omega_n = \begin{bmatrix} (\mu_4^\dagger - 3\mu_2^{\dagger 2})\omega_{nd}^\top \omega_{nd} & \mu_3^\dagger \omega_{nd}^\top Q_n & 0 \\ \mu_3^\dagger Q_n^\top \omega_{nd} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \Gamma_n, \tag{4.1}$$

where $\omega_{nd} = [\mathrm{vec}_d(P_{n1}), \cdots, \mathrm{vec}_d(P_{nk_p})]$ and

$$\Gamma_n = \begin{bmatrix} (\mu_2^{\dagger 2}/2)\omega_{ns}^\top \omega_{ns} & 0 \\ 0 & \Psi_n \end{bmatrix} \text{ with } \Psi_n = \begin{bmatrix} \mu_2^\dagger Q_n^\top Q_n & \mu_4^\dagger Q_n^\top \mathbf{1}_n \\ \mu_4^\dagger \mathbf{1}_n^\top Q_n & n\left(\mu_6^\dagger - \mu_3^{\dagger 2}\right) \end{bmatrix},$$

where $\omega_{ns} = [\mathrm{vec}(P_{n1}^s), \cdots, \mathrm{vec}(P_{nk_p}^s)]$, $\mu_2^\dagger = \left(\sigma_{v,o}^2 + (1-2/\pi)\sigma_{u,o}^2\right)$, and $\mu_3^\dagger = (1-4/\pi)\sqrt{2/\pi}\sigma_{u,o}^3$. Especially, even if $v_{in} \sim \mathcal{N}(0, \sigma_{v,o}^2)$, $\mu_4^\dagger - 3\mu_2^{\dagger 2} = 4(1/\pi + 3/\pi^2)\sigma_{u,o}^4 \neq 0$ by Lemma A.3, unless we further have $\sigma_{u,o} = 0$. Thus, that P_{nj}'s are selected from $\mathcal{P}_{1n}$ cannot guarantee $\Omega_n = \Gamma_n$. But, if P_{nj}'s are selected from $\mathcal{P}_{2n}$, Ω_n will equal to Γ_n since $\omega_{nd} = 0$.

As in Hansen (1982), the optimum distance matrix for $a_n^\top a_n$ is the inverse of $\frac{1}{n}\Omega_n$, so we require $\frac{1}{n}\Omega_n$ to be asymptotically invertible. Generally, Ω_n is singular if and only if the linear correlation exists among $g_n(\theta_o^\dagger)$. Recall that $g_{n2}(\theta_o^\dagger)$ in (3.3) is a cubic form in $\epsilon_{ni}^\dagger$, so no matter whether $\sigma_{u,o} = 0$ and $v_{in} \sim \mathcal{N}(0, \sigma_{v,o}^2)$, it must be independent of $g_{n1}(\kappa_o)$ in (3.2) that only includes linear and quadratic forms of $\epsilon_{ni}^\dagger$. The more details for correlation among elements of $g_{1n}(\kappa_o)$ can be found in Lee (2007a, p. 495). Thus, imposing the following conventional condition is always meaningful.

Assumption 13. The limit $\lim_{n\to\infty}\frac{1}{n}\Omega_n$ exists and is nonsingular.

Differently from $\frac{1}{n}\Omega_n$, the discussion, for whether the gradient matrix $\mathrm{E}[\partial g_n(\theta_o^\dagger)/\partial\theta^\dagger]$ has full rank, is slightly complex. We then analyze it by taking whether $\sigma_{u,o}=0$ into consideration.

4.1 As technical inefficiency exists: $\sigma_{u,o}>0$

Denote $-D_n^\dagger$ as the expectation of the gradient matrix $\partial g_n(\theta_o^\dagger)/\partial\theta^\dagger$. Then, by direct calculation, $D_n^\dagger$ is formulated as

$$D_n^\dagger=-\mathrm{E}\left[\frac{\partial g_n(\theta_o^\dagger)}{\partial\theta^\dagger}\right]=\begin{bmatrix}\mu_2^\dagger C_{k_pn} & (G_nX_n\beta_o^\dagger)^\top Q_n & 3\mu_2^\dagger(G_nX_n\beta_o^\dagger)^\top\mathbf{1}_n+\mu_3^\dagger\mathrm{tr}(G_n)\\ 0 & X_n^\top Q_n & 3\mu_2^\dagger X_n^\top\mathbf{1}_n\\ 0 & 0 & 3n\sqrt{\frac{2}{\pi}}(1-\frac{4}{\pi})\sigma_{u,o}^2\end{bmatrix}^\top \tag{4.2}$$

where $C_{k_pn}=[\mathrm{tr}(P_{n1}^sG_n),\cdots,\mathrm{tr}(P_{nk_p}^sG_n)]$ is k_p-dimensional row vector. It has not the full rank k_x+2 in the case of $\sigma_{u,o}=0$ (see the next subsection for details). Generally speaking, its full rank is guaranteed as $\sigma_{u,o}>0$ so that we can establish the asymptotic normality in the following proposition, which also suggests a limiting distribution for the estimator of $\beta_{1,o}$.

Proposition 3. *Supposed that the limit $\lim_{n\to\infty}a_n\left(\frac{1}{n}D_n^{\dagger\top}\right)$ exists and has the full rank k_x+1. Then, under 1–12 and $\sigma_{u,o}\neq0$, $\sqrt{n}(\hat\theta_{g,n}^\dagger-\theta_o^\dagger)\xrightarrow{d}\mathcal{N}(0,\Sigma_{g,n}^{-1})$, where*

$$\begin{aligned}\Sigma_{g,n}=\lim_{n\to\infty}&\left[\left(\frac{1}{n}D_n^{\dagger\top}\right)a_n^\top a_n\left(\frac{1}{n}D_n^{\dagger\top}\right)\right]^{-1}\left(\frac{1}{n}D_n^{\dagger\top}\right)a_n^\top a_n\left(\frac{1}{n}\Omega_n\right)\\&\times a_n^\top a_n\left(\frac{1}{n}D_n^{\dagger\top}\right)\left[\left(\frac{1}{n}D_n^{\dagger\top}\right)a_n^\top a_n\left(\frac{1}{n}D_n^{\dagger\top}\right)\right]^{-1}.\end{aligned}$$

Especially, $\sqrt{n}(\hat\beta_{1,g,n}^\dagger+\sqrt{2/\pi}\hat\sigma_{g,n})\xrightarrow{d}\mathcal{N}(\beta_{1,o},\mathbf{c}_{n,k_x+2}^\top\Sigma_{g,n}^{-1}\mathbf{c}_{n,k_x+2})$, where $\mathbf{c}_{n,k_x+2}=\left[0,1,\mathbf{0}_{k_x-1}^\top,\sqrt{2/\pi}\right]^\top$ is a (k_x+2)-dimensional column vector.

According to the generalized Schwartz inequality, the optimal selection of the distance matrix $a_n^\top a_n$ in Proposition 3 is $(\frac{1}{n}\Omega_n)^{-1}$. In practice, we cannot find the GMM estimator of $\theta_o^\dagger$ from solving $\min g_n^\top(\theta^\dagger)(\Omega_n)^{-1}g_n(\theta^\dagger)$, since the calculation of $(\frac{1}{n}\Omega_n)^{-1}$ refers to the unknown true value $\theta_o^\dagger$. Therefore, an initial consistent estimate of $\theta_o^\dagger$ is necessary, on which $\mu_2^\dagger$, $\mu_3^\dagger$, $\mu_4^\dagger$, and $\mu_6^\dagger$ can be consistently estimated based. Then $\frac{1}{n}\Omega_n$ can then be consistently estimated as $\frac{1}{n}\hat\Omega_n$, whereby we can derive the feasible optimum GMME (OGMME) $\hat\theta_{o,n}^\dagger$ from $\min g_n^\top(\theta^\dagger)(\hat\Omega_n)^{-1}g_n(\theta^\dagger)$. The following proposition portrays that $\hat\theta_{o,n}^\dagger$ has the same limiting distribution as the OGMME derived from the objective function based Ω_n. Moreover, an overidentification test statistic is provided.

Proposition 4. *Suppose that $\hat\Omega_n/n-\Omega/n=o_p(1)$. Then, under Assumptions 1–10, 12–13 and $\sigma_{u,o}\neq0$, the feasible OGMME $\hat\theta_{o,n}$ derived from $\min_{\theta\in\Theta}g_n^\top(\theta^\dagger)\hat\Omega_n^{-1}g_n(\theta^\dagger)$ is consistent for $\theta_o^\dagger$, and has the asymptotic distribution*

$$\sqrt{n}(\hat\theta_{o,n}^\dagger-\theta_o^\dagger)\xrightarrow{d}\mathcal{N}\left(0,\Sigma_{o,n}^{-1}\right),$$

where $\Sigma_{o,n}=\lim_{n\to\infty}\frac{1}{n}D_n^{\dagger\top}\Omega_n^{-1}D_n^\dagger$ is assumed to exist. Furthermore, we have

$$g_n^\top(\hat\theta_{o,n}^\dagger)\hat\Omega_n^{-1}g_n(\hat\theta_{o,n}^\dagger)\xrightarrow{d}\chi^2(k_q+k_p-k_x-1). \tag{4.3}$$

The best GMME (BGMME) corresponds to the best selection of the IV matrix Q_n and P_{nj}'s. The former can be derived from the 2SLS estimation, which yields

$$\left[\hat\lambda_{2sls,n},\hat\beta_{1,2sls,n}^\dagger,\hat\beta_{-1,2sls,n}^\top\right]^\top=[Z_n^\top P_qZ_n]^{-1}Z_n^\top P_qY_n,$$

where $P_q = Q_n(Q_n^\top Q_n)^{-1}Q_n^\top$, $Z_n = [W_nY_n, X_n]$, and $\hat{\beta}^\dagger_{1,2sls,n}$ is a consistent estimator for $\beta^\dagger_{1,o} = \beta_{1,o} - \sqrt{2/\pi}\sigma_{u,o}$. The corresponding asymptotic variance (Asy.var) matrix is

$$\mu_2^\dagger \cdot \lim_{n\to\infty} \left\{ \frac{1}{n} \left[G_nX_n\beta_o^\dagger, X_n\right]^\top P_q \left[G_nX_n\beta_o^\dagger, X_n\right] \right\}^{-1} \tag{4.4}$$

under Assumption 8(a) and $\lim_{n\to\infty} \frac{1}{n}Q_n^\top Q_n$ is nonsingular (Kelejian and Prucha, 1998). This implies that by the generalized Schwartz inequality applied to the Asy.var in (4.4), the best IV matrix Q_n would be $[G_nX_n\beta_o^\dagger, X_n]$, equivalently to $[G_n(X_n\beta_o + \sqrt{2/\pi}\sigma_{u,o}\mathbf{1}_n), X_n]$, determined by the characteristic of the SARSF model. For the case where $G_nX_n\beta_o^\dagger$ and X_n are linearly dependent, $G_nX_n\beta_o^\dagger$ is redundant under Assumption 4(i) and therefore, the best IV matrix would be X_n.

Next, it is of interest to select the best P_{nj}'s. As we said before, P_{nj}'s from $\mathcal{P}_{1n}$ cannot guarantee $\Omega_n = \Gamma_n$ (4.1). Thus, we only consider the case where P_{nj}'s be from $\mathcal{P}_{2n}$, then $\Omega_n = \Gamma_n$ always holds. In such a case, we have

$$D_n^{\dagger\top}\Gamma_n^{-1}D_n^\dagger = \begin{bmatrix} \frac{1}{2}C_{k_pn}(\omega_{ns}\omega_{ns}^\top)^{-1}C_{k_pn}^\top & 0 \\ 0 & 0 \end{bmatrix} + D_{n,23}^{\dagger\top}\Psi_n^{-1}D_{n,23}^\dagger\big|_{Q_n=[G_nX_n\beta_o^\dagger, X_n]}. \tag{4.5}$$

where $D_{n,23}^{\dagger\top}$ is the last k_q+1 columns of $D_n^{\dagger\top}$ in(4.2). Here the second term on the right-hand side of (4.5) does not involve P_{nj}'s, which implies that we just need to search the upper bound of the first term by choosing the best P_{nj}'s, so that the Asy.var $\lim_{n\to\infty} \frac{1}{n}(D_n^\top\Omega_n^{-1}D_n)^{-1}$ can arrive its lower bound as far as possible. Notice that below (4.2),

$$\mathrm{tr}(P_{nj}^sG_n) = \mathrm{tr}(P_{nj}^sG_n^d) = \frac{1}{2}\mathrm{tr}(P_{nj}^sG_n^{ds}) = \frac{1}{2}\mathrm{vec}^\top(G_n^{ds})\mathrm{vec}(P_{nj}^s),$$

where $G_n^{ds} = G_n^d + G_n^{d\top}$ and $G_n^d = G_n - \mathrm{Diag}(G_n)$ with $\mathrm{Diag}(G_n)$ being the diagonal matrix generated by diagonal elements of G_n. Thus, by the generalized Schwartz inequality, we can obtain

$$\frac{1}{2}C_{k_pn}(\omega_{ns}^\top\omega_{ns})^{-1}C_{k_pn}^\top \le \frac{1}{2}\mathrm{vec}^\top(G_n^{ds})\mathrm{vec}(G_n^{ds}) = \mathrm{vec}^\top(G_n^{ds})\mathrm{vec}(G_n) = \mathrm{vec}^\top(P_{nj}^s)\mathrm{vec}(G_n),$$

where $A_n \le B_n$ means $B_n - A_n$ is positive semi-definite. These tell us that the best P_{nj}'s would be G_n^{ds} ($k_p = 1$). In practice, with initial consistent estimators $(\hat{\lambda}_n, \hat{\beta}_1^\dagger, \hat{\beta}_{-1,n})$ of $(\lambda_o, \beta_{1,o}^\dagger, \beta_{-1,o})$, G_n and $G_nX_n\beta_o^\dagger$ can, respectively, estimated consistently by $\hat{G}_n \triangleq W_n(I_n - \hat{\lambda}_nW_n)^{-1}$ and $\hat{G}_nX_n\hat{\beta}_n^\dagger$. In addition, $\Omega_n = \Gamma_n$ can be estimates as $\hat{\Gamma}_n$. The following proposition portrays that the limiting distribution of the feasible BGMME is the same as that of the BGMME.

Proposition 5. *Under Assumptions 1–6, 9–10, 12 and $\sigma_{u,o} \neq 0$, suppose that $\hat{\theta}_n^\dagger$ is a $\sqrt{n}$-consistent estimator of $\theta_o^\dagger$. Then, within the class of GMMEs derived with $\mathcal{P}_{2n}$, the feasible BGMME $\hat{\theta}_{b,n}^\dagger$ has the asymptotic distribution $\sqrt{n}(\hat{\theta}_{b,n}^\dagger - \theta_o^\dagger) \xrightarrow{d} \mathcal{N}(0, \Sigma_{b,n}^{-1})$, where*

$$\Sigma_{b,n} = \lim_{n\to\infty} \frac{1}{n} \begin{bmatrix} \mathrm{tr}(G_n^{ds}G_n) & 0 \\ 0 & 0 \end{bmatrix} + \lim_{n\to\infty} \frac{1}{n}D_{n,23}^{\dagger\top}\Psi_n^{-1}D_{n,23}^\dagger\big|_{Q_n=[G_nX_n\beta_o^\dagger, X_n]}$$

is assumed to exist.

4.2 As technical inefficiency does not exist: $\sigma_{u,o} = 0$

In what follows of this subsection, we only consider the scenario where the IV matrix Q_n and P_{nj} are the best, like in Proposition 5. As $\sigma_{u,o} = 0$, $\frac{1}{n}\Omega_n$ in (4.1) is asymptotically nonsingular, but in (4.2),

$$\mathrm{E}\left[\frac{\partial g_n(\theta_o^\dagger)}{\partial\sigma_{u,o}}\right] = 3\sigma_{v,o}^2[\mathbf{1}_n^\top G_nX_n\beta_o, \mathbf{1}_n^\top X_n, 0] = 3\sigma_{v,o}^2\mathrm{E}\left[\frac{\partial g_n(\theta_o^\dagger)}{\partial\beta_1^\dagger}\right],$$

which indicates $\mathrm{E}[\partial g_n(\theta_o^\dagger)/\partial\theta^\dagger]$ has not the full rank k_x+2. Accordingly, all the asymptotic distributions in Propositions 3–5 do not work, and this phenomenon also appears in the ML estimation (Jin and Lee, 2020). To overcome this GMM framework's shortcomings, we suggest the reparameterization that $\tau=\sigma_u^3$. Denote $\theta^\ddagger=[\kappa^\top,\tau]^\top$ with its true value being $\theta_o^\ddagger=[\kappa_o^\top,\tau_o]$. Obviously, $\theta_o^\ddagger=[\lambda_o,\beta_o^\top,0]^\top$ since $\sigma_{u,o}=0$. Then the expectation of $\partial g_n(\theta_o^\ddagger)/\partial\theta^\ddagger$ at $\theta_o^\ddagger$ becomes

$$D_n^\ddagger \triangleq -\mathrm{E}\left[\frac{\partial g_n(\theta_o^\ddagger)}{\partial\theta^\ddagger}\right]=\begin{bmatrix}\sigma_{v,o}^2\mathrm{tr}(G_n^{ds}G_n) & (G_nX_n\beta_o)^\top Q_n & 3\sigma_{v,o}^2(G_nX_n\beta_o)^\top\mathbf{1}_n\\ 0 & X_n^\top Q_n & 3\sigma_{v,o}^2X_n^\top\mathbf{1}_n\\ 0 & 0 & n\sqrt{\frac{2}{\pi}}(1-\frac{4}{\pi})\end{bmatrix}^\top. \tag{4.6}$$

Generally speaking, $D_n^\ddagger$ has the full rank k_x+2, since $\sqrt{2/\pi}(1-4/\pi)\neq 0$. Let $\delta_{q,n}=\frac{1}{n}\mathbf{1}_n^\top[G_nX_n\beta_o,X_n]([G_nX_n\beta_o,X_n]^\top[G_nX_n\beta_o,X_n])^{-1}[G_nX_n\beta_o,X_n]^\top\mathbf{1}_n$. Then, by direct matrix calculation we have $\frac{1}{n}D_n^{\ddagger\top}\Omega_n^{-1}D_n^\ddagger=\Sigma_{b,n}^\ddagger=\Sigma_{b1,n}^\ddagger+\Sigma_{b2,n}^\ddagger$, where $\Sigma_{b1,n}^\ddagger$ equals

$$\frac{1}{n\sigma_{v,o}^2}\begin{bmatrix}\sigma_{v,o}^2\mathrm{tr}(G_n^{ds}G_n)+(G_nX_n\beta_o)^\top G_nX_n\beta_o & (G_nX_n\beta_o)^\top X_n & 0\\ X_n^\top G_nX_n\beta_o & X_n^\top X_n & 0\\ 0 & 0 & s_n\end{bmatrix}, \tag{4.7}$$

with $s_n=n\sigma_{u,o}^{-4}(15-9\delta_{q,n})^{-1}(2/\pi)(1-4/\pi)^2$, and $\Sigma_{b2,n}^\ddagger$ equals

$$\frac{1}{n\sigma_{v,o}^2}\begin{bmatrix}\gamma\cdot\frac{1}{n}(G_nX_n\beta_o)^\top\mathbf{1}_n\mathbf{1}_n^\top G_nX_n\beta_o & \gamma\cdot\frac{1}{n}(G_nX_n\beta_o)^\top\mathbf{1}_n\mathbf{1}_n^\top X_n & (3\sigma_{v,o}^4-\mu_{4,v})(G_nX_n\beta_o)^\top\mathbf{1}_n\\ \gamma\cdot\frac{1}{n}X_n^\top\mathbf{1}_n\mathbf{1}_n^\top G_nX_n\beta_o & \gamma\cdot\frac{1}{n}X_n^\top\mathbf{1}_n\mathbf{1}_n^\top X_n & (3\sigma_{v,o}^4-\mu_{4,v})X_n^\top\mathbf{1}_n\\ (3\sigma_{v,o}^4-\mu_{4,v})\mathbf{1}_n^\top G_nX_n\beta_o & (3\sigma_{v,o}^4-\mu_{4,v})\mathbf{1}_n^\top X_n & t_n-s_n\end{bmatrix}$$

with where $\mu_{4,v}=\mathrm{E}[v_{ni}^4]$, $\gamma=\mu_4^2/\sigma_{v,o}^2-6\mu_4\sigma_{v,o}^2+9\sigma_{v,o}^6$ and $t_n=n\sigma_{v,o}^2(\mu_6-\mu_4^2/\sigma_{v,o}^2\delta_{q,n})^{-1}(2/\pi)(1-4/\pi)^2$. Furthermore, if v_{ni}'s are normal, then $\Sigma_{b2,n}^\ddagger$ degenerates to a zero matrix, because $\gamma=0$, $\mu_{4,v}=3\sigma_{v,o}^4$ and $t_n=s_n$. Therefore, when $\sigma_{u,o}=0$, we can establish the limiting distribution as follows.

Proposition 6. *Under Assumptions 1–6, 9–10, 12 and $\sigma_{u,o}=0$, suppose that $\hat\theta_n^\ddagger$ is a $\sqrt{n}$-consistent estimator of $\theta_o^\ddagger$. Then, within the class of GMMEs derived with $\mathcal{P}_{2n}$, the feasible BGMME $\hat\theta_{b,n}^\ddagger$ has the asymptotic distribution $\sqrt{n}(\hat\theta_{b,n}^\ddagger-\theta_o^\ddagger)\xrightarrow{d}\mathcal{N}(0,(\lim_{n\to\infty}\Sigma_{b,n}^\ddagger)^{-1})$, where $\lim_{n\to\infty}\Sigma_{b,n}^\ddagger$ is assumed to exist. Especially, as v_{ni}'s are normal, we have $\sqrt{n}(\hat\theta_{b,n}^\ddagger-\theta_o^\ddagger)\xrightarrow{d}\mathcal{N}(0,(\lim_{n\to\infty}\Sigma_{b1,n}^\ddagger)^{-1})$.*

On the one hand, Proposition 6 shows the existence of limiting distribution of $\hat\theta_{b,n}^\ddagger$ as $\sigma_{u,o}=0$.[9] On the other hand, it follows from $\sqrt{n}(\hat\tau_{b,n})=O_p(1)$ that $n^{1/6}(\hat\sigma_{u,b,n})=O_p(1)$, which indicates $\hat\sigma_{u,b,n}$ have a slower convergence rate than one in the case of $\sigma_{u,o}>0$. Particularly, we can see from (4.7) that as v_{ni}'s are normal, $\hat\tau_{b,n}$ and $(\hat\lambda_{b,n},\hat\beta_{b,n}^\dagger)$ is uncorrelated. This implies $\mathrm{Asy.var}(\hat\tau_{b,n})=\lim_{n\to\infty}\sigma_{u,o}^6(15-9\delta_{q,n})\frac{2}{\pi}(1-\frac{4}{\pi})^2$. Here $|\delta_{q,n}|\le 1$ by Lemma A.5, and equality holds if and only if X_n includes the intercept term $\mathbf{x}_{n,1}=\mathbf{1}_n$.

When we test whether $\tau_o=0$, we find that it is not enough to reject by using the statistic derived from the proposition. In other words, the empirical size of the test is slightly lower than the nominal size we set previously, particularly as the sample size is small. Such phenomenon may be resulted from the convergence mentioned above rate $n^{1/6}$. To overcome such a problem to some extent, we consider a test statistic based on the third moment of the residuals. Indeed, $\hat\tau_{b,n}$ is derived by solving the optimization problem $\min_{\theta^\ddagger\in\Theta^\ddagger}\hat g_{b,n}^\top(\theta^\ddagger)\hat\Gamma_n^{-1}\hat g_{b,n}(\theta^\ddagger)$, which does not necessarily implies $g_{n2}(\theta_{b,n}^\ddagger)=0$. Thus $\sqrt{2/\pi}(1-4/\pi)\hat\tau_{b,n}$ may be not equal to $\frac{1}{n}\sum_{i=1}^n\epsilon_{ni}^{\dagger 3}(\hat\kappa_{b,n})$ numerically.

[9] However when the "wrong skew" problem occurs, we still set $\hat\tau_{g,n}=0$.

5 Testing for the existence of technical inefficiency

In this section, we pay attention to testing the existence of technical inefficiency, i.e., the null hypothesis is

$$H_0 : \sigma_{u,o} = 0. \tag{5.1}$$

Provided that (5.1) is true, the SARSF model in (2.1) reduces to the conventional SAR model $Y_n = \lambda W_n Y_n + X_n\beta_n + V_n$. Denote $\xi = [\lambda, \beta^\top]^\top$, its true value $\xi_o = [\lambda_o, \beta_o^\top]^\top$, and the BGMME $\hat{\xi}_{b,n} = [\hat{\lambda}_{b,n}, \hat{\beta}_{b,n}]$. Under Assumptions 2–6, $\hat{\xi}_{b,n}$ has the asymptotic distribution (Lee, 2007a)

$$\sqrt{n}(\hat{\xi}_{b,n} - \xi_o) = A_n \frac{1}{\sqrt{n}} \left[V_n^\top G_n^d V_n, V_n^\top G_n X_n \beta_o, V_n^\top X_n\right]^\top + o_p(1) \xrightarrow{d} \mathcal{N}\left(0, (\lim_{n\to\infty} \Sigma_{\xi,n})^{-1}\right), \tag{5.2}$$

where $A_n = \sigma_{v,o}^2 (\Sigma_{\xi,n})^{-1} D_{\xi,n}^\top \Omega_{\xi,n}^{-1}$ with

$$\Sigma_{\xi,n} = \frac{1}{n}\begin{bmatrix} \sigma_{v,o}^2 \mathrm{tr}(G_n^{ds} G_n) + (G_n X_n \beta_o)^\top G_n X_n \beta_o & (G_n X_n\beta_o)^\top X_n \\ X_n^\top G_n X_n \beta_o & X_n^\top X_n \end{bmatrix},$$

$$D_{\xi,n} = \frac{1}{n}\begin{bmatrix} \sigma_{v,o}^2 \mathrm{tr}(G_n^{ds} G_n) & (G_n X_n \beta_o)^\top G_n X_n \beta_o & (G_n X_n\beta_o)^\top X_n \\ 0 & X_n^\top G_n X_n \beta_o & X_n^\top X_n \end{bmatrix}^\top,$$

$$\Omega_{\xi,n} = \sigma_{v,o}^2 \cdot \frac{1}{n}\begin{bmatrix} \sigma_{v,o}^2 \mathrm{tr}(G_n^{ds} G_n) & 0 & 0 \\ 0 & (G_n X_n \beta_o)^\top G_n X_n \beta_o & (G_n X_n\beta_o)^\top X_n \\ 0 & X_n^\top G_n X_n \beta_o & X_n^\top X_n \end{bmatrix}.$$

Define $\mathbf{e}(\xi) = S_n(\lambda) Y_n - X_n\beta$ and obviously, the residual vector is $\hat{\mathbf{e}}_{b,n} = \mathbf{e}(\hat{\xi}_{b,n}) = S_n(\hat{\lambda}_{b,n}) Y_n - X_n \hat{\beta}_{b,n} \triangleq [\hat{e}_{b,n1}, \cdots, \hat{e}_{b,nn}]$. By the expansion, we have

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n \hat{e}_{b,ni}^3 =& \frac{1}{n}\sum_{i=1}^n v_{ni}^3 + \frac{3}{n}\sum_{i=1}^n v_{ni}^2 \mathbf{z}_{i,n}^\top (\xi_o - \hat{\xi}_{b,n}) + \frac{3}{n}\sum_{i=1}^n v_{ni} \left[\mathbf{z}_{i,n}^\top(\xi_o - \hat{\xi}_{b,n})\right]^2 \\ &+ \frac{1}{n}\sum_{i=1}^n \left[\mathbf{z}_{i,n}^\top (\xi_o - \hat{\xi}_{b,n})\right]^3, \end{aligned} \tag{5.3}$$

where $z_{ni} = [\mathbf{w}_{i,n} Y_n, \mathbf{x}_{i,n}^\top]$. By the Lindeberg-Lévy CLT, $\frac{1}{n}\sum_{i=1}^n v_{ni}^3 = O_p(n^{-1/2})$ since $\mathrm{E}[v_{ni}^3] = 0$. Let $h_{n,ij}$ be either 1 or any element of $\mathbf{z}_{i,n}^\top$ for $j = 1, 2, 3$. Then according to the generalized Hölder's inequality, under Assumptions 1–6, 9–10 and 12,

$$\sup_{i,n} \mathrm{E}\left[h_{n,i1} h_{n,i2} h_{n,i3}|\right] \le \sup_{i,n} \left\{(\mathrm{E}|h_{n,i1}|^3)^{1/3} (\mathrm{E}|h_{n,i2}|^3)^{1/3} (\mathrm{E}|h_{n,i3}|^3)^{1/3}\right\} < \infty$$

and

$$\sup_{i,n} \mathrm{E}\left[|v_{ni}^k h_{n,i1} h_{n,i2}|\right] \le \sup_{i,n} \left\{(\mathrm{E}|v_{ni}|^{3k})^{1/3} (\mathrm{E}|h_{n,i1}|^3)^{1/3} (\mathrm{E}|h_{n,i2}|^3)^{1/3}\right\} < \infty, \quad k = 0, 1, 2$$

hold. Thus, by the LLN in Jenisha and Prucha (2012), $\frac{1}{n}\sum_{i=1}^n v_{ni}^2 \mathbf{z}_{i,n}^\top = O_p(1)$, $\frac{1}{n}\sum_{i=1}^n v_{ni} \mathbf{z}_{i,n} \mathbf{z}_{i,n}^\top = O_p(1)$ and $\frac{1}{n}\sum_{i=1}^n \mathbf{z}_{i,n}\mathbf{z}_{i,n}^\top z_{n,ij} = O_p(1)$. Working with $\xi_o - \hat{\xi}_{b,n} = o_p(1)$, it yields that $\frac{1}{n}\sum_{i=1}^n \hat{e}_{b,ni}^3 = o_p(1)$.

We now consider the limiting distribution of $(1/\sqrt{n})\sum_{i=1}^n \hat{e}_{b,n1}^3$. It follows from (5.2) that

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n \hat{e}_{b,ni}^3 =& \frac{1}{n}\sum_{i=1}^n v_{ni}^3 - \frac{3}{\sqrt{n}} \sigma_{v,o}^2 \left[\frac{1}{n}\mathbf{1}_n^\top G_n X_n \beta_o, \frac{1}{n}\mathbf{1}_n^\top X_n\right] \\ & \times A_n \frac{1}{\sqrt{n}} \left[V_n^\top G_n^d V_n, V_n^\top G_n X_n\beta_o, V_n^\top X_n\right]^\top + o_p(n^{-1/2}) \\ \triangleq& R_n + o_p(n^{-1/2}), \end{aligned} \tag{5.4}$$

which shall be checked later in the proof of the subsequent Proposition 7. By using Theorem 1 again, we can derive the limiting distribution of $\sqrt{n}R_n$. Especially, as v_{ni}'s are normal, it follows from

$$\begin{aligned}
&\mathrm{E}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} v_{ni}^3 \cdot \frac{1}{\sqrt{n}}\left[V_n^\top G_n^d V_n, V_n^\top G_n X_n\beta_o, V_n^\top X_n\right] A_n^\top\right\}\\
=&\mu_{4,v}\sigma_{v,o}^2 \cdot \left[0, \frac{1}{n}\mathbf{1}_n^\top G_n X_n\beta_o, \frac{1}{n}\mathbf{1}_n^\top X_n\right]\Omega_{\xi,n}^{-1} D_{\xi,n}\Sigma_{\xi,n}^{-1}\\
=&3\sigma_{v,o}^4 \cdot \left[0, \frac{\sigma_{v,o}^2}{n}\mathbf{1}_n^\top G_n X_n\beta_o, \frac{\sigma_{v,o}^2}{n}\mathbf{1}_n^\top X_n\right]\Omega_{\xi,n}^{-\top} D_{\xi,n}\Sigma_{\xi,n}^{-1} = 3\sigma_{v,o}^4\left[\frac{1}{n}\mathbf{1}_n^\top G_n X_n\beta_o, \frac{1}{n}\mathbf{1}_n^\top X_n\right]\Sigma_{\xi,n}^{-1}
\end{aligned}$$

that $\mathrm{E}\left\{\sqrt{n}R_n \cdot \frac{1}{\sqrt{n}}\left[V_n^\top G_n^d V_n, V_n^\top G_n X_n\beta_o, V_n^\top X_n\right] A_n^\top\right\} = 0$. This implies $\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{e}_{b,n1}^3$ is asymptotically uncorrelated with $\sqrt{n}(\hat{\xi}_{b,n} - \xi_o)$, which conforms to the conclusion in Proposition 6 that $\hat{\tau}_{b,n}$ and $(\hat{\lambda}_{b,n}, \hat{\beta}_{b,n}^\dagger)$ is uncorrelated as v_{ni}'s are normal.[10] It leads to an easy-to-compute test statistic in the normality case.

Proposition 7. *Under under Assumptions 1–6, 9–10, 12, and* $\sigma_{u,o} = 0$, $\sqrt{n}R_n \xrightarrow{d} \mathcal{N}(0, \sigma_{R,n}^2)$, *where*

$$\begin{aligned}
\sigma_{R,n}^2 =&\sigma_{R,n}^2(\xi_o, \sigma_{v,o})\\
=&\mu_{6,v} + 9\lim_{n\to\infty}\sigma_{v,o}^6\left[\frac{1}{n}\mathbf{1}_n^\top G_n X_n\beta_o, \frac{1}{n}\mathbf{1}_n^\top X_n\right]\Sigma_{\xi,b}^{-1}\left[\frac{1}{n}G_n X_n\beta_o^\top\mathbf{1}_n, \frac{1}{n}X_n^\top\mathbf{1}_n\right]\\
&- 6\sigma_{v,o}^2\mu_{4,v}\lim_{n\to\infty}\left[0, \frac{\sigma_{v,o}^2}{n}\mathbf{1}_n^\top G_n X_n\beta_o, \frac{\sigma_{v,o}^2}{n}\mathbf{1}_n^\top X_n\right]\Omega_{\xi,n}^{-\top} D_{\xi,n}\Sigma_{\xi,n}^{-1}\left[\frac{1}{n}G_n X_n\beta_o^\top\mathbf{1}_n, \frac{1}{n}X_n^\top\mathbf{1}_n\right],
\end{aligned}$$

where $\mu_{6,v} = rmE[v_{ni}^6]$. *Moreover, we have*

$$\mathrm{T}_{b,n} = \frac{\sum_{i=1}^n \hat{e}_{b,ni}^3}{\sqrt{n}\sigma_{R,n}\left(\hat{\xi}_{b,n}, [\frac{1}{n}\hat{\mathbf{e}}_{b,n}^\top\hat{\mathbf{e}}_{b,n}]^{1/2}\right)} \xrightarrow{d} \mathcal{N}(0,1). \tag{5.5}$$

Especially, when v_{ni}*'s are normal,* $\mathrm{T}_{b,n}$ *in* (5.5) *simplifies to*

$$\frac{n\sum_{i=1}^n \hat{e}_{b,ni}^3}{(15 - 9\hat{\delta}_{q,n})^{1/2}(\sum_{i=1}^n \hat{e}_{b,ni}^2)^{3/2}} \xrightarrow{d} \mathcal{N}(0,1),$$

where $\hat{\delta}_{q,n} = \mathbf{1}_n^\top[\frac{1}{n}\hat{G}_n X_n\hat{\beta}_{b,n}, \frac{1}{n}X_n]\hat{\Sigma}_{\xi,n}^{-1}[\frac{1}{n}\hat{G}_n X_n\hat{\beta}_{b,n}, \frac{1}{n}X_n]^\top\mathbf{1}_n$.

Here $15 - 9\hat{\delta}_{q,n}$ must be positive, because $\hat{\delta}_{q,n} < 1$ strictly holds which can be proved by applying the singular value decomposition theorem for $[\frac{1}{n}\hat{G}_n X_n\hat{\beta}_{b,n}, \frac{1}{n}X_n]$. The test statistics in Proposition 7 is left-sided because of the alternative hypothesis $\sigma_{u,o} > 0$. Therefore, the null hypothesis (5.1) should be rejected if $\mathrm{T}_{b,n} \le c_\alpha$, where c_α is the quantile satisfying $\Phi(c_\alpha) = \alpha$ for a significance level given previously.

Jin and Lee (2020, Prop. 2.4) also derive a score (LM) test for (5.1), whose statistic is computationally simple by solving the MLE of the restricted SARSF (i.e., SAR) model. However, our numerical experiments in the next section suggest that their score test is not robust, and it strongly depends on not only the normality of V_n, but also the existence of the intercept included in X_n. Once anyone of both is violated, the reject rate of the statistic seems to be either higher or lower.

Finally, we briefly discuss marginal impacts generated by the generic regressor $\mathbf{x}_{n,k}$. By taking derivative of (3.1) with respect to $\mathbf{x}_{n,k}$, it follows that the marginal impact matrix

$$\frac{\partial Y_n}{\partial \mathbf{x}_{n,1}} = (I_n - \lambda_o W_n)^{-1}\beta_{k,o} \triangleq S_{n,k}(W_n), \quad k = 2, \cdots, k_x.$$

Based on the matrix $S_{n,k}(W_n)$, LeSage and Pace (2009) establish the concepts of marginal effects, including (1)

[10] This conclusion also holds in the ML estimation (Jin and Lee, 2020, p. 13)

direct effects, computed as the average of the diagonal elements of $S_{n,k}(W_n)$; (2) indirect effects, calculated as the average of the sum of the off-diagonal elements of $S_{n,k}(W_n)$; and (3) total effects equal to the sum of the direct and the indirect effects. To test whether triple effects exist significantly, the related standard errors or t-statistics can be computed by either the Bayesian MCMC simulation approach (LeSage and Pace, 2009), the Delta method (Taşpınar et al., 2018; Arbia et al., 2020), or structuring the Lagrange Multiplier and artificial regression test statistics (Deng and Wang, 2022).

6 Monte Carlo simulations

6.1 Experiment design

To investigate the finite sample properties of GMME in each case, we set the first data generated progress of model (2.1) with intercept

$$\textbf{DGP-I:}\ (I_n-\lambda_o W_n)^{-1}(\mathbf{1}_n\beta_{0,o}+\mathbf{x}_1\beta_{1,o}+\mathbf{x}_2\beta_{2,o}+V_n-U_n),$$

where the inefficiency term $U_n\sim\mathcal{N}^+(0,\sigma_{u,o}^2\cdot I_n)$, the sample size n varies among $\{100,225,400\}$, $\mathbf{x}_1$ and $\mathbf{x}_2$ are generated from $\mathcal{N}(\mathbf{0}_n,I_n)$, and slopes $(\beta_{0,o},\beta_{1,o},\beta_{2,o})=(0.5,1,1.5)$. Based on the symmetry of v_{in}'s distribution and the relative size of $\sigma_{u,o}$ and $\sigma_{v,o}$, we consider four cases as shown below.

- Case I: v_{in}'s are sampled from $\mathcal{N}(0,1)$ (i.e., $\sigma_{v,o}=1$), and the value of $\sigma_{u,o}$ is chosen from $\{0.5,1,1.5\}$,
- Case II: v_{in}'s are sampled from the Student's t distribution with t degrees of freedom $\mathcal{T}(4)$ (i.e., $\sigma_{v,o}=\sqrt{2}$), and the value of $\sigma_{u,o}$ is chosen from $\{1,\sqrt{2},2\}$,
- Case III: v_{in}'s are sampled from the Uniform distribution $\mathcal{U}(-3,3)$ (i.e., $\sigma_{v,o}=\sqrt{3}$), and the value of $\sigma_{u,o}$ is chosen from $\{1,\sqrt{3},2\}$,
- Case IV: v_{in}'s are sampled from the mixed distribution $\sqrt{3/2}\mathcal{N}(0,1)+\sqrt{1/2}\mathcal{T}(4)+\sqrt{1/2}\mathcal{U}(-3,3)$ (i.e., $\sigma_{v,o}=2$), and the value of $\sigma_{u,o}$ is chosen from $\{1.5,2,2.5\}$.

For the remaining setting, the spatial lag parameter $\lambda_o=-0.5$ and the spatial weight matrix W_n is based on the Queen criterion in Cases I–II, and $\lambda_o=0.5$ and W_n is based on the Rook criterion in Cases III–IV.[11] Here, we consider the estimation methods as follows. (1) C2SLS–the corrected 2SLS method with IV's X_n, W_nX_n, and $W_n^2X_n$; (2) OGMM–the optimum GMM approach with the distance matrix being the inverse of the related variance matrix based on C2SLS (i.e., $a_n^\top a_n=\hat\Omega_n(\theta_{2sls,n}^\dagger)^{-1}$), the IV matrix Q_n consisting of independent columns of $[X_n,W_nX_n,W_n^2X_n]$ for linear moments, and $P_{1n}=W_n$ and $P_{2n}=W_n^2-(\mathrm{tr}(W_n^2)/n)I_n$ for quadratic moments; (3) BGMM–the best GMM approach with $a_n^\top a_n=\hat\Omega_n(\theta_{2sls,n}^\dagger)^{-1}$, using the IV matrix $Q_n=[X_n,G_n(\hat\lambda_{2sls,n})X_n\hat\beta_{2sls,n}^\dagger]$ for linear moments, $P_n=G_n(\hat\lambda_{2sls,n})-\mathrm{Diag}(G_n(\hat\lambda_{2sls,n}))$ for quadratic moments. To facilitate the comparison of various estimators, we report some measures of central tendency and dispersions in Tables 1–2, including the standard deviation (SD), mean square error (MSE), and the interdecile range (IDR)[12] of the empirical distribution. For σ_u, we further report the percentages of estimates equal zero (Zero-Rate), SD, and MSE of the remaining estimates with all zeros excluded.

To investigate the finite sample performance of the BGMM-based test for the null hypothesis (5.1), besides DGP-I, we also consider the following DGP without intercept,

$$\textbf{DGP-II:}\ Y_n=(I_n-\lambda_o W_n)^{-1}(\mathbf{x}_1\beta_{1,o}+\mathbf{x}_2\beta_{2,o}+V_n-U_n),$$

where except the intercept term all the other settings are fully identical to those in DGP-I. We show the empirical sizes of our test and the score counterpart (Jin and Lee, 2020) in Figures 1–2, and plot their empirical powers (the nominal size is set to 5%) in Figures 3–4. All results are based on 5000 replications for each case, and the regressors are randomly redrawn for each repetition.

[11] See Kelejian and Piras (2017, p. 8) for details about the Queen and Rook criteria.
[12] The IDR is the difference between the 90% quantile and 10% quantile in the empirical distribution.

6.2 Results for estimation

Table 1 reports the estimation results of Case I (normality) and II (Student's t distribution). In Case I, no matter whether $\sigma_{u,o}$ is smaller, equal, or bigger than $\sigma_{v,o}$: (i) for λ, BGMME, and OGMME are the best with approximate performance, and C2SLSE performs the worst with the largest SDs, MSEs, and IDRs; this phenomenon always exists as the sample size n gets large gradually; (ii) for β_k with $k = 0, 1, 2$ and σ_v, BGMME has the slightly smaller MADs and RMSEs than those of C2SLSE and OGMME, but the triple perform similarly, i.e., they seem to be approximately equal as n tends to infinity; (iii) as for σ_u, the three estimators have similar performance, and the pattern of the estimates are obviously related to the relative size of $\sigma_{u,o}$ and $\sigma_{u,o}$. That is, nearly half of the scenarios correspond to all estimators equal zero when $\sigma_{u,o} < \sigma_{v,o}$, the Zero-Rate decreases to 20% when $\sigma_{u,o} = \sigma_{v,o}$, and such ratio becomes nor over than 1% as $\sigma_{u,o} > \sigma_{v,o}$. In Case II, whatever the relative size between $\sigma_{u,o}$ and $\sigma_{v,o}$, (i) for λ, β_0, σ_u, and σ_v, the BGMME and OGMME have smaller SDs, MSEs, and IDRs than those of the C2SLSEs in most cases; (ii) for β_1 and β_2, we can find that BGMME outperforms slightly better than the other two, but none of them has a dominating performance in terms of Sds, MSEs, and IDRs.

Table 2 reports the estimation results of Case III (Uniform distribution) and IV (mixed symmetric distribution). It is found that in most cases, BGMME performs still the best, and C2SLSE produces the larger SDs, MSEs, and IDRs, which follows the pattern in Table 1. Particularly, relative to β_0 and σ_u, the three estimators for the remaining parameters possess somewhat larger SDs, MSEs, and IDRs. In addition, by comparing different results in the different sizes between $\sigma_{u,o}$ and $\sigma_{v,o}$, we can summarize the pattern as follows (take the mixed distribution as an example): (i) in the case of $\sigma_{u,o} < \sigma_{v,o}$, the Zero-Rate is approximately 40% as $n = 100$, then declines from 35% as $n = 225$ to 31% as $n = 400$; (ii) in the case of $\sigma_{u,o} = \sigma_{v,o}$, the rate nears 32% as $n = 100$, then declines from 24% as $n = 225$ to 18% as $n = 400$; This ratio falls apparently faster than that in the case of $\sigma_{u,o} < \sigma_{v,o}$; (iii) in the case of $\sigma_{u,o} > \sigma_{v,o}$, such speed continues to fast. This suggests that a relatively severe wrong skew problem arises in the case of $\sigma_{u,o} < \sigma_{v,o}$, and this problem is gradually improved as the size relationship between $\sigma_{u,o}$ and $\sigma_{v,o}$ changed from "=" to ">".

In a word, for estimating the partially distribution-free SARSF model, BGMME generates the smallest measures of central tendency and dispersions and therefore performs the best.

6.3 Results for testing

For the sizes of the two tests in finite samples, our main findings are as follows.

- **DGP-I (with intercept):** (i) The three figures in the first column of Figure 1 portray that in Case I (normality), the performance of the BGMM and score tests is satisfactory and is almost identical, because both plots are very close to the diagonal line, which indicates their empirical sizes are approximately equal to the nominal sizes (say α). (ii) We can see from the figures in the second column that in Case II (Student's t distribution), the BGMM test outperforms the score test which is rather disappointing; concretely, the plot of the BGMM test lies slightly below the diagonal line meaning slightly inflated type I errors when $\alpha \in [0, 0.5]$, and lies slightly up the diagonal line when $\alpha \in (0.5, 1)$ indicating slightly deflated type I errors; the plot for the score test strongly deviate from the diagonal line. (iii) Turning to figures in the third column, the plot of the BGMM test is still close to the diagonal line in Case III (Uniform distribution), even almost coinciding; although the counterpart of the score test deviates from the diagonal lines again, the deviation degree is smaller than that in Case II, and the deviation direction is contrary to that in Case II. (iv) It is found from the fourth column figures that in Case IV (mixed symmetric distribution), the performance of the score test is somewhat worse than that of the BGMM test.

- **DGP-II (without intercept):** (i) The first column figures in Figure 2 show that in Case I (normality), the plot for the score test obviously deviates from the diagonal line; it lies below the diagonal line corresponding to inflated type I errors as $\alpha \in (0, 0.5)$, and lies up the diagonal line when $\alpha \in (0.5, 1)$ corresponding to deflated type I errors. (ii) The pattern in the remaining cases is the same as in Figure 1.

For the powers of the two tests in finite samples, the starting points (i.e., $\sigma_{u,o} = 0$) of all plots in Figures 3–4 should correspond to 5% of type I error, because we default $\alpha = 0.05$. Our main findings are summarized as follows. (i) The first column of Figure 3 suggests that both tests have similar empirical powers only in Case I of DGP-I. (ii) The other parts in Figure 3 tell us that the score test rejects too often in Case II and slightly often in Case IV and rejects not enough in Case III. (iii) Figure 4 summarizes that as the intercept does not exist in the SARSF model, the score test is more likely to over-reject in Cases I, II, and IV and to reject not enough in Cases III.

In summary, our GMM is robust against the violation of normality (symmetry still holds), and the score test strongly relies on normality. Moreover, our GMM test is satisfactory no matter whether the intercept exists in the SARSF model, while the score test relies on the existence of the intercept.

7 Conclusion

This paper mainly pays attention to estimating a SARSF model within the GMM framework, whose inefficiency term follows the half-normal distribution. The distributional specification regarding the disturbance term is relaxed to symmetry around zero, and our model, therefore, is partially distribution-free. To establish the asymptotic theory, we first generalize the CLT in Kelejian and Prucha (2001) designed for linear-quadratic processes to the cubic case. As $\sigma_{u,o} \neq 0$, GMMEs of all model parameters is $\sqrt{n}$-consistent. As $\sigma_{u,o} = 0$, the convergence rate of BGMME for $\sigma_{u,o}$ is $n^{1/6}$, and BGMMEs for the remaining parameters are still $\sqrt{n}$-consistent. Finally, a BGMME-based test statistic is provided for testing the existence of technical inefficiency.

Monte Carlo simulations suggest that in estimating the SARSF model, the BGMME outperforms the C2SLSE, GMME, and OGMME. Regarding whether $\sigma_{u,o} = 0$, our BGMM-based test is more robust against the non-normal symmetric disturbance than the score test (Jin and Lee, 2020), and does not rely upon whether the intercept exists in the SARSF model.

The half-normal distribution for the inefficiency term in the SARSF model is a single-parameter setting. Stevenson (1980) thought the half-normal density is most concentrated near zero, which may limit the model's explanatory power. Greene (1990) pointed out that it is a bit inflexible relative to a two-parameter distribution. Therefore, we are also interested in considering other SARSF models where the inefficiency term follows the truncated normal distribution (Stevenson, 1980) or the Gamma counterpart (Greene, 1990).

Appendix A Useful Lemmas

Recall that $\theta_{1,*} = [\hat{\lambda}_{m-sar,n}, \hat{\beta}_{m-sar,n}^\top, 0, \hat{\sigma}_{m-sar,n}^2]^\top$, where $\hat{\lambda}_{m-sar,n}$, $\hat{\beta}_{m-sar,n}^\top$, and $\hat{\sigma}_{m-sar,n}^2$ is the MLE of the SAR model $Y_n = \lambda W_n Y_n + X_n \beta + V_n$. Denote $\hat{s}_n^2 = \frac{1}{n}\mathbf{e}_{1,*}^\top \mathbf{e}_{1,*}$ as the sample variance of the residuals $\mathbf{e}_{1,*} = S_n(\hat{\lambda}_{m-sar,n})Y_n - X_n \hat{\beta}_{m-sar,n}^\top$.

Lemma A.1. *0 is a eigenvalue of the empirical Hessian matrix of* (2.2) *at the point* $\theta = \theta_{1,*}$*, i.e.,* $H_n(\theta_{1,*})$*, and the corresponding eigenvalue is* $[0, \sqrt{2/\pi}\hat{s}_n, \mathbf{0}_{k_x-1}^\top, 1, 0]^\top$.

Proof. By the second-order derivatives, as given by Jin and Lee (2020, p. 27–28), it follows that

$$H_n(\theta_{1,*}) = \begin{bmatrix} h_{11,n} & -\hat{s}_n^{-2}(W_nY_n)^\top X_n & \hat{s}_n^{-1}\sqrt{2/\pi}(W_nY_n)^\top \mathbf{1}_n & -\hat{s}_n^{-4}(W_nY_n)^\top \mathbf{e}_{1,*} \\ -\hat{s}_n^{-2}X_n^\top W_nY_n & -\hat{s}_n^{-2}X_n^\top X_n & \hat{s}_n^{-1}\sqrt{2/\pi}X_n^\top \mathbf{1}_n & -\hat{s}_n^{-4}X_n^\top \mathbf{e}_{1,*} \\ \hat{s}_n^{-1}\sqrt{2/\pi}\mathbf{1}_n^\top W_nY_n & \hat{s}_n^{-1}\sqrt{2/\pi}\mathbf{1}_n^\top X_n & -2n/\pi & (1/2)\hat{s}_n^{-3}\sqrt{2/\pi}\mathbf{1}_n^\top \mathbf{e}_{1,*} \\ -\hat{s}_n^{-4}\mathbf{e}_{1,*}^\top W_nY_n & -\hat{s}_n^{-4}\mathbf{e}_{1,*}^\top X_n & (1/2)\hat{s}_n^{-3}\sqrt{2/\pi}\mathbf{e}_{1,*}^\top \mathbf{1}_n & -2n\hat{s}_n^{-4} \end{bmatrix},$$

where $h_{11,n} = -\mathrm{tr}[G_n(\lambda_{1,*})^2] - \hat{s}_n^{-2}(W_nY_n)^\top W_nY_n$. Denote $H_n(\theta_{1,*}) = [\mathbf{v}_1^\top, \mathbf{v}_2^\top, \mathbf{v}_3^\top, \mathbf{v}_4^\top, \mathbf{v}_5^\top]^\top$, where $\mathbf{v}_2$ is a $(k_x-1)\times(k_x+3)$ matrix, and $\mathbf{v}_l$ for $l = 1, 2, 4, 5$ are (k_x+3)-dimensional row vectors. Since $\mathbf{e}_{1,*}^\top \mathbf{1}_n = 0$, $\mathbf{v}_2$ and $\mathbf{v}_4$ are linear dependent and satisfy $\mathbf{v}_2 = -\sqrt{2/\pi}\hat{s}_n\mathbf{v}_4$. Although generally there exist not any linear dependent among row vectors of $[\mathbf{v}_1^\top, \mathbf{v}_3^\top, \mathbf{v}_5^\top]^\top$, $H_n(\theta_{1,*})$ is non-positive definite, with k_x+2 eigenvalues being negative and the last one being zero. Here it is of interest to only consider the eigenvector, say $\mathbf{d} = [d_1, d_2, \mathbf{d}_3, d_4, d_5]^\top$, corresponding to eigenvalue-0. By the definition of matrix eigenvector, we have $H_n(\theta_{1,*})\mathbf{d} = \mathbf{0}$, i.e.,

$$d_1\mathbf{v}_1 + d_2\mathbf{v}_2 + \mathbf{d}_3\mathbf{v}_3 + d_4\mathbf{v}_4 + d_5\mathbf{v}_5 = 0.$$

Because $rank(H_n(\theta_{1,*})) = k_x+2$, the rank of the solution set to the above equation is equal to one. Recall that $[\mathbf{v}_1^\top, \mathbf{v}_3^\top, \mathbf{v}_5^\top]^\top$ has the full rank k_x+1 and its row vectors are independent linearly of each other and of $\mathbf{v}_2$ and $\mathbf{v}_4$, so there must have $d_1 = \mathbf{d}_3 = d_5 = 0$. Otherwise, by a counter argument it would generate a contradiction to the fact that $[\mathbf{v}_1^\top, \mathbf{v}_3^\top, \mathbf{v}_5^\top]^\top$ has full row rank. By utilizing the relationship $\mathbf{v}_2 = -\sqrt{2/\pi}\hat{s}_n\mathbf{v}_4$, we can take $d_2 = \sqrt{2/\pi}\hat{s}_n$ and $d_4 = 1$. In summary, $\mathbf{d} = [0, \sqrt{2/\pi}\hat{s}_n, \mathbf{0}_{k_x-1}^\top, 1, 0]^\top$. The proof is accomplished. $\square$

For any random variable t with a finite p^{th} absolute moment, where $p \geq 1$, denote its L_p-norm by $\|t\|_p = [\mathrm{E}|t|^p]^{1/p}$. Let $g = \{g_{ni}, i \in D_n, n \geq 1\}$ and $v = \{v_{ni}, i \in D_n, n \geq 1\}$ be two random fields, where D_n satisfies Assumption 1. Assume that $\sup_{i,n}\|g_{ni}\|_p < \infty$. The random field g is said to be L_p-NED on v if

$$\|g_{ni} - \mathrm{E}\left(g_{ni} \mid \mathcal{F}_{ni}(s)\right)\|_p \leq d_{ni}\psi(s)$$

for some arrays of finite positive constants $\{d_{ni}, i \in D_n, n \geq 1\}$ and for some sequence $\psi(s) \geq 0$ such that $\lim_{s\to\infty}\psi(s) = 0$, where $\mathcal{F}_{ni}(s)$ is the σ-field generated by the random variables v_{nj} 's with units j 's located within the ball $B_i(s)$ that is centered at i with radius s, and $\psi(s)$ is the NED coefficient. If we further have $\sup_n \sup_{i\in D_n} d_{ni} < \infty$, then g_{ni} is said to be uniformly L_p-NED on v.

Lemma A.2. *Under Assumptions 1–6 and 9–10, $\sup_{\kappa\in\Theta_\kappa}\mathbf{z}_{i,n}^\top(\kappa_o-\kappa) = \mathbf{w}_{i,n}Y_n(\lambda_o-\lambda) + \mathbf{x}_{i,n}^\top(\beta_o^\dagger - \beta^\dagger)$ (i) is uniformly L_{2r}-norm bounded for some $r > 2$, and (ii) is uniformly L_2-NED on $\{\mathbf{x}_{i,n}, \epsilon_{ni}^\dagger\}$, where Θ_κ is the parameter space of κ.*

Proof. Obviously, Θ_κ is compact under Assumption 6. We first show Y_{ni} and $\mathbf{w}_{i,n}Y_n$ are uniformly L_4-norm bounded. Notice that

$$[t_{ij,n}] := abs(S_n^{-1}) = abs\left[\sum_{i=0}^{\infty}(\lambda_oW_n)^i\right] \leq^* \sum_{i=0}^{\infty} abs\left(\lambda_oW_n\right)^i = [I_n - abs(\lambda_oW_n)]^{-1} := [m_{ij,n}],$$

where $[t_{ij,n}] \leq^* [m_{ij,n}]$ means that $t_{ij,n} \leq m_{ij,n}$ for all i, j, so by the Minkoeski equality,

$$\sup_{i,n}\|y_{ni}\|_{4+\iota} \leq \sup_{i,n}\sum_{j=1}^{n} m_{ij,n}\left(|\beta_{1,o}^\dagger| + \sum_{k=2}^{k_x}|x_{ik,n}|^{4+\iota}|\beta_{k,o}| + \|\epsilon_{ni}^\dagger\|_{4+\iota}\right) < \infty$$

as $\lambda_m \sup_n\|W_n\|_\infty < \infty$, $\sup_{i,n}\|x_{i,n}\|_{4+\iota} < \infty$, and $\sup_{j,n}\|\epsilon_{nj}^\dagger\|_6 < \infty$. So does the $\mathbf{w}_{i,n}Y_n$. Moreover, by the Minkoeski equality

$$\sup_{i,n}\|\mathbf{z}_{i,n}(\kappa_o-\kappa)\|_{4+\iota} = \sup_{i,n}\|\mathbf{w}_{i,n}Y_n(\lambda_o-\lambda) + \mathbf{x}_{i,n}^\top(\beta_o^\dagger - \beta^\dagger)\|_{4+\iota} < \infty,$$

because $\mathbf{z}_{i,n}(\kappa_o-\kappa)$ is linear in $\mathbf{w}_{i,n}Y_n$ and $\mathbf{x}_{i,n}$. As the parameter space Θ_κ is compact, $\sup_{\kappa\in\Theta_\kappa}\mathbf{z}_{i,n}^\top(\kappa_o-\kappa) = \mathbf{w}_{i,n}Y_n(\lambda_o-\lambda) + \mathbf{x}_{i,n}^\top(\beta_o^\dagger - \beta^\dagger)$ (i) is uniformly L_{2r}-norm bounded for $r = 2+\iota/2 > 2$.

We now prove (ii). It is seen from the proof of Xu and Lee (2015, Prop. 1), $\sup_{i,n}\sum_{j:d(i,j)>s} m_{ij,n} \leq cc_1^{s/d_0}$ under Assumption 9(a) and $\sup_{i,n}\sum_{j:d(i,j)>s} m_{ij,n} \leq cs^{-(\alpha-d)}$ under Assumption 9(b), both of which imply

$\lim_{s\to\infty}\sup_{i,n}\sum_{j:d(i,j)>s} m_{n,ij}=0$. Then, working with $\sup_{1\le k\le k_x,i,n}\|x_{ik,n}\|_2<\infty$, it yields that $\{y_{ni}\}_{i=1}^n$ is uniformly L_2-NED on $\{\mathbf{x}_{i,n},\epsilon_{ni}^\dagger\}_{i=1}^n$ by Jenisha and Prucha (2012, Prop. 1). By the proof of Xu and Lee (2015, Prop. 1), $\{\mathbf{w}_{i,n}Y_n\}_{i=1}^n$ also follows from the NED property of $\{y_{ni}\}_{i=1}^n$ with the same NED coefficient as that of the latter. Ditto for $\mathbf{z}_{i,n}(\kappa_o-\kappa)$, because of its linear form in and the compactness of Θ_κ. ☐

Recall that $\epsilon_{ni}=v_{ni}-u_{ni}$ and $\epsilon_{ni}^\dagger=\epsilon_{ni}+\mathrm{E}[u_{ni}]$, where $v_{ni}\sim\mathcal{N}(0,\sigma_{v,o}^2)$ and $u_{ni}\sim\mathcal{N}^+(0,\sigma_{u,o}^2)$.

Lemma A.3. *The first six central moments of the composed error ϵ_{ni}, i.e., the moments around zero of $\epsilon_{ni}^\dagger$, are*

$$\begin{aligned}
&\mu^\dagger=\mathrm{E}[\epsilon_{ni}^\dagger]=0,\quad \mu_2^\dagger=\mathrm{E}\left[\epsilon_{ni}^{\dagger 2}\right]=\sigma_{v,o}^2+(1-2/\pi)\sigma_{u,o}^2,\quad \mu_3^\dagger=\mathrm{E}\left[\epsilon_{ni}^{\dagger 3}\right]=(1-4/\pi)\sqrt{2/\pi}\sigma_{u,o}^3,\\
&\mu_4^\dagger=\mathrm{E}\left[\epsilon_{ni}^{\dagger 4}\right]=3\left[\sigma_{v,o}^2+(1-2/\pi)\sigma_{u,o}^2\right]^2-4\left(1/\pi+3/\pi^2\right)\sigma_{u,o}^4,\\
&\mu_5^\dagger=\mathrm{E}\left[\epsilon_{ni}^{\dagger 5}\right]=10\sqrt{2/\pi}\,(1-4/\pi)\,\sigma_{v,o}^2\sigma_{u,o}^2+\sqrt{2/\pi}\left(7-20/\pi-16/\pi^2\right)\sigma_{u,o}^5,\ \text{and}\\
&\mu_6^\dagger=\mathrm{E}\left[\epsilon_{ni}^{\dagger 6}\right]=15\left[\sigma_{v,o}^2+(1-2/\pi)\sigma_{u,o}^2\right]^3+120\left(1/\pi-3/\pi^2\right)\sigma_{v,o}^2\sigma_{u,o}^4+4\left(21/\pi-70/\pi^2+20/\pi^3\right)\sigma_{u,o}^6.
\end{aligned}$$

It follows that the variances of $\epsilon_{ni}^\dagger$, $\epsilon_{ni}^{\dagger 2}$, and $\epsilon_{ni}^{\dagger 3}$ are, respectively,

$$\begin{aligned}
&\mathrm{Var}[\epsilon_{ni}^\dagger]=\mu_2^\dagger,\quad \mathrm{Var}\left[\epsilon_{ni}^{\dagger 2}\right]=2\left[\sigma_{v,o}^2+(1-2/\pi)\sigma_{u,o}^2\right]^2-4\left(1/\pi+3/\pi^2\right)\sigma_{u,o}^4,\ \text{and}\\
&\mathrm{Var}\left[\epsilon_{ni}^{\dagger 3}\right]=15\left[\sigma_{v,o}^2+(1-2/\pi)\sigma_{u,o}^2\right]^3+120\left(1/\pi-3/\pi^2\right)\sigma_{v,o}^2\sigma_{u,o}^4+2\left(41/\pi-132/\pi^2+24/\pi^3\right)\sigma_{u,o}^6.
\end{aligned}$$

Proof. These results can be straightly derived from the expansion

$$\epsilon_{ni}^{\dagger r}=(\epsilon_{ni}-\mathrm{E}[\epsilon_{ni}])^r=[v_{ni}-(u_{ni}-\mathrm{E}[u_{ni}])]^r=\sum_{i=0}^r C_r^i v_{ni}^i\left(u_{ni}+\sqrt{1-2/\pi}\sigma_{u,o}\right)^{r-i}$$

for some positive inter r, and some moments of v_{ni} and u_{ni}, where $C_r^i=r!/[i!(r-i)!]$ is the combinatorial number. Here we omit these tedious calculations. ☐

Lemma A.4. *Let $A_n=(A_{n,ij})$ and $B_n=(B_{n,ij})$ be two arbitrary n-dimensional square matrices, and Q_n be the $n\times k_q$ IV matrix. Then*

$$\mathrm{E}\left[\epsilon_n^{\dagger\top}A_n\epsilon_n^\dagger\cdot\sum_{i=1}^n\epsilon_{ni}^{\dagger 3}\right]=[\mu_5^\dagger+(n-1)\mu_2^\dagger\mu_3^\dagger]\cdot\mathrm{tr}(A_n),\quad \mathrm{E}\left[Q_n^\top\epsilon_n^\dagger\cdot\sum_{i=1}^n\epsilon_{ni}^{\dagger 3}\right]=\mu_4^\dagger\cdot Q_n^\top\mathbf{1}_n,$$

and $\mathrm{E}\left[\epsilon_n^{\dagger\top}A_n\epsilon_n^\dagger\cdot\epsilon_n^{\dagger\top}B_n\epsilon_n^\dagger\right]=4(1/\pi+3/\pi^2)\sigma_{u,o}^4\mathrm{vec}_D^\top(A_n)\mathrm{vec}_D(B_n)+\mu_2^\dagger\left[\mathrm{tr}(A_n)\mathrm{tr}(B_n)+\mathrm{tr}\left(A_nB_n^s\right)\right]$.

Proof. By direct calculation, we can obtain

$$\begin{aligned}
\mathrm{E}\left[\epsilon_n^{\dagger\top}A_n\epsilon_n^\dagger\cdot\sum_{i=1}^n\epsilon_{ni}^{\dagger 3}\right]&=\mathrm{E}\left[\sum_{i,j=1}^n A_{n,ij}\epsilon_{ni}^\dagger\epsilon_{nj}^\dagger\sum_{k=1}^n\epsilon_{nk}^{\dagger 3}\right]=\mathrm{E}\left[\sum_{i=1}^n A_{n,ii}\epsilon_{ni}^{\dagger 5}+\sum_{i=1}^n A_{n,ii}\epsilon_{ni}^{\dagger 2}\sum_{k=1,k\neq i}^n\xi_{nk}^3\right]\\
&=\mu_5^\dagger\cdot\mathrm{tr}(A_n)+(n-1)\mu_2^\dagger\mu_3^\dagger\cdot\mathrm{tr}(A_n)\quad\text{and}
\end{aligned}$$

$$\mathrm{E}\left[Q_n^\top\epsilon_n^\dagger\cdot\sum_{i=1}^n\epsilon_{ni}^{\dagger 3}\right]=\mathrm{E}\left[\sum_{i=1}^n Q_{n,i}\epsilon_{ni}^\dagger\sum_{j=1}^n\epsilon_{nj}^{\dagger 3}\right]=\mathrm{E}\left[\sum_{i=1}^n Q_{n,i}\epsilon_{ni}^{\dagger 4}\right]=\mu_4^\dagger\cdot Q_n^\top\mathbf{1}_n.$$

The last can be found in Lee (2001, Lemma 2.1). ☐

Lemma A.5. *Let Q_n be $[G_nX_n\beta_o,X_n]$. Then $|\delta_q|\le 1$, where $\delta_q=\frac{1}{n}\mathbf{1}_n^\top Q_n(Q_n^\top Q_n)^{-1}Q_n^\top\mathbf{1}_n$.*

Proof. Define the projection matrix $P_q=Q_n(Q_n^\top Q_n)^{-1}Q_n^\top$. Under Assumptions 7 and 4, P_q is uniformly bounded in both row and column sums (Lee, 2001, Lemma A.10). By utilizing that P_q is symmetric and

$rank(P_q) = k_x + 1$, consider its orthogonal decomposition, i.e., there exists an orthogonal matrix M_n and a diagonal matrix Λ_n such that $P_q = M_n \Lambda_n M_n^\top$, where the diagonal element of Λ_n is either one $(k_x + 1)$ or zero $(n - k_x - 1)$. So we have $\mathbf{1}_n^\top P_q \mathbf{1}_n = \mathbf{1}_n^\top M_n \Lambda_n M_n^\top \mathbf{1}_n$. Let $\mathbf{m}_n = M_n^\top \mathbf{1}_n = [m_1, \cdots, m_n]$, then $\|\mathbf{m}_n\|_2 = \|M_n^\top \mathbf{1}_n\|_2 = \|\mathbf{1}_n\|_2 = \sqrt{n}$ by the orthogonality of M_n. It follows that

$$|\mathbf{1}_n^\top P_q \mathbf{1}_n| = \|\mathbf{m}_n^\top \Lambda_n \mathbf{m}_n\|_2 = \sum_{i=1}^{k_x+1} m_i^2 \leq \|\mathbf{m}_n\|_2^2 = n$$

by the property of Λ_n. The proof is accomplished.

□

Lemma A.6. *(Lee, 2007a,* Lemmas A.3–A.4*) Suppose that A_n are uniformly bounded in both row and column sums in absolute value. The ε_{ni}'s are i.i.d. with zero-mean, and its fourth moment exists. Then,* $\mathrm{E}[\varepsilon_n^\top A_n \varepsilon_n] = O(n)$, $\mathrm{Var}[\varepsilon_n^\top A_n \varepsilon_n] = O(n)$, $\varepsilon_n^\top A_n \varepsilon_n = O_p(n)$, *and* $\frac{1}{n}\varepsilon_n^\top A_n \varepsilon_n - \frac{1}{n}\mathrm{E}[\varepsilon_n^\top A_n \varepsilon_n] = o_p(1)$.

Suppose that elements of the $n \times k$ matrix C_n are uniformly bounded. Then $(1/\sqrt{n})C_n^\top A_n \varepsilon_n = O_p(1)$, $\frac{1}{n}C_n^\top A_n \varepsilon_n = o_p(1)$. *Furthermore, if the limit of $\frac{1}{n}C_n^\top A_n A_n^\top C_n$ exists and is positive definite, then* $(1/\sqrt{n})C_n^\top A_n \varepsilon_n \xrightarrow{d} \mathcal{N}(0, \mathrm{Var}[\varepsilon_{ni}] \lim_{n\to\infty} \frac{1}{n}C_n^\top A_n A_n^\top C_n)$.

Lemma A.7. *(Lee, 2007a,* Lemma A.6*) Suppose that $\frac{1}{n}(\mathcal{Q}_n(\theta) - \bar{Q}_n(\theta))$ converges in probability to zero uniformly in $\theta \in \Theta$ which is a convex set, and $\{\frac{1}{n}\bar{Q}_n(\theta)\}$ satisfies the identification uniqueness condition at θ_o. Let θ_n and θ_n^* be, respectively, the minimizers of $\mathcal{Q}_n(\theta)$ and $\mathcal{Q}_n^*(\theta)$. If $\frac{1}{n}(\bar{Q}_n(\theta) - \mathcal{Q}_n^*(\theta)) = o_p(1)$ uniformly in $\theta \in \Theta$, then both θ_n and θ_n^* converge in probability to θ_o.*

In addition, suppose that $\frac{1}{n}\partial^2 Q_n(\theta)/\partial\theta\partial\theta^\top$ converges in probability to a well-defined limiting matrix, uniformly in $\theta \in \Theta$, which is nonsingular at θ_o, and $(1/\sqrt{n})\partial Q_n(\theta)/\partial\theta = O_p(1)$. If $\frac{1}{n}(\partial^2 Q_n^*(\theta)/\partial\theta\partial\theta^\top - \partial^2 Q_n(\theta)/\partial\theta\partial\theta^\top)$ uniformly in $\theta \in \Theta$ and $\frac{1}{n}(\partial Q_n^*(\theta)/\partial\theta - \partial Q_n(\theta)/\partial\theta) = o_p(1)$, then $\sqrt{n}(\theta_n^* - \theta_o)$ and $\sqrt{n}(\theta_n - \theta_o)$ have the same limiting distribution.

Let $\{X_{i,n}, 1 \leq i \leq k_n, n \geq 1\}$ $(k_n \to \infty$ as $n \to \infty)$ be an array of random variables defined on a probability space $(\Omega, \mathfrak{F}; P)$ with $\mathrm{E}|X_{i,n}| < \infty$. Let $\{\mathfrak{F}_{i,n}, 1 \leq i \leq k_n, n \geq 1\}$ be an array of sub-sigma fields with $\mathfrak{F}_{i-1,n} \subseteq \mathfrak{F}_{i,n}$ $(\mathfrak{F}_{0,n} = \{\varnothing, \Omega\})$. We then call $\{X_{i,n}, \mathfrak{F}_{i,n}, 1 \leq i \leq k_n, n \geq 1\}$ a martingale difference array if $X_{i,n}$ is $\mathfrak{F}_{i,n}$-measurable and $\mathrm{E}[X_{i,n}|\mathfrak{F}_{i,n}] = 0$.

Lemma A.8. *(Kelejian and Prucha, 2001, Lemma A.1) Let $\{X_{i,n}, \mathfrak{F}_{i,n}, 1 \leq i \leq k_n, n \geq 1\}$ be a square integrable martingale difference array. Suppose that for all $\delta > 0$,*

$$\sum_{i=1}^{k_n} \mathrm{E}\left[X_{i,n}^2 I_{\{|X_{i,n}|>\delta\}} \mid \mathfrak{F}_{i-1,n}\right] \xrightarrow{p} 0 \tag{A.1}$$

$$\text{and} \qquad \sum_{i=1}^{k_n} \mathrm{E}\left[X_{i,n}^2 \mid \mathfrak{F}_{i-1,n}\right] \xrightarrow{p} 1. \tag{A.2}$$

Then, $\sum_{i=1}^{k_n} X_{i,n} \xrightarrow{d} \mathcal{N}(0,1)$.

Appendix B Proofs of main results

Proof of Proposition 1: Within the GMM framework, the identification condition means the uniqueness of the solution $\theta^\dagger = \theta_o^\dagger$ to the equation $a_o \lim_{n\to\infty} \frac{1}{n}\mathrm{E}[g_n(\theta^\dagger)] = 0$. It follows from Assumption 8 that equation $a_o x = 0$ has a unique root at $x = 0$. Therefore, the proof remains to be accomplished by proving that $\lim_{n\to\infty} \frac{1}{n}\mathrm{E}[g_n(\theta^\dagger)] = 0$ is uniquely zero at $\theta^\dagger = \theta_o^\dagger$.

For any possible value $\theta^\dagger \in \Theta$,

$$\mathrm{E}[g_{1n}(\kappa)] = \begin{bmatrix} d_n^\top(\kappa)P_{1n}d_n(\kappa) + \mu_2^\dagger \cdot \mathrm{tr}(S_n^{-\top}S_n^\top(\lambda)P_{1n}S_n(\lambda)S_n^{-1}) \\ \vdots \\ d_n^\top(\kappa)P_{1n}d_n(\kappa) + \mu_2^\dagger \cdot \mathrm{tr}(S_n^{-\top}S_n^\top(\lambda)P_{k_pn}S_n(\lambda)S_n^{-1}) \\ Q_n^\top d_n(\kappa) \end{bmatrix}$$
$$\mathrm{E}[g_{2n}(\theta^\dagger)] = f(\kappa) + n\sqrt{2/\pi}(1-4/\pi)(\sigma_{u,o}^3 - \sigma_u^3),$$

where $f(\kappa_o) = 0$,[13] $\mu_2^\dagger = \sigma_{v,o}^2 + (1-2/\pi)\sigma_{u,o}^2$ and

$$d(\kappa) = \mathbf{1}_n(\beta_{1,o} - \beta_1^\dagger) + X_{-1,n}(\beta_{-1,o} - \beta_{-1}) + (\lambda_o - \lambda)G_nX_n\beta_o^\dagger. \tag{B.1}$$

Obviously, $\lim_{n\to\infty} \frac{1}{n}\mathrm{E}[g_n(\kappa_o)] = 0$ since $d(\kappa_o) = 0$. We now prove the uniqueness. The equation $\lim_{n\to\infty} \frac{1}{n}Q_n^\top d_n(\theta^\dagger) = 0$ is equivalent to

$$\lim_{n\to\infty} \frac{1}{n}\left[Q_n^\top G_nX_n\beta_o^\dagger, Q_n^\top \mathbf{1}_n, Q_n^\top X_{-1,n}\right] \cdot \begin{bmatrix} \lambda_o - \lambda \\ \beta_{1,o}^\dagger - \beta_1^\dagger \\ \beta_{-1,o} - \beta_{-1} \end{bmatrix} = 0.$$

Notice that the coefficient matrix has full column ranks in Assumption 8(a), so the above has the unique root vector where $\lambda = \lambda_o$, and $\beta^\dagger = \beta_o^\dagger = [\beta_{1,o}^\dagger, \beta_{-1,o}^\top]^\top$. The identification of $\beta_{1,o}^\dagger = \beta_{1,o} - \sqrt{2/\pi}\sigma_{u,o}$ does not imply that of $\beta_{1,o}$ and $\sqrt{2/\pi}\sigma_{u,o}$. After κ_o and λ_o are identified, the second moment equation $\lim_{n\to\infty} \frac{1}{n}\mathrm{E}[g_{2n}(\theta_o^\dagger)] = 0$ simplifies to $\sqrt{2/\pi}(1-4/\pi)(\sigma_{u,o}^3 - \sigma_u^3) = 0$, which is uniquely zero at $\sigma_u^3 = \sigma_{u,o}^3$. Then, $\beta_{1,o}$ is identifiable.

Sometimes there is multicollinearity among column vectors consisting of X_n and $G_nX_n\beta_o^\dagger$, such as in the scenario where any element of $\beta_{-1,o}$ is zero, and other cases in Kelejian and Prucha (1998). In these cases, Assumption 8(a) do not hold, and there exists some $c \in \mathbb{R}^{k_x}$ such that $G_nX_n\beta_o^\dagger = X_nc$ from Assumption 4, which indicates the modified reduced form (3.1) becomes $Y_n = X_n(\beta_o^\dagger + c\lambda_o) + \zeta_n$ with $\zeta_n = S_n^{-1}\epsilon_n^\dagger$. One can see from Lee (2001) that Assumption 8(b) is sufficient for the identification of λ_o in the pure SAR process $\zeta_n = \lambda_oW_n\zeta_n + \epsilon_n^\dagger$, and here we omit these. The proof is accomplished. ■

Proof of Proposition 2: The consistency of the GMME $\hat\theta_{g,n}^\dagger$ follows from the uniformly convergence that $\sup_{\theta^\dagger\in\Theta}\{(1/n^2)g_n^\top(\theta^\dagger)a_n^\top a_ng_n(\theta^\dagger) - (1/n^2)\mathrm{E}[g_n^\top(\theta^\dagger)]a_n^\top a_n\mathrm{E}[g_n(\theta^\dagger)]\} = o_p(1)$, and the identification condition that $(1/n^2)\mathrm{E}[g_n^\top(\theta^\dagger)]a_n^\top a_n\mathrm{E}[g_n(\theta^\dagger)]$ is uniquely maximized at $\theta_o^\dagger \in \Theta$ (White, 1994). Recall that in Lemma A.2, Θ_κ is the compact parameter space of κ_o.

We first show that $\frac{1}{n}a_ng_n(\theta^\dagger) - \frac{1}{n}a_n\mathrm{E}[g_n(\theta^\dagger)] = o_p(1)$ uniformly in $\theta^\dagger \in \Theta$. Denote $\mathbf{a}_{i,n} = [a_{i1,n}, a_{i2,n}, \cdots, a_{ik_p,n}, \mathbf{a}_{ik_q,n}, a_{ipq,n}]$ as the i^{th} row of a_n, where $a_{ij,n}$, $j = 1, \cdots, k_p$, and $a_{ipq,n}$ are scalars, and $\mathbf{a}_{ik_q,n}$ is a $1 \times k_q$ row subvector. It is sufficient to consider the uniform convergence of $\frac{1}{n}\mathbf{a}_{i,n}g_n(\theta^\dagger)$ for each i. Notice that

$$\frac{1}{n}\mathbf{a}_{i,n}g_n(\theta^\dagger) = \frac{1}{n}\epsilon_n^{\dagger\top}(\kappa)\left(\sum_{j=1}^{k_p} a_{ij,n}P_{nj}\right)\epsilon_n^\dagger(\kappa) + \frac{1}{n}\mathbf{a}_{ik_q,n}Q_n^\top\epsilon_n^\dagger(\kappa) + \frac{1}{n}a_{ipq,n}g_{2n}(\theta^\dagger) \tag{B.2}$$

and $\epsilon_n^\dagger(\kappa) = d_n(\kappa) + \epsilon_n^\dagger + (\lambda_o - \lambda)G_n\epsilon_n^\dagger$, where $d_n(\kappa) = X_n(\beta_o^\dagger - \beta) + (\lambda_o - \lambda)G_nX_n\beta_o^\dagger$ is a rewriting form of (B.1).

For the first term on the right-hand side of (B.2), it follows that

$$\frac{1}{n}\epsilon_n^{\dagger\top}(\kappa)\left(\sum_{j=1}^{k_p} a_{ij,n}P_{nj}\right)\epsilon_n^\dagger(\kappa) = \frac{1}{n}d_n^\top(\kappa)\left(\sum_{j=1}^{k_p} a_{ij,n}P_{nj}\right)d_n(\kappa) + \frac{1}{n}l_n(\kappa) + \frac{1}{n}q_n(\kappa),$$

[13]Here we do not need to specify the expression of $f(\kappa)$.

where $l_n(\kappa) = d_n^\top(\kappa)(\sum_{j=1}^m a_{nj}P_{nj}^s)[\epsilon_n^\dagger + (\lambda_o - \lambda)G_n\epsilon_n^\dagger]$ and $q_n(\kappa) = [\epsilon_n^\dagger + (\lambda_o - \lambda)G_n\epsilon_n^\dagger]^\top(\sum_{j=1}^m a_{nj}P_{nj})[\epsilon_n^\dagger + (\lambda_o - \lambda)G_n\epsilon_n^\dagger]$. By expansion,

$$
\begin{aligned}
\frac{1}{n}l_n(\kappa) =& (\lambda_o - \lambda)\frac{1}{n}\left(X_n\beta_o^\dagger\right)^\top G_n^\top\left(\sum_{j=1}^m a_{nj}P_{nj}^s\right)\epsilon_n^\dagger + \left(\beta_o^\dagger - \beta^\dagger\right)^\top\frac{1}{n}X_n^\top\left(\sum_{j=1}^m a_{nj}P_{nj}^s\right)\epsilon_n^\dagger \\
&+ (\lambda_o - \lambda)^2\frac{1}{n}\left(X_n\beta_o^\dagger\right)^\top G_n^\top\left(\sum_{j=1}^m a_{nj}P_{nj}^s\right)G_n\epsilon_n^\dagger \\
&+ (\lambda_o - \lambda)\left(\beta_o^\dagger - \beta^\dagger\right)^\top\frac{1}{n}X_n^\top\left(\sum_{j=1}^m a_{nj}P_{nj}^s\right)G_n\epsilon_n^\dagger
\end{aligned}
$$

by Lemma A.6, uniformly in $\kappa \in \Theta_\kappa$. The uniform convergence is because $l_n(\kappa)$ are quadratic forms in κ and Θ_κ is a bounded set. Similarly, $\frac{1}{n}q_n(\kappa) - \frac{1}{n}\mathrm{E}[q_n(\kappa)] = o_p(1)$ uniformly in $\kappa \in \Theta_\kappa$. So, $\frac{1}{n}\epsilon_n^{\dagger\top}(\kappa)(\sum_{j=1}^{k_p} a_{ij,n}P_{nj})\epsilon_n^\dagger(\kappa) - \frac{1}{n}\mathrm{E}[\epsilon_n^{\dagger\top}(\kappa)(\sum_{j=1}^{k_p} a_{ij,n}P_{nj})\epsilon_n^\dagger(\kappa)] = o_p(1)$ uniformly in $\kappa \in \Theta_\kappa$.

For the second term on the right-hand side of (B.2), it is similar to show that

$$
\begin{aligned}
\frac{1}{n}\mathbf{a}_{ik_q,n}Q_n^\top\epsilon_n^\dagger(\kappa) &= \frac{1}{n}\mathbf{a}_{ik_q,n}Q_n^\top d_n(\kappa) + \frac{1}{n}\mathbf{a}_{ik_q,n}Q_n^\top\epsilon_n^\dagger + \frac{1}{n}(\lambda_o - \lambda)\mathbf{a}_{ik_q,n}Q_n^\top G_n\epsilon_n^\dagger \\
&= \frac{1}{n}\mathbf{a}_{ik_q,n}Q_n^\top d_n(\kappa) + o_p(1),
\end{aligned}
$$

uniformly in $\kappa \in \Theta_\kappa$, whereby $\frac{1}{n}\mathbf{a}_{ik_q,n}Q_n^\top\epsilon_n^\dagger(\kappa) - \frac{1}{n}\mathrm{E}[\mathbf{a}_{ik_q,n}Q_n^\top\epsilon_n^\dagger(\kappa)] = o_p(1)$ uniformly in $\kappa \in \Theta_\kappa$.

For the third term on the right-hand side of (B.2), denote $Z_n = [W_nY_n, X_n]$ with the i^{th} row being $\mathbf{z}_{i,n}^\top = [\mathbf{w}_{i,n}Y_n, \mathbf{x}_{i,n}^\top]$, $i = 1, \cdots, n$. Then, $\epsilon_n^\dagger(\kappa) = \epsilon_n^\dagger + (\lambda_o - \lambda)W_nY_n + X_n(\beta_o^\dagger - \beta^\dagger) = \epsilon_n^\dagger + Z_n(\kappa_o - \kappa)$. By expansion, we have

$$
\begin{aligned}
\frac{1}{n}a_{ipq,n}g_{2n}(\theta^\dagger) =& a_{ipq,n}\left[\frac{1}{n}\sum_{i=1}^n\epsilon_{ni}^{\dagger 3} - \sqrt{\frac{2}{\pi}}(1 - \frac{4}{\pi})\sigma_{u,o}^3\right] + a_{ipq,n}\sqrt{\frac{2}{\pi}}(1 - \frac{4}{\pi})\left(\sigma_{u,o}^3 - \sigma_u^3\right) \\
&+ 3a_{ipq,n}\frac{1}{n}\sum_{i=1}^n\epsilon_{ni}^{\dagger 2}\mathbf{z}_{i,n}^\top(\kappa_o - \kappa) + 3a_{ipq,n}\frac{1}{n}\sum_{i=1}^n\epsilon_{ni}^\dagger\left[\mathbf{z}_{i,n}^\top(\kappa_o - \kappa)\right]^2 \qquad \text{(B.3)} \\
&+ a_{ipq,n}\frac{1}{n}\sum_{i=1}^n\left[\mathbf{z}_{i,n}^\top(\kappa_o - \kappa)\right]^3.
\end{aligned}
$$

Lindeberg-Levý CLT tells us that $\frac{1}{n}\sum_{i=1}^n\epsilon_{ni}^{\dagger 3} - \sqrt{2/\pi}(1 - 4/\pi)\sigma_{u,o}^3 = O_p(n^{-1/2}) = o_p(1)$ under Assumptions 1′ and 3. By Lemma A.2, $\mathbf{z}_{i,n}^\top(\kappa_o - \kappa)$ is uniformly L_{2r}-norm bounded for $r = 2 + \iota/2 > 2$ and uniformly L_2-NED on $\{\mathbf{x}_{i,n}, \epsilon_{ni}^\dagger\}$. So does the $\epsilon_{ni}^\dagger\mathbf{z}_{i,n}^\top(\kappa_o - \kappa)$ since $\sup_{i,n}\|\epsilon_{ni}^\dagger\|_{4+\iota} < \infty$ by Lemma A.3. Thus $3a_{ipq,n}\epsilon_{ni}^{\dagger 2}\mathbf{z}_{i,n}^\top(\kappa_o - \kappa)$ is still uniformly L_{2r}-norm bounded and uniformly L_2-NED on $\{\mathbf{x}_{i,n}, \epsilon_{ni}^\dagger\}$ as $\sup_{i,n}\|\epsilon_{ni}^\dagger\|_{4+\iota} < \infty$. Then $3a_{ipq,n}\frac{1}{n}\sum_{i=1}^n\{\epsilon_{ni}^{\dagger 2}\mathbf{z}_{i,n}^\top(\kappa_o - \kappa) - \mathrm{E}[\epsilon_{ni}^{\dagger 2}\mathbf{z}_{i,n}^\top(\kappa_o - \kappa)]\} = o_p(1)$ by Jenisha and Prucha (2012, Prop. 1), uniformly in $\kappa \in \Theta_\kappa$. By Xu and Lee (2015, Lemma A.2), $[\mathbf{z}_{i,n}^\top(\kappa_o - \kappa)]^2$ is uniformly L_2-NED on $\{\mathbf{x}_{i,n}, \epsilon_{ni}^\dagger\}$ and so is $\epsilon_{ni}^\dagger[\mathbf{z}_{i,n}^\top(\kappa_o - \kappa)]^2$ as $\sup_{i,n}\|\epsilon_{ni}^\dagger\|_{4+\iota} < \infty$. Taking together with $\sup_{i,n}\|\epsilon_{ni}^\dagger[\mathbf{z}_{i,n}^\top(\kappa_o - \kappa)]^2\|_4 \leq \sup_{i,n}\|\epsilon_{ni}^\dagger\|_4 \cdot \|\mathbf{z}_{i,n}^\top(\kappa_o - \kappa)\|_4^2 < \infty$, it follows that $3a_{i\xi,n}\frac{1}{n}\sum_{i=1}^n(\epsilon_{ni}^\dagger[\mathbf{z}_{i,n}^\top(\kappa_o - \kappa)]^2 - \mathrm{E}[\epsilon_{ni}^\dagger[\mathbf{z}_{i,n}^\top(\kappa_o - \kappa)]^2]) = o_p(1)$ by Jenisha and Prucha (2012, Prop. 1), uniformly in $\kappa \in \Theta_\kappa$. Ditto for the last term on the right-hand side of (B.3). To sum up, $\frac{1}{n}\left\{a_{ipq,n}g_{2n}(\theta^\dagger) - \frac{1}{n}\mathrm{E}[a_{ipq,n}g_{2n}(\theta^\dagger)]\right\} = o_p(1)$ uniformly in $\theta^\dagger \in \Theta$.

From above analysis, $\frac{1}{n}a_n\{g_n(\theta^\dagger) - \mathrm{E}[g_n(\theta^\dagger)]\} = o_p(1)$ uniformly in $\theta^\dagger \in \Theta$. As a cubic function in $\theta^\dagger$, $\frac{1}{n}a_ng_n(\theta^\dagger)$ is uniformly equicontinuous on Θ under Assumption 6. This implies that the identification uniqueness condition for $\{(1/n^2)g_n^\top(\theta^\dagger)a_n^\top a_ng_n(\theta^\dagger)$ must be satisfied, otherwise it would generate a contradiction to the conclusion in Lemma 1 by a counter argument. The consistency of $\hat{\theta}_{g,n}$ holds.

$\hat{\beta}^{\dagger}_{1,g,n} + \sqrt{2/\pi}\hat{\sigma}_{u,g,n} \xrightarrow{p} \beta_{1,o}$ follows from

$$\hat{\beta}^{\dagger}_{1,g,n} - (\beta_{1,o} - \sqrt{2/\pi}\hat{\sigma}_{u,o}) + \sqrt{2/\pi}\hat{\sigma}_{u,g,n} - \sqrt{2/\pi}\hat{\sigma}_{u,o} = o_p(1).$$

Recall that $\mathbf{e}_{g,n} = S_n(\hat{\lambda}_{g,n}) - X_n\hat{\beta}_{g,n} = \epsilon^{\dagger}_n - Z_n(\hat{\kappa}_{g,n} - \kappa_o)$ is the generated residual vector, then

$$\begin{aligned}\frac{1}{n}\mathbf{e}^{\top}_{g,n}\mathbf{e}_{g,n} &= \frac{1}{n}\left[\epsilon^{\dagger}_n - Z_n(\hat{\kappa}_{g,n} - \kappa_o)\right]^{\top}\left[\epsilon^{\dagger}_n - Z_n(\hat{\kappa}_{g,n} - \kappa_o)\right] \\ &= \frac{1}{n}\epsilon^{\dagger\top}_n\epsilon^{\dagger}_n - 2\left(\hat{\kappa}_{g,n} - \kappa_o\right)^{\top}\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_{i,n}\epsilon^{\dagger}_{ni} + \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\kappa}_{g,n} - \kappa_o\right)^{\top}\mathbf{z}_{i,n}\mathbf{z}^{\top}_{i,n}\left(\hat{\kappa}_{g,n} - \kappa_o\right).\end{aligned}$$

Lindeberg-Levy CLT tells us that $\frac{1}{n}\sum_{i=1}^{n}\epsilon^{\dagger\top}_n\epsilon^{\dagger}_n = \mu^{\dagger}_2 + O_p(n^{-1/2}) = \sigma^2_{v,o} + (1-2/\pi)\sigma^2_{u,o} + o_p(1)$ under Assumptions $1'$ and 3. Let $h_{n,ij}$ be either ϵ_{ni} or any element of $\mathbf{z}^{\top}_{i,n}$ for $j = 1, 2$. Then by the generalized Hölder's inequality,

$$\sup_{i,n}\mathrm{E}\left[|h_{n,i1}h_{n,i2}|\right] \le \sup_{i,n}\left\{\left(\mathrm{E}h^2_{n,i1}\right)^{1/2}\left(\mathrm{E}h^2_{n,i2}\right)^{1/2}\right\} < \infty,$$

according to Lemma A.2. Thus, by the LLN in Jenisha and Prucha (2012), $\frac{1}{n}\sum_{i=1}^{n}\epsilon_{ni}\mathbf{z}^{\top}_{i,n} = O_p(1)$, $\frac{1}{n}\sum_{i=1}^{n}\mathbf{z}_{i,n}\mathbf{z}^{\top}_{i,n} = O_p(1)$. Working with $\hat{\kappa}_{g,n} - \kappa_o = o_p(1)$, it yields that $\frac{1}{n}\mathbf{e}^{\top}_{g,n}\mathbf{e}_{g,n} = \sigma^2_{v,o} + (1-2/\pi)\sigma^2_{u,o} + o_p(1)$. Hence $\frac{1}{n}\mathbf{e}^{\top}_{g,n}\mathbf{e}_{g,n} - (1-2/\pi)\hat{\sigma}^2_{u,g,n} \xrightarrow{p} \sigma^2_{v,o}$. The proof is accomplished. ■

Proof of Theorem 1: We follow the framework in Kelejian and Prucha (2001). That is, checking the conditions in Lemma A.8. Notice that the condition (A.1) follows from

$$\sum_{i=1}^{k_n}\mathrm{E}\left\{\mathrm{E}\left[X^2_{i,n} \mid \mathfrak{F}_{i-1,n}\right]\right\} \xrightarrow{p} 0, \tag{B.4}$$

so it only remains to check (B.4) and (A.2) (see ibid, p. 240 for details).

Notice $\varepsilon^{\top}_n A_n\varepsilon_n = (1/2)\varepsilon^{\top}_n A^s_n\varepsilon_n$, so without loss of generality, we can reasonably assume that A_n is symmetric. Apparently $\mu_{F,n} = \mu_{\varepsilon,3}\sum_{i=1}^{n} c_{ni}$, because of the independence of the ε_{ni}'s and $\mathrm{E}[\varepsilon_{ni}] = 0$. Let

$$F_n - \mu_{F,n} = \sum_{i=1}^{n} Y_{i,n} = \sum_{i=1}^{n}\left[c_{ni}(\varepsilon^3_{ni} - \mu_{\varepsilon,3}) + 2\varepsilon_{ni}\sum_{j=1}^{i-1}a_{ij,n}\varepsilon_{nj} + b_{ni}\varepsilon_{ni}\right].$$

Consider the σ-fields $\mathfrak{F}_{0,n} = \{\varnothing, \Omega\}$, $\mathfrak{F}_{i,n} = \sigma(\varepsilon_{n1}, \cdots, \varepsilon_{ni-1})$, $1 \le i \le n$. Obviously, $\mathfrak{F}_{i-1,n} \subseteq \mathfrak{F}_{i,n}$, Y_{ni} is $\mathfrak{F}_{i,n}$-measurable, and $\mathrm{E}[Y_{ni}|\mathfrak{F}_{i,n}] = 0$. Therefore, $\{Y_{i,n}, \mathfrak{F}_{i,n}, 1 \le i \le n, n \ge 1\}$ forms a martingale difference array. Consequently $\sigma^2_{F,n} = \sum_{i=1}^{n}\mathrm{E}[Y^2_{i,n}]$, where

$$\begin{aligned}Y^2_{i,n} =& c^2_{ni}(\varepsilon^3_{ni} - \mu_{\varepsilon,3})^2 + 4\varepsilon^2_{ni}\left(\sum_{j=1}^{i-1}a_{ij,n}\varepsilon_{nj}\right)^2 + b^2_{ni}\varepsilon^2_{ni} \\ &+ 4c_{ni}\varepsilon_{ni}(\varepsilon^3_{ni} - \mu_{\varepsilon,3})\sum_{j=1}^{i-1}a_{ij,n}\varepsilon_{nj} + 4b_{ni}\varepsilon^2_{ni}\sum_{j=1}^{i-1}a_{ij,n}\varepsilon_{nj} + 2b_{ni}c_{ni}\varepsilon_{ni}(\varepsilon^3_{ni} - \mu_{\varepsilon,3}).\end{aligned} \tag{B.5}$$

It follows that

$$\mathrm{E}[Y^2_{i,n}] = c^2_{ni}(\mu_{\varepsilon,6} - \mu^2_{\varepsilon,3}) + 4\sigma^4_{\varepsilon}\sum_{j=1}^{i-1}a^2_{ij,n} + b^2_{ni}\sigma^2_{\varepsilon} + 2b_{ni}c_{ni}\mu_{\varepsilon,4}. \tag{B.6}$$

Let $X_{i,n} = Y_{i,n}/\sigma_{F,n}$, then $\{X_{i,n}, \mathfrak{F}_{i,n}, 1 \le i \le n, n \ge 1\}$ also forms a martingale difference array. We now prove that

$$\frac{F_n - \mu_{F,n}}{\sigma_{F,n}} = \sum_{i=1}^{n}X_{i,n} \xrightarrow{d} \mathcal{N}(0,1)$$

by checking that $X_{i,n}$'s satisfy conditions (B.4) and (A.2) under our assumptions.

We consider the case where $0 < \eta \leqslant \frac{1}{2}\min(\eta_1, \eta_2, \eta_3)$, which make Conditions (i)–(iii) simultaneously holds. Under Condition (i) on the ε_{ni}, there exists then some finite constant K_e such that $\mathrm{E}\,|\varepsilon_{ni}|^s \leqslant K_e$ for $s = 1, \cdots, 6$, and $\mathrm{E}\,|\varepsilon_{ni}|^t\,\mathrm{E}\,|\varepsilon_{ni}|^t \leqslant K_e$, and $\mathrm{E}\left|\varepsilon_{ni}^3 - \mu_{\varepsilon,3}\right|^t \leqslant K_e$ for $t \leqslant 2+\eta$. Similarly, under Conditions (ii)–(iii) on the $(a_{ij,n}, b_{ni}, c_{ni})$, there also exists some finite constant K_p such that $\sum_{j=1}^n |a_{ij,n}| \leqslant K_p$, $n^{-1}\sum_{j=1}^n |b_{ni}|^t \leqslant K_p$ and $n^{-1}\sum_{j=1}^n |c_{ni}|^t \leqslant K_p$ for $t \leqslant 2+\eta$. Observe that $\sum_{j=1}^n |a_{ij,n}|^r \leqslant K_p^r$ for $r \geqslant 1$ and therefore, $\sum_{k=1}^n |a_{ik,n}| \cdot |a_{jk,n}| \leqslant [\sum_{k=1}^n a_{ik,n}^2]^{1/2}[\sum_{k=1}^n a_{jk,n}^2]^{1/2} \leqslant K_p^2$ by the Cauchy-Schwarz inequality.

In what follows, let $q = 2+\eta$ and there must exist $p > 1$ such that $1/q + 1/p = 1$. By applying the inequality $|a+b|^q \leq 2^q(|a|^q + |b|^q)$ twice, the triangle inequality, and Hölder's inequality, we have

$$
\begin{aligned}
|Y_{i,n}|^q &= |c_{ni}(\varepsilon_{ni}^3 - \mu_{\varepsilon,3}) + 2\varepsilon_{ni}\sum_{j=1}^{i-1} a_{ij,n}\varepsilon_{nj} + b_{ni}|\varepsilon_{ni}|^q \\
&\leq 2^q\left[|c_{ni}|\cdot|\varepsilon_{ni}^3 - \mu_{\varepsilon,3}| + 2\sum_{j=1}^{i-1}|a_{ij,n}|^{1/p}\cdot|a_{ij,n}|^{1/q}\cdot|\varepsilon_{ni}|\cdot|\varepsilon_{nj}|\right]^q + 2^q|b_{ni}|^q\cdot|\varepsilon_{ni}|^q \\
&\leq 2^{2q}\left[|c_{ni}|^q\cdot|\varepsilon_{ni}^3 - \mu_{\varepsilon,3}|^q + 2^q\left(\sum_{j=1}^{i-1}|a_{ij,n}|^{1/p}\cdot|a_{ij,n}|^{1/q}\cdot|\varepsilon_{ni}|\cdot|\varepsilon_{nj}|\right)^q\right] + 2^q|b_{ni}|^q\cdot|\varepsilon_{ni}|^q \\
&\leq 2^{3q}\left[|c_{ni}|^q\cdot|\varepsilon_{ni}^3 - \mu_{\varepsilon,3}|^q + \left(\sum_{j=1}^{i-1}|a_{ij,n}|^{p/p}\right)^{q/p}\cdot\left(\sum_{j=1}^{i-1}|a_{ij,n}|\cdot|\varepsilon_{ni}|^q\cdot|\varepsilon_{nj}|^q\right)\right] + 2^q|b_{ni}|^q\cdot|\varepsilon_{ni}|^q \\
&\leq 2^{3q}\left[|c_{ni}|^q\cdot|\varepsilon_{ni}^3 - \mu_{\varepsilon,3}|^q + K_p^{q/p}\cdot\sum_{j=1}^{i-1}|a_{ij,n}|\cdot|\varepsilon_{ni}|^q\cdot|\varepsilon_{nj}|^q\right] + 2^q|b_{ni}|^q\cdot|\varepsilon_{ni}|^q.
\end{aligned}
$$

It yields that

$$
\begin{aligned}
&\sum_{i=1}^n \mathrm{E}\left\{\mathrm{E}\left[|Y_{i,n}|^q \mid \mathfrak{F}_{i,n}\right]\right\} \\
\leq &2^{3q}\sum_{i=1}^n\left[|c_{ni}|^q\cdot\mathrm{E}[|\varepsilon_{ni}^3 - \mu_{\varepsilon,3}|^q] + K_p^{q/p}\sum_{j=1}^{i-1}|a_{ij,n}|\cdot\mathrm{E}[|\varepsilon_{ni}|^q]\cdot\mathrm{E}[|\varepsilon_{nj}|^q]\right] + 2^q\sum_{i=1}^n|b_{ni}|^q\mathrm{E}[|\varepsilon_{ni}|^q] \\
\leq &2^{3q}\left(nK_eK_p + nK_eK_p^{1+q/p}\right) + 2^qK_e\sum_{i=1}^n|b_{ni}|^q \\
\leq &n\cdot K_eK_p\left(2^{3q} + 2^{3q}K_p^{q/p} + 2^q\right).
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
0 \leq \sum_{i=1}^n \mathrm{E}\left\{\mathrm{E}[|X_{i,n}|^{2+\eta} \mid \mathfrak{F}_{i,n}]\right\} &= \frac{1}{(\sigma_{F,n}^2)^{1+\eta/2}}\sum_{i=1}^n \mathrm{E}\left\{\mathrm{E}[|Y_{i,n}|^q \mid \mathfrak{F}_{i,n}]\right\} \\
&\leq \frac{1}{n^{\eta/2}}\cdot\frac{1}{(n^{-1}\sigma_{F,n}^2)^{1+\eta/2}}\cdot K_eK_p\left(2^{3q} + 2^{3q}K_p^{q/p} + 2^q\right) \\
&\leq \frac{1}{n^{\eta/2}}\cdot O(1) \to 0 \quad (\text{as } n \to \infty),
\end{aligned}
$$

where the last inequality follows from $0 < \nu \leq n^{-1}\sigma_{F,n}^2$. This proves that condition (B.4) holds.

We now check (A.2). Noticing that the ε_{ni}'s are independent with zero mean, it follows from (B.5) that

$$
\begin{aligned}
&\mathrm{E}[|Y_{i,n}|^2 \mid \mathfrak{F}_{i-1,n}] = \mathrm{E}\left[|Y_{i,n}|^2 \mid \varepsilon_{n1}, \cdots, \varepsilon_{ni-1}\right] \\
=&c_{ni}^2(\mu_{\varepsilon,6} - \mu_{\varepsilon,3}^2) + 4\sigma_\varepsilon^2 \sum_{j=1}^{i-1}\sum_{k=1}^{i-1} a_{ij,n}a_{ik,n}\varepsilon_{nj}\varepsilon_{nk} + 4(c_{ni}\mu_{\varepsilon,4} + b_{ni}\sigma_\varepsilon^2)\sum_{j=1}^{i-1} a_{ij,n}\varepsilon_{nj} + b_{ni}\sigma_\varepsilon^2 + 2b_{ni}c_{ni}\mu_{\varepsilon,4}.
\end{aligned}
$$

Recalling that $\sigma_{F,n}^2 = \sum_{i=1}^n \mathrm{E}[Y_{ni}^2]$ and working with (B.5), it yields that

$$
\begin{aligned}
&\sum_{i=1}^n \mathrm{E}\left[X_{i,n}^2 \mid \mathfrak{F}_{i-1,n}\right] - 1 = \frac{1}{\sigma_{F,n}^2}\sum_{i=1}^n \left\{\mathrm{E}\left[Y_{i,n}^2 \mid \mathfrak{F}_{i-1,n}\right] - \mathrm{E}[Y_{ni}^2]\right\} \\
=&\frac{1}{\sigma_{F,n}^2}\sum_{i=1}^n \left\{4\sigma_\varepsilon^2 \sum_{j=1}^{i-1}\sum_{k=1,k\neq j}^{i-1} a_{ij,n}a_{ik,n}\varepsilon_{nj}\varepsilon_{nk} + 4\sigma_\varepsilon^2\sum_{j=1}^{i-1} a_{ij,n}^2(\varepsilon_{nj}^2 - \sigma_\varepsilon^2) + 4(c_{ni}\mu_{\varepsilon,4} + b_{ni}\sigma_\varepsilon^2)\sum_{j=1}^{i-1} a_{ij,n}\varepsilon_{nj}\right\} \\
\triangleq&\frac{1}{n^{-1}\sigma_{F,n}^2}\left[4H_{1,n} + 4H_{2,n} + 4H_{3,n}\right].
\end{aligned}
$$

Since $0 < \nu \leq n^{-1}\sigma_{F,n}^2$ condition (A.2) holds if $H_{i,n} = o_p(1)$ for $i = 1, 2, 3$. Here $4H_{1,n}$ and $4H_{2,n}$ are fully identical to those in Kelejian and Prucha (2001, p. 245-246). Hence, we omit analysis for both, and the proof will be accomplished only by proving the last $H_{3,n} = o_p(1)$. Observe that

$$
H_{3,n} = \sum_{i=1}^n \varphi_{i,n}\varepsilon_{ni}, \quad \varphi_{i,n} = \frac{1}{n}\sum_{j=i+1}^n a_{ji,n}(c_{nj}\mu_{\varepsilon,4} + b_{nj}\sigma_\varepsilon^2),
$$

where $\varphi_{i,n}\varepsilon_{ni}/|\varphi_{i,n}|$'s are then independent with zero-mean and uniformly integrable under our moment assumptions. Furthermore, we have

$$
\begin{aligned}
\limsup_{n\to\infty}\sum_{i=1}^{n-1}|\varphi_{i,n}| &\leq \limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^n \left(|c_{nj}|\mu_{\varepsilon,4} + |b_{nj}|\sigma_\varepsilon^2\right)\cdot|a_{ji,n}| \\
&\leq K_e \limsup_{n\to\infty}\frac{1}{n}\sum_{j=1}^n(|c_{nj}| + |b_{nj}|)\sum_{i=1}^n|a_{ji,n}| \\
&\leq 2K_eK_p^2 < \infty
\end{aligned}
$$

and

$$
\begin{aligned}
\lim_{n\to\infty}\sum_{i=1}^{n-1}\varphi_{i,n}^2 &\leq \lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n-1}\left[\sum_{j=i+1}^n \left(|c_{nj}|\mu_{\varepsilon,4} + |b_{nj}|\sigma_\varepsilon^2\right)\cdot|a_{ji,n}|\right]^2 \\
&\leq K_e^2\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n-1}\left[\sum_{j=i+1}^n (|c_{nj}| + |b_{nj}|)\cdot|a_{ji,n}|\right]^2 \\
&\leq K_e^2\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n-1}\left[\sum_{j=1}^n(|c_{nj}|\cdot|a_{ji,n}| + |b_{nj}|\cdot|a_{ji,n}|)\right]^2.
\end{aligned}
\tag{B.7}
$$

Observing that

$$
\sum_{j=1}^n|b_{nj}|\cdot|a_{ji,n}| \leq n^{1/q}\left(\frac{1}{n}\sum_{j=1}^n|b_{nj}|^q\right)^{1/q}\cdot\left(\sum_{j=1}^n|a_{ji,n}|^p\right)^{1/p} \leq n^{1/q}K_p^{1+1/q}
$$

by the Hölder's inequality, and so does $\sum_{j=1}^n |c_{nj}| \cdot |a_{ji,n}|$, (B.7) simplifies to

$$\lim_{n\to\infty}\sum_{i=1}^{n-1}\varphi_{i,n}^2 \le K_e^2 \lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}\left(2n^{1/q}K_p^{1+1/q}\right)^2 \le 4K_e^2K_p^{2(1+1/q)}\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^{n}n^{2/q}$$
$$= 4K_e^2K_p^{2(1+1/q)}\lim_{n\to\infty}n^{2/q-1} = 0,$$

where the final "=" follows from $q > 2$. By the weak law of large numbers (for martingale difference arrays) in Davidson (1994, p. 299), $H_{3,n} = o_p(1)$ holds. This demonstrated that also condition (A.2) holds. The proof is accomplished. ■

Proof of Proposition 3: For the asymptotic distribution of $\hat{\theta}_{g,n}^{\dagger}$, one shall be derived by the Taylor expansion of $[\partial g_n^{\top}(\hat{\theta}_{g,n}^{\dagger})/\partial\theta^{\dagger}]a_n^{\top}a_ng_n(\hat{\theta}_{g,n}^{\dagger})$ at $\theta_o^{\dagger}$,

$$\sqrt{n}(\hat{\theta}_{g,n}^{\dagger}-\theta_o^{\dagger}) = -\left[\frac{1}{n}\frac{\partial g_n^{\top}(\hat{\theta}_{g,n}^{\dagger})}{\partial\theta^{\dagger}}a_n^{\top}a_n\frac{\partial g_n^{\top}(\bar{\theta}_n^{\dagger})}{\partial\theta^{\dagger}}\right]^{-1}\frac{1}{n}\frac{\partial g_n^{\top}(\hat{\theta}_{g,n}^{\dagger})}{\partial\theta^{\dagger}}a_n^{\top}\frac{1}{\sqrt{n}}a_ng_n(\theta_o^{\dagger}), \tag{B.8}$$

where $\bar{\theta}_n^{\dagger}$ satisfies $\|\bar{\theta}_n^{\dagger}-\theta_o^{\dagger}\| \le \|\hat{\theta}_{g,n}^{\dagger}-\theta_o^{\dagger}\|$, and therefore $\bar{\theta}^{\dagger}{}_n - \theta_o^{\dagger} = o_p(1)$. On one hand, for the i^{th} row of a_n, one can see from (B.2) that

$$\mathbf{a}_{i,n}g_n(\theta_o^{\dagger}) = a_{ipq,n}\sum_{i=1}^{n}\left(\epsilon_{ni}^{\dagger 3}-\mu_3^{\dagger}\right) + \epsilon_n^{\dagger\top}\left(\sum_{j=1}^{k_p}a_{ij,n}P_{nj}\right)\epsilon_n^{\dagger} + \mathbf{a}_{ik_q,n}Q_n^{\top}\epsilon_n^{\dagger}.$$

Theorem 1 implies that

$$\frac{1}{\sqrt{n}}\mathbf{a}_{i,n}g_n(\theta_o^{\dagger}) = a_{ipq,n}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\epsilon_{ni}^{\dagger 3}-\mu_3^{\dagger}) + \frac{1}{\sqrt{n}}\epsilon_n^{\dagger\top}\left(\sum_{j=1}^{k_p}a_{ij,n}P_{nj}\right)\epsilon_n^{\dagger} + \frac{1}{\sqrt{n}}\mathbf{a}_{ik_q,n}Q_n^{\top}\epsilon_n^{\dagger}$$
$$\xrightarrow{d}\mathcal{N}(0, \lim_{n\to\infty}\frac{1}{n}\mathbf{a}_{i,n}\Omega_n\mathbf{a}_{i,n}^{\top}),$$

whereby we can derive

$$\frac{1}{\sqrt{n}}a_ng_n(\theta_o^{\dagger}) \xrightarrow{d} \mathcal{N}(0, \lim_{n\to\infty}\frac{1}{n}a_n\Omega_na_n^{\top}). \tag{B.9}$$

On the other hand, as $\partial\epsilon_n^{\dagger}(\kappa)/\partial\theta^{\dagger\top} = -[W_nY_n, X_n, 0]$, it follows that $\partial g_{1n}(\kappa_o)/\partial\theta^{\dagger\top} = -[P_{1n}^s\epsilon_n^{\dagger}, \cdots, P_{k_pn}^s\epsilon_n^{\dagger}, Q_n]^{\top}[W_nY_n, X_n, 0]$ and $\partial g_{2n}(\theta_o^{\dagger})/\partial\theta^{\dagger\top} = -3\sum_{i=1}^n[\epsilon_{ni}^{\dagger 2}\mathbf{w}_{i,n}Y_n, \epsilon_{ni}^{\dagger 2}\mathbf{x}_{i,n}^{\top}, \sqrt{2/\pi}(1-4/\pi)\sigma_{u,o}^2]$. In the former, by expansion we have $\epsilon_n^{\dagger\top}P_{nj}^sW_nY_n = \epsilon_n^{\dagger\top}P_{nj}^sG_nX_n\beta_o^{\dagger} + \epsilon_n^{\dagger\top}P_{nj}^sG_n\epsilon_n^{\dagger}$. Lemma A.6 tells us that $\frac{1}{n}\epsilon_n^{\dagger\top}P_{nj}^sG_nX_n\beta_o^{\dagger} = o_p(1)$ and $\frac{1}{n}\epsilon_n^{\dagger\top}P_{nj}^sG_n\epsilon_n^{\dagger} = \frac{1}{n}\mu_2^{\dagger}\text{tr}(P_{nj}^sG_n) + o_p(1)$, so $\frac{1}{n}\epsilon_n^{\dagger\top}P_{nj}^sW_nY_n = \frac{1}{n}\mu_2^{\dagger}\text{tr}(P_{nj}^sG_n) + o_p(1)$. Similarly, $\frac{1}{n}\epsilon_n^{\dagger\top}P_{nj}^sX_n = o_p(1)$, $\frac{1}{n}Q_n^{\top}W_nY_n = \frac{1}{n}Q_n^{\top}G_nX_n\beta_o^{\dagger} + \frac{1}{n}Q_n^{\top}G_n\epsilon_n^{\dagger} = \frac{1}{n}Q_n^{\top}G_nX_n\beta_o^{\dagger} + o_p(1)$. In the latter, as the proof in Lemma A.2 $\{\mathbf{w}_{i,n}Y_n\}$ is at least uniformly L_{2r}-norm bounded for $r > 2$, and uniformly L_2-NED on $\{\mathbf{x}_{i,n}, \epsilon_{ni}^{\dagger}\}$, and so does $\{\epsilon_{ni}^{\dagger 2}\mathbf{w}_{i,n}Y_n\}$ by using $\sup_{i,n}\|\epsilon_{ni}^{\dagger}\|_{4+\iota} < \infty$ for some $\iota > 0$ twice. Then according to Jenisha and Prucha (2012, Prop. 1), $\frac{1}{n}\sum_{i=1}^n\epsilon_{ni}^{\dagger 2}\mathbf{w}_{i,n}Y_n = \frac{1}{n}\sum_{i=1}^n\text{E}[\epsilon_{ni}^{\dagger 2}\mathbf{w}_{i,n}Y_n] + o_p(1) = \frac{1}{n}\text{E}[\mu_2^{\dagger}(G_nX_n\beta_o^{\dagger})^{\top}\mathbf{1}_n + \mu_3^{\dagger}\text{tr}(G_n)] + o_p(1)$. Under Assumption 4, $\text{Var}[\frac{1}{n}\sum_{i=1}^n\epsilon_{ni}^{\dagger 2}\mathbf{x}_{i,n}] = \frac{1}{n}\text{Var}[\epsilon_{ni}^{\dagger 2}]\cdot\frac{1}{n}\|G_nX_n\beta_o\|_2^2 = \frac{1}{n}\cdot O(1) = o(1)$. The Chebyshev LLN tells us that $\frac{1}{n}\sum_{i=1}^n\epsilon_n^{\dagger 2}\mathbf{x}_{i,n} = \frac{1}{n}\mu_2^{\dagger}X_n^{\top}\mathbf{1}_n + o_p(1)$ under Assumptions 1′ and 3. In conclusion,

$$\frac{1}{n}\frac{\partial g_n^{\top}(\hat{\theta}_{g,n}^{\dagger})}{\partial\theta^{\dagger}} = -\frac{1}{n}D_n^{\dagger} + o_p(1). \tag{B.10}$$

Similarly, $\frac{1}{n}\partial g_n^{\top}(\bar{\theta}^{\dagger})/\partial\theta^{\dagger} = -\frac{1}{n}D_n^{\dagger} + o_p(1)$. Substituting (B.9) and (B.10) into (B.8), it yields the asymptotic distribution of $\sqrt{n}(\hat{\theta}_{g,n}^{\dagger}-\theta_o^{\dagger})$.

Finally, notice that $\beta_{1,o} = \beta_{1,o}^{\dagger} + \sqrt{2/\pi}\sigma_{u,o} = \mathbf{c}_{n,k_x+2}^{\top}\theta_o^{\dagger}$, so the limiting distribution of $\sqrt{n}(\hat{\beta}_{1,g,n}^{\dagger} + \sqrt{2/\pi}\hat{\sigma}_{g,n})$ follows from the Delta method. The proof is accomplished. ■

Proof of Proposition 4: The generalized Schwartz inequality implies that the optimal selection for the distance matrix $a_n^\top a_n$ in Proposition 3 is $(\frac{1}{n}\Omega_n)^{-1}$. For consistency, notice that

$$\frac{1}{n}g_n^\top(\theta^\dagger)\hat{\Omega}_n^{-1}g_n(\theta^\dagger) = \frac{1}{n}g_n^\top(\theta^\dagger)\Omega_n^{-1}g_n(\theta^\dagger) + \frac{1}{n}g_n^\top(\theta^\dagger)(\hat{\Omega}_n^{-1} - \Omega_n^{-1})g_n(\theta^\dagger).$$

Under Assumption 13, $a_o = \lim_{n\to\infty}(\frac{1}{n}\Omega_n)^{-1/2}$ exists. As a_o is nonsingular, the identification condition means the existence and uniqueness of the root to the equation $\lim_{n\to\infty}\frac{1}{n}\mathrm{E}[g_n^\top(\theta^\dagger)] = 0$, which has been proved in Lemma 1. In addition, the uniform convergence in probability of $g_n^\top(\theta^\dagger)\Omega_n^{-1}g_n(\theta^\dagger)$ follows by a similar argument in the proof of Proposition 2. It suffices to show that $\frac{1}{n}g_n^\top(\theta^\dagger)(\hat{\Omega}_n^{-1} - \Omega_n^{-1})g_n(\theta^\dagger) = o_p(1)$ uniformly in $\theta^\dagger \in \Theta$, and then $\frac{1}{n}g_n^\top(\theta^\dagger)\hat{\Omega}_n^{-1}g_n(\theta^\dagger)$ could heir the identification and uniform convergence conditions of $g_n^\top(\theta^\dagger)\Omega_n^{-1}g_n(\theta^\dagger)$. Noticing that

$$\left\|\frac{1}{n}g_n^\top(\theta^\dagger)(\hat{\Omega}_n^{-1} - \Omega_n^{-1})g_n(\theta^\dagger)\right\| \le \left\|\frac{1}{n}g_n(\theta^\dagger)\right\|^2 \cdot \left\|\left(\frac{1}{n}\hat{\Omega}_n\right)^{-1} - \left(\frac{1}{n}\Omega_n\right)^{-1}\right\|, \tag{B.11}$$

where the second factor on the right-hand side of (B.11) satisfies

$$\begin{aligned}\left\|\left(\frac{1}{n}\hat{\Omega}_n\right)^{-1} - \left(\frac{1}{n}\Omega_n\right)^{-1}\right\| &= \left\|\left(\frac{1}{n}\Omega_n\right)^{-1}\left(\frac{1}{n}\hat{\Omega}_n - \frac{1}{n}\Omega_n\right)\left(\frac{1}{n}\hat{\Omega}_n\right)^{-1}\right\| \\ &\le \left\|\left(\frac{1}{n}\Omega_n\right)^{-1}\right\| \cdot \left\|\frac{1}{n}\hat{\Omega}_n - \frac{1}{n}\Omega_n\right\| \cdot \left\|\left(\frac{1}{n}\hat{\Omega}_n\right)^{-1}\right\| = o_p(1)\end{aligned}$$

uniformly in $\theta^\dagger \in \Theta$, so it remains to show that $\left\|\frac{1}{n}g_n(\theta^\dagger)\right\| = O_p(1)$ uniformly in $\theta^\dagger \in \Theta$. From the proof of Proposition 2, $\frac{1}{n}(g_n(\theta^\dagger) - \mathrm{E}[g_n(\theta^\dagger)]) = o_p(1)$ uniformly in $\theta^\dagger \in \Theta$. Such uniform convergence indicates $\frac{1}{n}\mathrm{E}[g_n(\theta^\dagger)]$ is continuous on Θ, then it is also uniformly bounded in $\theta^\dagger \in \Theta$ because of the compactness of Θ. It follows that $\|\frac{1}{n}g_n(\theta^\dagger)\| \le \|\frac{1}{n}(g_n(\theta^\dagger) - \mathrm{E}[g_n(\theta^\dagger)])\| + \|\frac{1}{n}\mathrm{E}[g_n(\theta^\dagger)])\| = O_p(1)$ by the triangle inequality. Thus, $\frac{1}{n}g_n^\top(\theta^\dagger)(\hat{\Omega}_n^{-1} - \Omega_n^{-1})g_n(\theta^\dagger) = o_p(1)$ uniformly in $\theta^\dagger \in \Theta$. The consistency of the feasible optimum GMME $\hat{\theta}_{o,n}^\dagger$ follows.

For the limiting distribution, as $\frac{1}{n}\partial g_n^\top(\hat{\theta}_{o,n}^\dagger)/\partial\theta^\dagger = -\frac{1}{n}D_n^\dagger + o_p(1)$ from the proof of Proposition 3, one follows from the expansion

$$\begin{aligned}\sqrt{n}(\hat{\theta}_{o,n}^\dagger - \theta_o) &= -\left[\frac{1}{n}\frac{\partial g_n^\top(\hat{\theta}_{o,n}^\dagger)}{\partial\theta^\dagger}\left(\frac{\hat{\Omega}_n}{n}\right)^{-1}\frac{1}{n}\frac{\partial g_n^\top(\bar{\theta}_n^\dagger)}{\partial\theta^\dagger}\right]^{-1}\frac{1}{n}\frac{\partial g_n^\top(\hat{\theta}_{o,n}^\dagger)}{\partial\theta^\dagger}\left(\frac{\hat{\Omega}_n}{n}\right)^{-1}a_n\frac{1}{\sqrt{n}}g_n(\theta_o^\dagger) \\ &= \left[\frac{D_n^{\top\dagger}}{n}\left(\frac{\Omega_n}{n}\right)^{-1}\frac{D_n^\dagger}{n}\right]^{-1}\frac{D_n^{\top\dagger}}{n}\left(\frac{\Omega_n}{n}\right)^{-1}\frac{1}{\sqrt{n}}g_n(\theta_o^\dagger) + o_p(1).\end{aligned} \tag{B.12}$$

For the overidentification test, from the Taylor expansion and (B.12) we can derive

$$\frac{1}{\sqrt{n}}g_n^\top(\hat{\theta}_{o,n}^\dagger) = \frac{1}{\sqrt{n}}g_n(\theta_o^\dagger) + \frac{1}{n}\frac{\partial g_n^\top(\bar{\theta}_n^\dagger)}{\partial\theta^\dagger}\sqrt{n}(\hat{\theta}_{o,n}^\dagger - \theta_o) = A_n\frac{g_n(\theta_o^\dagger)}{\sqrt{n}} + o_p(1),$$

where $A_n = I_n - (D_n^\dagger/n)[(D_n^{\top\dagger}/n)(\Omega_n/n)^{-1}(D_n^\dagger/n)]^{-1}(D_n^{\top\dagger}/n)(\Omega_n/n)^{-1}$. It follows that

$$\begin{aligned}g_n^\top(\hat{\theta}_{o,n}^\dagger)\hat{\Omega}_n^{-1}g_n(\hat{\theta}_{o,n}^\dagger) &= \frac{g_n^\top(\theta_o^\dagger)}{\sqrt{n}}A_n^\top\left(\frac{\Omega_n}{n}\right)^{-1}A_n\frac{g_n(\theta_o^\dagger)}{\sqrt{n}} + O_p(1) \\ =&\frac{g_n^\top(\theta_o^\dagger)}{\sqrt{n}}\left(\frac{\Omega_n}{n}\right)^{-1/2}(I_{pq} - P_\Omega)\left(\frac{\Omega_n}{n}\right)^{-1/2}\frac{g_n(\theta_o)}{\sqrt{n}} + o_p(1),\end{aligned}$$

where $I_{pq} \in \mathbb{R}^{k_p+k_q+1}$ is the identity matrix and

$$P_\Omega = (\Omega_n/n)^{-1/2}\,(D_n^\dagger/n)\left[(D_n^{\top\dagger}/n)(\Omega_n/n)^{-1}(D_n^\dagger/n)\right]^{-1}(D_n^{\top\dagger}/n)(\Omega_n/n)^{-1/2}$$

satisfies $\mathrm{tr}(P_\Omega) = k_x + 2$. As $(g_n^\top(\theta_o^\dagger)/\sqrt{n})\,(\Omega_n/n)^{-1/2} \xrightarrow{d} \mathcal{N}(0, I_n)$ by $(1/\sqrt{n})g_n^\top(\theta_o^\dagger) \xrightarrow{d} \mathcal{N}(0, (\Omega_n/n)^{-1/2})$, (4.3) follows from the fact $\mathrm{tr}(I_{pq} - P_\Omega) = k_p + k_q + 1 - (k_x + 2)$. ■

Proof of Proposition 5: For simplicity we only consider the case where $G_n X_n \beta_o^\dagger$ is independent linearly of X_n, and the discussion for the multicollinearity case is analogous. It is seen from Proposition 3 that the BGMME derived from $\min_{\theta^\dagger \in \Theta} \mathcal{Q}_n(\theta^\dagger) := g_{b,n}(\theta^\dagger)^\top \Gamma_n^{-1} g_{b,n}(\theta^\dagger)$ is consistent and is distributed as asymptotically normal, where $g_{b,n}(\theta^\dagger) = [\epsilon_n^{\dagger\top}(\theta^\dagger) G_n^d \epsilon_n^\dagger(\theta^\dagger), \epsilon_n^{\dagger\top}(\theta^\dagger) G_n X_n \beta_o^\dagger, \epsilon_n^{\dagger\top}(\theta^\dagger) X_n, g_{2n}(\theta^\dagger)]^\top$, and Γ_n is blocked as

$$\Gamma_n := \left[\begin{array}{c|c} \Gamma_{11,n} & \Gamma_{12,n} \\ \hline \Gamma_{21,n} & n(\mu_6^\dagger - \mu_3^{\dagger 2}) \end{array}\right] \text{ with } \Gamma_{11,n} = \begin{bmatrix} \mu_2^{\dagger 2}\mathrm{tr}(G_n^{ds} G_n) & 0 \\ 0 & \mu_2^\dagger \mathrm{E}[G_n X_n \beta_o^\dagger, X_n]^\top [G_n X_n \beta_o^\dagger, X_n] \end{bmatrix}$$

and $\Gamma_{21,n} = \mu_4^\dagger \mathbf{1}_n^\top [0, G_n X_n \beta_o^\dagger, X_n]$. For the feasible BGMME $\hat{\theta}_{b,n}^\dagger$, one is derived from $\min_{\theta^\dagger \in \Theta} \mathcal{Q}_n^*(\theta^\dagger) = \hat{g}_{b,n}(\theta^\dagger)^\top \hat{\Gamma}_n^{-1} \hat{g}_{b,n}(\theta^\dagger)$, where $\hat{g}_{b,n}(\theta^\dagger) = [\epsilon_n^{\dagger\top}(\theta^\dagger) \hat{G}_n^d \epsilon_n^\dagger(\theta^\dagger), \epsilon_n^{\dagger\top}(\theta^\dagger) \hat{G}_n X_n \beta_o^\dagger, \epsilon_n^{\dagger\top}(\theta^\dagger) X_n, g_{2n}(\theta^\dagger)]^\top$, and $\hat{\Gamma}_n$ is accordingly blocked as

$$\hat{\Gamma}_n := \left[\begin{array}{c|c} \hat{\Gamma}_{11,n} & \hat{\Gamma}_{12,n} \\ \hline \hat{\Gamma}_{21,n} & n(\hat{\mu}_6^\dagger - \hat{\mu}_3^{\dagger 2}) \end{array}\right] \text{ with } \hat{\Gamma}_{11,n} = \begin{bmatrix} \hat{\mu}_2^{\dagger 2}\mathrm{tr}(\hat{G}_n^{ds} \hat{G}_n) & 0 \\ 0 & \hat{\mu}_2^{\dagger 2} [\hat{G}_n X_n \hat{\beta}_n^\dagger, X_n]^\top [\hat{G}_n X_n \hat{\beta}_n^\dagger, X_n] \end{bmatrix}$$

and $\hat{\Gamma}_{21,n} = \hat{\mu}_4^\dagger \mathbf{1}_n^\top [0, \hat{G}_n X_n \hat{\beta}_n^\dagger, X_n]$.

For the consistency of $\hat{\theta}_{b,n}^\dagger$, it follows from $\lim_{n\to\infty} \frac{1}{n}[\hat{g}_n^\top(\theta^\dagger) \hat{\Gamma}_n^{-1} \hat{g}_n(\theta^\dagger) - g_{b,n}(\theta^\dagger)^\top \Gamma_n^{-1} g_{b,n}(\theta^\dagger)] = o_p(1)$, uniformly in $\theta^\dagger \in \Theta$, and we now prove it. Notice that

$$\begin{aligned}
&\frac{1}{n}\hat{g}_n^\top(\theta^\dagger)\hat{\Gamma}_n^{-1}\hat{g}_n(\theta^\dagger) - g_{b,n}^\top(\theta^\dagger)\Gamma_n^{-1} g_{b,n}(\theta^\dagger) \\
=&\frac{1}{n}\hat{g}_n^\top(\theta^\dagger)\hat{\Gamma}_n^{-1}\left[\hat{g}_n(\theta^\dagger) - g_{b,n}(\theta^\dagger)\right] + \frac{1}{n}\hat{g}_n^\top(\theta^\dagger)\Gamma_n^{-1}\left[\hat{g}_n(\theta^\dagger) - g_{b,n}(\theta^\dagger)\right] + \frac{1}{n} g_{b,n}^\top(\theta^\dagger)(\hat{\Gamma}_n^{-1} - \Gamma_n^{-1})\hat{g}_n(\theta^\dagger),
\end{aligned}$$

$\sup_{\theta^\dagger \in \Theta} \frac{1}{n}\|g_{b,n}(\theta^\dagger)\| = O_p(1)$ by the proof of Proposition 4, and $\frac{1}{n}\|\Gamma_n\| = O(1)$, so it suffices to show $\sup_{\theta^\dagger \in \Theta} \frac{1}{n}\|\hat{g}_n(\theta^\dagger) - g_{b,n}(\theta^\dagger)\| = o_p(1)$, $\frac{1}{n}\|\hat{\Gamma}_n - \Gamma_n\| = o_p(1)$, and $\sup_{\theta^\dagger \in \Theta} \frac{1}{n}\|\hat{g}_n(\theta^\dagger)\| = O_p(1)$. By the expressions of $\hat{g}_n(\theta^\dagger)$ and $g_{b,n}(\theta^\dagger)$, we know the last element in the vector of their difference is zero, so the work for proving such vector equals $o_p(1)$ is similar to the counterpart in Lee (2007a, p. 511) and therefore is omitted. For showing $\frac{1}{n}(\hat{\Gamma}_n - \Gamma_n)$, notice that the $(1,1)$ block satisfies $\frac{1}{n}\|\hat{\Gamma}_{11,n} - \Gamma_{11,n}\| = o_p(1)$ (ibid, equivalent to $\frac{1}{n}(\hat{V}_n - V_n) = o_p(1)$ proved in p. 512), so it remains to show $\frac{1}{n}(\hat{\Gamma}_{21,n} - \Gamma_{21,n}) = o_p(1)$ and $\hat{\mu}_6^\dagger - \hat{\mu}_3^{\dagger 2} \xrightarrow{p} \mu_6^\dagger - \mu_3^{\dagger 2}$. The latter follows by the continuous mapping theorem. For the former, it can be formulated as $\frac{1}{n}(\hat{\Gamma}_{21,n} - \Gamma_{21,n}) = \frac{1}{n}\left[0, \hat{\mu}_4^\dagger \mathbf{1}_n^\top \hat{G}_n X_n \hat{\beta}_n^\dagger - \mu_4^\dagger \mathbf{1}_n^\top G_n X_n \beta_o^\dagger, \hat{\mu}_4^\dagger \mathbf{1}_n^\top X_n - \mu_4^\dagger \mathbf{1}_n^\top X_n\right]$. Here by direct calculation, we have $\frac{1}{n}(\hat{\mu}_4^\dagger \mathbf{1}_n^\top X_n - \mu_4^\dagger \mathbf{1}_n^\top X_n) = o_p(1)$, $(\hat{\mu}_4^\dagger - \mu_4^\dagger)\frac{1}{n}\mathbf{1}_n^\top X_n = o_p(1)$ and

$$\begin{aligned}
&\frac{1}{n}\left(\hat{\mu}_4^\dagger \mathbf{1}_n^\top \hat{G}_n X_n \hat{\beta}_n^\dagger - \mu_4^\dagger \mathbf{1}_n^\top G_n X_n \beta_o^\dagger\right) \\
=&(\hat{\mu}_4^\dagger - \mu_4^\dagger)\frac{1}{n}\mathbf{1}_n^\top G_n X_n \beta_o^\dagger + \frac{1}{n}\hat{\mu}_4^\dagger\left(\mathbf{1}_n^\top \hat{G}_n X_n \hat{\beta}_n^\dagger - \mathbf{1}_n^\top G_n X_n \beta_o^\dagger\right) \\
=&o_p(1) + \hat{\mu}_4^\dagger\left[\frac{1}{n}\mathbf{1}_n^\top \hat{G}_n X_n(\hat{\beta}_n^\dagger - \beta_o^\dagger) + \frac{1}{n}\mathbf{1}_n^\top(\hat{G}_n - G_n)X_n\beta_o^\dagger\right] = o_p(1).
\end{aligned}$$

Lastly, $\frac{1}{n}\|\hat{g}_{b,n}(\theta^\dagger)\| \le \frac{1}{n}\|\hat{g}_{b,n}(\theta^\dagger) - g_{b,n}(\theta^\dagger)\| + \frac{1}{n}\|g_{b,n}(\theta^\dagger)\| = o_p(1) + O_p(1) = O_p(1)$, uniformly in $\theta^\dagger \in \Theta$. In summary, $\sup_{\theta^\dagger \in \Theta} \lim_{n\to\infty} \frac{1}{n}[Q_n^*(\theta^\dagger) - Q_n(\theta^\dagger)] = o_p(1)$, and then the consistency of $\hat{\theta}_{b,n}^\dagger$ follows.

For the limiting distribution of $\hat{\theta}_{b,n}^\dagger$, according to Lemma A.7, we need to check whether $\frac{1}{n}[\partial^2 Q_n^*(\theta^\dagger)/\partial\theta^\dagger\partial\theta^{\dagger\top} - \partial^2 Q_n(\theta^\dagger)/\partial\theta^\dagger\partial\theta^{\dagger\top}]$ uniformly in $\theta^\dagger \in \Theta$ and $\frac{1}{n}[\partial Q_n^*(\theta_o^\dagger)/\partial\theta^\dagger - \partial Q_n(\theta_o^\dagger)/\partial\theta^\dagger] = o_p(1)$

hold one by one. To this end, we first compute $\partial g_{b,n}(\theta^\dagger)/\partial\theta^\dagger$ as

$$\frac{\partial g_{b,n}(\theta^\dagger)}{\partial\theta^{\dagger\top}} = \begin{bmatrix} \epsilon_n^{\dagger\top}(\kappa)G_n^{ds}\frac{\partial\epsilon_n^\dagger(\kappa)}{\partial\kappa^\top} & 0 \\ (G_nX_n\beta_o^\dagger)^\top\frac{\partial\epsilon_n^\dagger(\kappa)}{\partial\kappa^\top} & 0 \\ X_n^\top\frac{\partial\epsilon_n^\dagger(\kappa)}{\partial\kappa^\top} & 0 \\ 3\sum_{i=1}^n\epsilon_{ni}^{\dagger2}(\kappa)\frac{\partial\epsilon_n^\dagger(\kappa)}{\partial\kappa^\top} & 3n\sqrt{\frac{2}{\pi}}(1-\frac{4}{\pi})\sigma_u^2 \end{bmatrix},$$

and elements of $\partial^2 g_{b,n}(\theta^\dagger)/\partial\theta^\dagger\partial\theta^{\dagger\top}$ are

$$\frac{\partial^2 g_{b,n}(\theta^\dagger)}{\partial\theta^{\dagger\top}\partial\lambda} = \begin{bmatrix} -(W_nY_n)^\top G_n^{ds}\frac{\partial\epsilon_n^\dagger(\kappa)}{\partial\kappa^\top} & 0 \\ 0 & 0 \\ 0 & 0 \\ -6\sum_{i=1}^n\epsilon_{ni}^\dagger(\kappa)(W_nY_n)^\top\frac{\partial\epsilon_n^\dagger(\kappa)}{\partial\kappa^\top} & 0 \end{bmatrix}, \frac{\partial^2 g_{b,n}(\theta^\dagger)}{\partial\theta^{\dagger\top}\partial\sigma_{u,o}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 6n\sqrt{\frac{2}{\pi}}(1-\frac{4}{\pi})\sigma_u \end{bmatrix},$$

$$\text{and } \frac{\partial^2 g_{b,n}(\theta^\dagger)}{\partial\theta^{\dagger\top}\partial\beta_l} = \begin{bmatrix} -\mathbf{x}_{l,n}^\top G_n^{ds}\frac{\partial\epsilon_n^\dagger(\kappa)}{\partial\kappa^\top} & 0 \\ 0 & 0 \\ 0 & 0 \\ -6\sum_{i=1}^n\epsilon_{ni}^\dagger(\kappa)\mathbf{x}_{l,n}^\top\frac{\partial\epsilon_n^\dagger(\kappa)}{\partial\kappa^\top} & 0 \end{bmatrix},$$

where β_l is the l^{th} element of $\beta^\dagger = [\beta_1^\dagger, \beta_{-1}^\top]^\top$. Apparently, except for $\epsilon_n^{\dagger\top}(\kappa)G_n^{ds}[\partial\epsilon_n^\dagger(\kappa)/\partial\kappa^\top]$, $(G_nX_n\beta_o^\dagger)^\top[\partial\epsilon_n^\dagger(\kappa)/\partial\kappa^\top]$ and $[\partial\epsilon_n^\dagger(\kappa)/\partial\kappa]G_n^{ds}[\partial\epsilon_n^\dagger(\kappa)/\partial\kappa^\top]$, the remaining elements do not refer to the G_n and $G_nX_n\beta_o^\dagger$, which tells us that in the expressions of $\frac{1}{n}(\frac{\partial\hat{g}_n(\theta_o^\dagger)}{\partial\theta^{\dagger\top}} - \frac{\partial g_{b,n}(\theta_o^\dagger)}{\partial\theta^{\dagger\top}})$ and $\frac{1}{n}(\frac{\partial^2\hat{g}_n(\theta^\dagger)}{\partial\theta^\dagger\partial\theta^{\dagger\top}} - \frac{\partial^2 g_{b,n}(\theta^\dagger)}{\partial\theta^\dagger\partial\theta^{\dagger\top}})$, the corresponding elements must be zero. Therefore, their (uniform) convergence only involves the aforementioned three terms, which have been proved in Lee (2007a, p. 511). Similarly, $(1/\sqrt{n})[\hat{g}_{b,n}(\theta_o^\dagger) - g_{b,n}(\theta_o^\dagger)] = o_p(1)$. From these, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\left(\frac{\partial Q_n^*(\theta_o^\dagger)}{\partial\theta} - \frac{\partial Q_n(\theta_o^\dagger)}{\partial\theta}\right) \\
=&2\left\{\frac{g_{b,n}(\theta_o^\dagger)}{\sqrt{n}}\left[\frac{\partial\hat{g}_{b,n}^\top(\theta_o^\dagger)}{\partial\theta^\dagger}\hat{\Gamma}_n^{-1} - \frac{\partial g_{b,n}^\top(\theta_o^\dagger)}{\partial\theta^\dagger}\Gamma_n^{-1}\right] + \frac{\partial\hat{g}_{b,n}^\top(\theta_o^\dagger)}{\partial\theta^\dagger}\hat{\Gamma}_n^{-1}\frac{\hat{g}_{b,n}(\theta_o^\dagger) - g_{b,n}(\theta_o^\dagger)}{\sqrt{n}}\right\} \\
=&2\left\{\frac{g_{b,n}(\theta_o^\dagger)}{\sqrt{n}}\left[\frac{1}{n}\frac{\partial\hat{g}_{b,n}^\top(\theta_o^\dagger)}{\partial\theta^\dagger}\left(\left(\frac{1}{n}\hat{\Gamma}_n\right)^{-1} - \left(\frac{1}{n}\Gamma_n\right)^{-1}\right) + \frac{1}{n}\left(\frac{\hat{g}_{b,n}^\top(\theta_o^\dagger)}{\partial\theta^\dagger} - \frac{\partial g_{b,n}^\top(\theta_o^\dagger)}{\partial\theta^\dagger}\right)\left(\frac{\Gamma_n}{n}\right)^{-1}\right]\right. \\
&\left.+\frac{1}{n}\frac{\partial\hat{g}_{b,n}^\top(\theta_o^\dagger)}{\partial\theta^\dagger}\left(\frac{1}{n}\hat{\Gamma}_n\right)^{-1}\frac{\hat{g}_{b,n}(\theta_o^\dagger) - g_{b,n}(\theta_o^\dagger)}{\sqrt{n}}\right\} = o_p(1),
\end{aligned}$$

and for each component $\theta_l^\dagger$ of $\theta^\dagger$,

$$\begin{aligned}
\frac{1}{n}\left(\frac{\partial^2 Q_n^*(\theta_l^\dagger)}{\partial\theta_l^\dagger\partial\theta^{\dagger\top}} - \frac{\partial^2 Q_n(\theta_l^\dagger)}{\partial\theta_l^\dagger\partial\theta^{\dagger\top}}\right) =& \frac{2}{n}\left[\frac{\hat{g}_{b,n}^\top(\theta^\dagger)}{\partial\theta_l^\dagger}\hat{\Gamma}_n^{-1}\frac{\partial\hat{g}_{b,n}(\theta^\dagger)}{\partial\theta^\dagger} + \hat{g}_{b,n}^\top(\theta^\dagger)\hat{\Gamma}_n^{-1}\frac{\partial^2\hat{g}_{b,n}(\theta^\dagger)}{\partial\theta_l^\dagger\partial\theta^{\dagger\top}}\right] \\
&-\frac{2}{n}\left[\frac{g_{b,n}^\top(\theta^\dagger)}{\partial\theta_l^\dagger}\Gamma_n^{-1}\frac{\partial g_{b,n}(\theta^\dagger)}{\partial\theta^\dagger} + g_{b,n}^\top(\theta^\dagger)\Gamma_n^{-1}\frac{\partial^2 g_{b,n}(\theta^\dagger)}{\partial\theta_l^\dagger\partial\theta^{\dagger\top}}\right] = o_p(1),
\end{aligned}$$

where the last "=" follows by adding and subtracting some same interaction terms. Then the proof is accomplished. ■

Proof of Proposition 6: Notice that the gradient matrix $D_n^\ddagger$ in (4.6) has the full rank $k_x + 2$, which is different from that of $D_n^\dagger$ in (4.2). Except for this point, the remaining analysis is accomplished by a similar argument in the proof of Proposition 5. ■

Proof of Proposition 7: The asymptotic distribution of $(1/\sqrt{n})\sum_{i=1}^n\hat{e}_{b,n1}^3$ follows the equality (5.4), and we now prove it. Applying the mean-value theorem to $\frac{1}{n}\sum_{i=1}^n \hat{e}_{b,n1}^3 = \frac{1}{n}\sum_{i=1}^n e_{ni}(\hat{\xi}_{b,n})$ at ξ_o, it yields that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n \hat{e}_{b,ni}^3 &= \frac{1}{n}\sum_{i=1}^n v_{ni}^3 - \frac{3}{n^{3/2}}\sum_{i=1}^n e_{ni}^2(\bar{\xi}_n)[\mathbf{w}_{i,n}Y_n, \mathbf{x}_{i,n}^\top] \cdot \sqrt{n}(\hat{\xi}_{b,n} - \xi_o) \\ &= \frac{1}{n}\sum_{i=1}^n v_{ni}^3 - \frac{3}{n^{1/2}}\left[\frac{1}{n}\sum_{i=1}^n e_{ni}^2(\bar{\xi}_n)\mathbf{w}_{i,n}Y_n, \frac{1}{n}\sum_{i=1}^n e_{ni}^2(\bar{\xi}_n)\mathbf{x}_{i,n}^\top\right] \cdot \sqrt{n}(\hat{\xi}_{b,n} - \xi_o)\end{aligned}$$

where $\bar{\xi}_n$ satisfies $\|\bar{\xi}_n - \xi_o\| \leq \|\hat{\xi}_{b,n} - \xi_o\|$ so $\bar{\xi}_n - \xi_o = o_p(1)$. By expansion we can obtain

$$\frac{1}{n}\sum_{i=1}^n e_{ni}^2(\bar{\xi}_n)\mathbf{w}_{i,n}Y_n = \frac{1}{n}\sum_{i=1}^n \epsilon_{ni}^{\dagger 2}\mathbf{w}_{i,n}Y_n + \frac{1}{n}\sum_{i=1}^n \left[\mathbf{z}_{i,n}^\top(\xi_o - \bar{\xi}_n)\right]^2 \mathbf{w}_{i,n}Y_n + \frac{2}{n}\sum_{i=1}^n \epsilon_{ni}^\dagger \left[\mathbf{z}_{i,n}^\top(\xi_o - \bar{\xi}_n)\right] \mathbf{w}_{i,n}Y_n.$$

As in the proof of Lemma A.2, any element of $\mathbf{z}_{i,n}^\top = [\mathbf{w}_{i,n}Y_n, \mathbf{x}_{i,n}^\top]$ are uniformly L_{2r}-norm bounded for some $r > 2$, and uniformly L_2-NED on $\{\mathbf{x}_{i,n}, \epsilon_{ni}^\dagger\}$. Based on this, we can further conclude that $\epsilon_{ni}^{\dagger 2}\mathbf{w}_{i,n}Y_n$ is uniformly at least L_4-norm bounded, and uniformly L_2-NED on $\{\mathbf{x}_{i,n}, \epsilon_{ni}^\dagger\}$, thereby $\frac{1}{n}\sum_{i=1}^n \epsilon_{ni}^{\dagger 2}\mathbf{w}_{i,n}Y_n - \mathrm{E}[\frac{1}{n}\mathbf{1}_n^\top G_n X_n\beta_o] = o_p(1)$ by the LLN in Jenisha and Prucha (2012) and the fact $\mathrm{E}[\frac{1}{n}\sum_{i=1}^n \epsilon_{ni}^{\dagger 2}\mathbf{w}_{i,n}Y_n] = \mathrm{E}[\frac{1}{n}\sum_{i=1}^n \epsilon_{ni}^{\dagger 2}\mathbf{w}_{i,n}S^{-1}X_n\beta_o] + \mathrm{E}[\frac{1}{n}\sum_{i=1}^n \epsilon_{ni}^{\dagger 2}\mathbf{w}_{i,n}S^{-1}\epsilon_n^\dagger] = \mathrm{E}[\frac{1}{n}\mathbf{1}_n^\top G_n X_n\beta_o]$. Ditto for $\frac{1}{n}\sum_{i=1}^n z_{ij,n}^\top \mathbf{w}_{i,n}Y_n$ and $\frac{1}{n}\sum_{i=1}^n \epsilon_{ni}^\dagger z_{ij,n}\mathbf{w}_{i,n}Y_n$. Taking with $\xi_o - \bar{\xi}_n = o_p(1)$ together, $\frac{1}{n}\sum_{i=1}^n \left[\mathbf{z}_{i,n}^\top(\xi_o - \bar{\xi}_n)\right]^2 \mathbf{w}_{i,n}Y_n$ and $\frac{1}{n}\sum_{i=1}^n \epsilon_{ni}^\dagger \left[\mathbf{z}_{i,n}^\top(\xi_o - \bar{\xi}_n)\right] \mathbf{w}_{i,n}Y_n$ are $o_p(1)$. Therefore, $\frac{1}{n}\sum_{i=1}^n e_{ni}^2(\bar{\xi}_n)\mathbf{w}_{i,n}Y_n - \mathrm{E}[\frac{1}{n}\mathbf{1}_n^\top G_n X_n\beta_o] = o_p(1)$.

By a similar argument for

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n e_{ni}^2(\bar{\xi}_n)\mathbf{x}_{i,n} = \frac{1}{n}\sum_{i=1}^n [\epsilon_{ni}^\dagger + \mathbf{z}_{i,n}^\top(\xi_o - \bar{\xi}_n)]^2 \mathbf{x}_{i,n} \\ =&\frac{1}{n}\sum_{i=1}^n \epsilon_{ni}^{\dagger 2}\mathbf{x}_{i,n} + \frac{1}{n}\sum_{i=1}^n \left[\mathbf{z}_{i,n}^\top(\xi_o - \bar{\xi}_n)\right]^2 \mathbf{x}_{i,n} + \frac{2}{n}\sum_{i=1}^n \epsilon_{ni}^\dagger \left[\mathbf{z}_{i,n}^\top(\xi_o - \bar{\xi}_n)\right] \mathbf{x}_{i,n},\end{aligned}$$

the last two terms on the right-hand side satisfy $\frac{1}{n}\sum_{i=1}^n \left[\mathbf{z}_{i,n}^\top(\xi_o - \bar{\xi}_n)\right]^2 \mathbf{x}_{i,n} = o_p(1)$ and $\frac{2}{n}\sum_{i=1}^n \epsilon_{ni}^\dagger \left[\mathbf{z}_{i,n}^\top(\xi_o - \bar{\xi}_n)\right] \mathbf{x}_{i,n} = o_p(1)$. Working with $\frac{1}{n}\sum_{i=1}^n \epsilon_{ni}^{\dagger 2}\mathbf{x}_{i,n} - \mathrm{E}[\frac{1}{n}\sigma_{v,o}^2\mathbf{1}_n^\top X_n] = o_p(1)$ by the Chebyshev LLN, it yields that $\frac{1}{n}\sum_{i=1}^n e_{ni}^2(\bar{\xi}_n)\mathbf{x}_{i,n} = \mathrm{E}[\frac{1}{n}\sigma_{v,o}^2\mathbf{1}_n^\top X_n] + o_p(1)$. Therefore, (5.4) holds. Derivation for the limiting distribution of $\sqrt{n}R_n$ is a corollary of Theorem 1. Then (5.5) follows by the Slutsky theorem.

As v_{ni}'s are normal, $\sigma_{R,n}$ simplifies to $(15 - 9\tilde{\delta}_{q,n})\sigma_{v,o}^6$. Notice $\frac{1}{n}\sum_{i=1}^n \hat{e}_{b,ni}^2 \xrightarrow{p} \sigma_{v,o}^2$, so

$$\frac{n\sum_{i=1}^n \hat{e}_{b,ni}^3}{(15 - 9\tilde{\delta}_{q,n})^{1/2}(\sum_{i=1}^n \hat{e}_{b,ni}^2)^{3/2}} - \frac{\sum_{i=1}^n \hat{e}_{b,ni}^3}{\sqrt{n}(15 - 9\tilde{\delta}_{q,n})^{1/2}\sigma_{v,o}^3} = o_p(1).$$

The proof is accomplished. ■

Appendix C Experimental results

Table 1. Experimental results for estimation in DGP-I

$\lambda_o = -0.5$		Case I: $v_{in} \sim \mathcal{N}(0,1)$			Case II: $v_{in} \sim \mathcal{T}(4)$		
$\sigma_{u,o} < \sigma_{v,o}$		$n = 100$	$n = 225$	$n = 400$	$n = 100$	$n = 225$	$n = 400$
λ_o	C2SLS	0.169(0.029)[0.431]	0.110(0.012)[0.279]	0.087(0.008)[0.221]	0.249(0.063)[0.631]	0.163(0.027)[0.418]	0.124(0.016)[0.315]
	OGMM	0.136(0.018)[0.343]	0.087(0.008)[0.221]	0.068(0.005)[0.173]	0.166(0.028)[0.423]	0.105(0.011)[0.266]	0.082(0.007)[0.209]
	BGMM	0.131(0.017)[0.336]	0.086(0.007)[0.225]	0.067(0.005)[0.173]	0.159(0.026)[0.403]	0.103(0.011)[0.263]	0.080(0.006)[0.206]
$\beta_{0,o}$	C2SLS	0.417(0.174)[1.044]	0.363(0.132)[0.902]	0.332(0.113)[0.818]	0.909(0.829)[2.005]	0.871(0.761)[1.965]	0.868(0.754)[1.885]
	OGMM	0.417(0.174)[1.048]	0.362(0.132)[0.901]	0.331(0.113)[0.814]	0.842(0.709)[2.016]	0.826(0.682)[1.972]	0.817(0.668)[1.892]
	BGMM	0.414(0.172)[1.042]	0.362(0.131)[0.900]	0.331(0.112)[0.814]	0.833(0.694)[1.998]	0.822(0.676)[1.966]	0.816(0.666)[1.891]
$\beta_{1,o}$	C2SLS	0.109(0.012)[0.278]	0.070(0.005)[0.179]	0.053(0.003)[0.135]	0.158(0.025)[0.406]	0.106(0.011)[0.266]	0.080(0.006)[0.206]
	OGMM	0.117(0.014)[0.282]	0.071(0.005)[0.180]	0.053(0.003)[0.135]	0.159(0.025)[0.405]	0.106(0.011)[0.266]	0.080(0.006)[0.206]
	BGMM	0.109(0.012)[0.278]	0.070(0.005)[0.180]	0.053(0.003)[0.135]	0.158(0.025)[0.403]	0.105(0.011)[0.264]	0.080(0.006)[0.206]
$\beta_{2,o}$	C2SLS	0.111(0.012)[0.286]	0.070(0.005)[0.178]	0.052(0.003)[0.133]	0.162(0.026)[0.409]	0.105(0.011)[0.268]	0.078(0.006)[0.199]
	OGMM	0.111(0.012)[0.287]	0.071(0.005)[0.182]	0.052(0.003)[0.133]	0.161(0.026)[0.408]	0.104(0.011)[0.268]	0.077(0.006)[0.199]
	BGMM	0.111(0.012)[0.287]	0.070(0.005)[0.182]	0.052(0.003)[0.133]	0.160(0.026)[0.407]	0.104(0.011)[0.265]	0.077(0.006)[0.199]
$\sigma_{u,o}$	C2SLS	0.503(0.253)[1.176]	0.446(0.200)[1.057]	0.411(0.173)[0.969]	1.189(1.418)[2.505]	1.130(1.280)[2.470]	1.118(1.249)[2.389]
		0.291(0.256){0.468}	0.245(0.163){0.426}	0.224(0.116){0.428}	0.968(1.747){0.441}	0.869(1.495){0.435}	0.883(1.441){0.435}
	OGMM	0.505(0.255)[1.171]	0.447(0.200)[1.054]	0.411(0.173)[0.966]	1.080(1.165)[2.458]	1.056(1.115)[2.437]	1.040(1.082)[2.354]
		0.308(0.260){0.470}	0.251(0.164){0.426}	0.224(0.116){0.427}	0.775(1.299){0.444}	0.739(1.204){0.436}	0.742(1.145){0.434}
	BGMM	0.502(0.252)[1.170]	0.446(0.200)[1.054]	0.411(0.173)[0.967]	1.074(1.154)[2.451]	1.053(1.110)[2.436]	1.040(1.081)[2.359]
		0.293(0.254){0.470}	0.247(0.163){0.426}	0.224(0.116){0.426}	0.759(1.277){0.442}	0.730(1.195){0.436}	0.739(1.143){0.434}
$\sigma_{v,o}$	C2SLS	0.142(0.026)[0.361]	0.103(0.013)[0.265]	0.084(0.008)[0.215]	0.490(0.320)[1.270]	0.473(0.282)[1.151]	0.443(0.240)[0.994]
	OGMM	0.145(0.027)[0.361]	0.102(0.013)[0.263]	0.083(0.008)[0.215]	0.481(0.311)[1.213]	0.466(0.274)[1.098]	0.438(0.235)[0.971]
	BGMM	0.141(0.026)[0.361]	0.102(0.013)[0.264]	0.083(0.008)[0.215]	0.481(0.313)[1.214]	0.466(0.274)[1.108]	0.438(0.235)[0.970]

Note: (i) $(\sigma_{u,o}, \sigma_{v,o}) = (0.5, 1)$ in Case I, and $(\sigma_{u,o}, \sigma_{v,o}) = (1, \sqrt{2})$ in Case II".

(ii) In each cell, the black font represents "SD(MSE)[IDR]", and the blue font represents "SD(MSE){Zero − Rate}".

(continued on the next page)

Table 1 (continued from the previous page)

$\lambda_o = -0.5$		Case I: $v_{in} \sim \mathcal{N}(0,1)$			Case II: $v_{in} \sim \mathcal{T}(4)$		
$\sigma_{u,o} = \sigma_{v,o}$		$n = 100$	$n = 225$	$n = 400$	$n = 100$	$n = 225$	$n = 400$
λ_o	C2SLS	0.188(0.035)[0.482]	0.128(0.016)[0.327]	0.096(0.009)[0.245]	0.260(0.068)[0.652]	0.177(0.031)[0.448]	0.136(0.019)[0.345]
	OGMM	0.145(0.021)[0.368]	0.096(0.009)[0.248]	0.072(0.005)[0.182]	0.171(0.029)[0.436]	0.111(0.012)[0.282]	0.084(0.007)[0.216]
	BGMM	0.140(0.020)[0.358]	0.095(0.009)[0.242]	0.071(0.005)[0.181]	0.163(0.027)[0.420]	0.109(0.012)[0.276]	0.083(0.007)[0.215]
$\beta_{0,o}$	C2SLS	0.475(0.256)[1.246]	0.409(0.191)[1.106]	0.341(0.128)[1.008]	0.914(0.864)[2.107]	0.879(0.795)[2.011]	0.921(0.870)[1.998]
	OGMM	0.476(0.258)[1.238]	0.407(0.190)[1.101]	0.340(0.128)[1.007]	0.868(0.807)[2.101]	0.829(0.722)[2.009]	0.826(0.715)[1.987]
	BGMM	0.473(0.256)[1.235]	0.407(0.190)[1.101]	0.340(0.128)[1.006]	0.864(0.800)[2.093]	0.825(0.716)[2.000]	0.825(0.714)[1.986]
$\beta_{1,o}$	C2SLS	0.120(0.014)[0.309]	0.080(0.006)[0.205]	0.059(0.003)[0.150]	0.170(0.029)[0.430]	0.111(0.012)[0.283]	0.084(0.007)[0.212]
	OGMM	0.121(0.015)[0.311]	0.080(0.006)[0.205]	0.059(0.003)[0.149]	0.170(0.029)[0.431]	0.111(0.012)[0.286]	0.083(0.007)[0.212]
	BGMM	0.120(0.014)[0.309]	0.080(0.006)[0.205]	0.059(0.003)[0.150]	0.168(0.029)[0.426]	0.110(0.012)[0.283]	0.083(0.007)[0.212]
$\beta_{2,o}$	C2SLS	0.123(0.015)[0.310]	0.082(0.007)[0.209]	0.060(0.004)[0.154]	0.174(0.030)[0.445]	0.113(0.013)[0.285]	0.084(0.007)[0.210]
	OGMM	0.122(0.015)[0.309]	0.081(0.007)[0.207]	0.060(0.004)[0.153]	0.173(0.030)[0.441]	0.112(0.013)[0.283]	0.082(0.007)[0.207]
	BGMM	0.121(0.015)[0.305]	0.081(0.007)[0.206]	0.060(0.004)[0.152]	0.172(0.030)[0.440]	0.111(0.012)[0.281]	0.082(0.007)[0.205]
$\sigma_{u,o}$	C2SLS	0.580(0.383)[1.471]	0.503(0.291)[1.350]	0.422(0.197)[1.296]	1.183(1.448)[2.593]	1.132(1.319)[2.511]	1.176(1.418)[2.501]
		0.333(0.123){0.296}	0.289(0.085){0.225}	0.256(0.066){0.140}	0.883(1.090){0.394}	0.837(0.939){0.358}	0.933(1.101){0.352}
	OGMM	0.584(0.390)[1.469]	0.503(0.290)[1.351]	0.422(0.196)[1.294]	1.107(1.295)[2.542]	1.058(1.166)[2.491]	1.046(1.143)[2.463]
		0.344(0.130){0.298}	0.290(0.086){0.224}	0.256(0.065){0.140}	0.762(0.828){0.395}	0.704(0.699){0.358}	0.697(0.674){0.352}
	BGMM	0.580(0.385)[1.465]	0.502(0.290)[1.350]	0.421(0.196)[1.293]	1.106(1.290)[2.545]	1.055(1.161)[2.492]	1.047(1.143)[2.465]
		0.334(0.124){0.298}	0.289(0.085){0.223}	0.256(0.065){0.140}	0.758(0.825){0.394}	0.701(0.693){0.357}	0.699(0.676){0.351}
$\sigma_{v,o}$	C2SLS	0.176(0.032)[0.441]	0.132(0.017)[0.342]	0.108(0.012)[0.288]	0.499(0.287)[1.192]	0.464(0.239)[1.043]	0.447(0.219)[0.979]
	OGMM	0.176(0.032)[0.443]	0.132(0.017)[0.342]	0.108(0.012)[0.287]	0.489(0.277)[1.152]	0.455(0.231)[1.022]	0.440(0.212)[0.960]
	BGMM	0.176(0.032)[0.440]	0.132(0.017)[0.342]	0.108(0.012)[0.287]	0.488(0.278)[1.152]	0.455(0.232)[1.019]	0.440(0.212)[0.961]

Note: (i) $(\sigma_{u,o}, \sigma_{v,o}) = (1, 1)$ in Case I, and $(\sigma_{u,o}, \sigma_{v,o}) = (\sqrt{2}, \sqrt{2})$ in Case II".

(ii) In each cell, the black font represents "SD(MSE)[IDR]", and the blue font represents "SD(MSE){Zero − Rate}".

(continued on the next page)

Table 1 (continued from the previous page)

$\lambda_o = -0.5$		Case I: $v_{in} \sim \mathcal{N}(0,1)$			Case II: $v_{in} \sim \mathcal{T}(4)$		
$\sigma_{u,o} > \sigma_{v,o}$		$n = 100$	$n = 225$	$n = 400$	$n = 100$	$n = 225$	$n = 400$
λ_o	C2SLS	0.216(0.047)[0.547]	0.147(0.022)[0.378]	0.110(0.012)[0.280]	0.264(0.071)[0.669]	0.181(0.033)[0.459]	0.138(0.019)[0.353]
	OGMM	0.160(0.026)[0.401]	0.104(0.011)[0.266]	0.077(0.006)[0.199]	0.168(0.028)[0.428]	0.112(0.013)[0.288]	0.085(0.007)[0.218]
	BGMM	0.151(0.023)[0.384]	0.103(0.011)[0.262]	0.076(0.006)[0.197]	0.161(0.027)[0.410]	0.109(0.012)[0.279]	0.084(0.007)[0.214]
$\beta_{0,o}$	C2SLS	0.533(0.324)[1.486]	0.355(0.135)[0.773]	0.242(0.061)[0.556]	0.959(0.942)[2.188]	0.893(0.826)[2.093]	0.894(0.823)[2.006]
	OGMM	0.525(0.320)[1.483]	0.352(0.134)[0.756]	0.238(0.059)[0.544]	0.888(0.837)[2.167]	0.853(0.770)[2.082]	0.824(0.715)[1.998]
	BGMM	0.520(0.318)[1.478]	0.351(0.134)[0.752]	0.238(0.059)[0.542]	0.878(0.821)[2.140]	0.849(0.765)[2.071]	0.822(0.712)[1.997]
$\beta_{1,o}$	C2SLS	0.142(0.020)[0.358]	0.090(0.008)[0.233]	0.068(0.005)[0.175]	0.175(0.031)[0.443]	0.113(0.013)[0.286]	0.085(0.007)[0.214]
	OGMM	0.141(0.020)[0.359]	0.090(0.008)[0.232]	0.068(0.005)[0.173]	0.173(0.030)[0.439]	0.112(0.013)[0.285]	0.085(0.007)[0.213]
	BGMM	0.140(0.020)[0.356]	0.090(0.008)[0.229]	0.068(0.005)[0.174]	0.173(0.030)[0.439]	0.112(0.013)[0.283]	0.085(0.007)[0.213]
$\beta_{2,o}$	C2SLS	0.141(0.020)[0.360]	0.092(0.009)[0.236]	0.070(0.005)[0.178]	0.176(0.031)[0.439]	0.115(0.013)[0.291]	0.086(0.007)[0.217]
	OGMM	0.139(0.019)[0.360]	0.092(0.009)[0.235]	0.069(0.005)[0.177]	0.174(0.030)[0.436]	0.114(0.013)[0.288]	0.084(0.007)[0.214]
	BGMM	0.138(0.019)[0.355]	0.092(0.008)[0.236]	0.069(0.005)[0.176]	0.173(0.030)[0.435]	0.113(0.013)[0.287]	0.084(0.007)[0.213]
$\sigma_{u,o}$	C2SLS	0.627(0.459)[1.921]	0.426(0.196)[0.907]	0.283(0.084)[0.638]	1.237(1.572)[2.680]	1.152(1.373)[2.630]	1.144(1.348)[2.496]
		0.400(0.162){0.142}	0.318(0.105){0.043}	0.252(0.066){0.008}	0.951(1.187){0.362}	0.817(0.898){0.351}	0.863(0.919){0.322}
	OGMM	0.629(0.465)[1.910]	0.426(0.197)[0.904]	0.283(0.083)[0.635]	1.127(1.335)[2.633]	1.085(1.237)[2.596]	1.046(1.146)[2.472]
		0.406(0.168){0.142}	0.313(0.102){0.044}	0.252(0.066){0.008}	0.776(0.811){0.361}	0.701(0.684){0.351}	0.682(0.613){0.324}
	BGMM	0.623(0.455)[1.909]	0.425(0.196)[0.907]	0.282(0.083)[0.636]	1.121(1.321)[2.623]	1.084(1.233)[2.594]	1.045(1.144)[2.471]
		0.400(0.164){0.140}	0.313(0.101){0.044}	0.252(0.066){0.008}	0.762(0.791){0.360}	0.701(0.682){0.350}	0.681(0.611){0.323}
$\sigma_{v,o}$	C2SLS	0.234(0.055)[0.572]	0.160(0.026)[0.400]	0.121(0.015)[0.303]	0.518(0.309)[1.298]	0.484(0.258)[1.115]	0.447(0.217)[0.996]
	OGMM	0.234(0.055)[0.577]	0.160(0.025)[0.399]	0.120(0.014)[0.301]	0.504(0.295)[1.226]	0.476(0.250)[1.083]	0.439(0.210)[0.977]
	BGMM	0.232(0.054)[0.572]	0.159(0.025)[0.399]	0.120(0.014)[0.302]	0.504(0.295)[1.222]	0.475(0.250)[1.087]	0.439(0.210)[0.976]

Note: (i) $(\sigma_{u,o}, \sigma_{v,o}) = (1.5, 1)$ in Case I, and $(\sigma_{u,o}, \sigma_{v,o}) = (2, \sqrt{2})$ in Case II".

(ii) In each cell, the black font represents "SD(MSE)[IDR]", and the blue font represents "SD(MSE){Zero − Rate}".

Table 2. Experimental results for estimation in Cases III-IV

$\lambda_o = 0.5$		Case III: $v_{in} \sim \mathcal{U}(-3,3)$			Case IV: v_{in} is mixed		
$\sigma_{u,o} < \sigma_{v,o}$		$n = 100$	$n = 225$	$n = 400$	$n = 100$	$n = 225$	$n = 400$
λ_o	C2SLS	0.162(0.026)[0.408]	0.107(0.011)[0.270]	0.082(0.007)[0.207]	0.196(0.038)[0.485]	0.130(0.017)[0.330]	0.097(0.009)[0.245]
	OGMM	0.103(0.011)[0.253]	0.064(0.004)[0.164]	0.048(0.002)[0.123]	0.107(0.011)[0.266]	0.068(0.005)[0.174]	0.049(0.002)[0.127]
	BGMM	0.096(0.010)[0.243]	0.067(0.005)[0.162]	0.048(0.002)[0.121]	0.101(0.011)[0.255]	0.066(0.004)[0.170]	0.049(0.002)[0.126]
$\beta_{0,o}$	C2SLS	0.746(0.579)[1.872]	0.630(0.426)[1.584]	0.554(0.331)[1.415]	0.964(0.976)[2.389]	0.858(0.777)[2.103]	0.792(0.673)[1.965]
	OGMM	0.715(0.529)[1.783]	0.606(0.392)[1.525]	0.537(0.309)[1.376]	0.930(0.923)[2.342]	0.829(0.735)[2.070]	0.769(0.643)[1.930]
	BGMM	0.713(0.526)[1.791]	0.606(0.392)[1.527]	0.537(0.310)[1.376]	0.928(0.922)[2.339]	0.829(0.736)[2.069]	0.769(0.645)[1.932]
$\beta_{1,o}$	C2SLS	0.192(0.037)[0.492]	0.125(0.016)[0.320]	0.094(0.009)[0.243]	0.227(0.052)[0.584]	0.151(0.023)[0.391]	0.110(0.012)[0.285]
	OGMM	0.202(0.041)[0.496]	0.125(0.016)[0.322]	0.094(0.009)[0.241]	0.227(0.052)[0.587]	0.150(0.022)[0.385]	0.109(0.012)[0.280]
	BGMM	0.191(0.037)[0.491]	0.125(0.016)[0.319]	0.093(0.009)[0.241]	0.225(0.051)[0.581]	0.149(0.022)[0.384]	0.109(0.012)[0.281]
$\beta_{2,o}$	C2SLS	0.195(0.038)[0.499]	0.127(0.016)[0.327]	0.094(0.009)[0.240]	0.231(0.054)[0.580]	0.150(0.022)[0.387]	0.114(0.013)[0.293]
	OGMM	0.196(0.038)[0.496]	0.125(0.016)[0.323]	0.093(0.009)[0.241]	0.227(0.052)[0.573]	0.148(0.022)[0.379]	0.112(0.013)[0.286]
	BGMM	0.192(0.037)[0.492]	0.125(0.016)[0.322]	0.093(0.009)[0.239]	0.226(0.051)[0.572]	0.147(0.022)[0.378]	0.112(0.013)[0.285]
$\sigma_{u,o}$	C2SLS	0.787(0.654)[1.849]	0.693(0.525)[1.670]	0.623(0.426)[1.542]	1.148(1.391)[2.667]	1.056(1.179)[2.482]	0.979(1.029)[2.349]
		0.441(0.388){0.434}	0.379(0.226){0.386}	0.341(0.152){0.322}	0.712(0.813){0.402}	0.672(0.615){0.345}	0.605(0.461){0.318}
	OGMM	0.790(0.665)[1.839]	0.694(0.527)[1.664]	0.622(0.424)[1.540]	1.135(1.370)[2.648]	1.029(1.128)[2.466]	0.959(0.995)[2.332]
		0.473(0.402){0.440}	0.384(0.228){0.388}	0.341(0.152){0.321}	0.711(0.780){0.400}	0.622(0.536){0.345}	0.569(0.410){0.317}
	BGMM	0.785(0.656)[1.840]	0.693(0.525)[1.663]	0.622(0.424)[1.541]	1.131(1.355)[2.651]	1.028(1.127)[2.465]	0.959(0.995)[2.332]
		0.450(0.387){0.439}	0.381(0.226){0.386}	0.340(0.151){0.321}	0.688(0.760){0.399}	0.617(0.533){0.345}	0.567(0.409){0.318}
$\sigma_{v,o}$	C2SLS	0.188(0.040)[0.478]	0.136(0.019)[0.347]	0.111(0.012)[0.286]	0.371(0.152)[0.840]	0.304(0.097)[0.670]	0.261(0.070)[0.580]
	OGMM	0.190(0.042)[0.470]	0.134(0.019)[0.343]	0.109(0.012)[0.282]	0.362(0.149)[0.837]	0.290(0.090)[0.665]	0.249(0.064)[0.566]
	BGMM	0.183(0.039)[0.471]	0.135(0.019)[0.344]	0.109(0.012)[0.283]	0.361(0.147)[0.833]	0.291(0.090)[0.663]	0.249(0.064)[0.565]

Note: (i) $(\sigma_{u,o}, \sigma_{v,o}) = \left(1, \sqrt{3}\right)$ in Case I, and $(\sigma_{u,o}, \sigma_{v,o}) = (1.5, 2)$ in Case II".

(ii) In each cell, the black font represents "SD(MSE)[IDR]", and the blue font represents "SD(MSE){Zero − Rate}".

(continued on the next page)

Table 2 (continued from the previous page)

$\lambda_o = 0.5$	$\sigma_{u,o} = \sigma_{v,o}$	Case III: $v_{in} \sim \mathcal{U}(-3,3)$			Case IV: v_{in} is mixed		
		$n = 100$	$n = 225$	$n = 400$	$n = 100$	$n = 225$	$n = 400$
λ_o	C2SLS	0.180(0.033)[0.445]	0.120(0.015)[0.304]	0.091(0.008)[0.229]	0.208(0.043)[0.507]	0.137(0.019)[0.338]	0.102(0.010)[0.258]
	OGMM	0.102(0.011)[0.263]	0.066(0.004)[0.169]	0.048(0.002)[0.123]	0.106(0.011)[0.267]	0.067(0.005)[0.172]	0.050(0.003)[0.128]
	BGMM	0.098(0.010)[0.250]	0.064(0.004)[0.166]	0.048(0.002)[0.121]	0.101(0.010)[0.254]	0.065(0.004)[0.167]	0.049(0.002)[0.126]
$\beta_{0,o}$	C2SLS	0.877(0.873)[2.270]	0.672(0.505)[1.829]	0.516(0.284)[1.222]	1.100(1.323)[2.782]	0.924(0.943)[2.364]	0.790(0.681)[2.147]
	OGMM	0.816(0.772)[2.140]	0.643(0.466)[1.813]	0.494(0.262)[1.122]	0.999(1.143)[2.571]	0.869(0.858)[2.277]	0.768(0.650)[2.116]
	BGMM	0.815(0.778)[2.136]	0.640(0.464)[1.810]	0.493(0.263)[1.125]	0.996(1.150)[2.559]	0.868(0.861)[2.281]	0.767(0.650)[2.117]
$\beta_{1,o}$	C2SLS	0.206(0.043)[0.534]	0.136(0.019)[0.349]	0.104(0.011)[0.270]	0.244(0.060)[0.614]	0.161(0.026)[0.417]	0.117(0.014)[0.298]
	OGMM	0.205(0.042)[0.531]	0.135(0.018)[0.345]	0.104(0.011)[0.269]	0.243(0.059)[0.620]	0.159(0.025)[0.412]	0.116(0.013)[0.296]
	BGMM	0.203(0.041)[0.527]	0.135(0.018)[0.343]	0.104(0.011)[0.268]	0.241(0.058)[0.612]	0.159(0.025)[0.414]	0.116(0.013)[0.295]
$\beta_{2,o}$	C2SLS	0.214(0.046)[0.544]	0.138(0.019)[0.353]	0.103(0.011)[0.263]	0.246(0.061)[0.631]	0.163(0.027)[0.426]	0.121(0.015)[0.311]
	OGMM	0.212(0.045)[0.534]	0.136(0.019)[0.351]	0.102(0.010)[0.261]	0.242(0.059)[0.624]	0.161(0.026)[0.418]	0.119(0.014)[0.305]
	BGMM	0.209(0.044)[0.534]	0.136(0.018)[0.352]	0.102(0.010)[0.260]	0.241(0.058)[0.620]	0.160(0.026)[0.415]	0.119(0.014)[0.305]
$\sigma_{u,o}$	C2SLS	0.918(1.008)[2.387]	0.737(0.623)[2.206]	0.565(0.347)[1.250]	1.234(1.724)[2.994]	1.096(1.338)[2.790]	0.958(1.005)[2.670]
		0.531(0.287){0.266}	0.452(0.205){0.150}	0.386(0.152){0.068}	0.757(0.652){0.320}	0.671(0.475){0.245}	0.581(0.344){0.181}
	OGMM	0.921(1.011)[2.382]	0.736(0.621)[2.200]	0.561(0.342)[1.212]	1.208(1.667)[2.987]	1.068(1.286)[2.778]	0.955(1.001)[2.668]
		0.541(0.298){0.264}	0.452(0.205){0.149}	0.381(0.148){0.068}	0.733(0.600){0.313}	0.626(0.412){0.243}	0.570(0.331){0.182}
	BGMM	0.917(0.999)[2.380]	0.734(0.615)[2.203]	0.560(0.340)[1.209]	1.204(1.650)[2.997]	1.067(1.281)[2.776]	0.955(0.999)[2.670]
		0.535(0.292){0.261}	0.448(0.201){0.148}	0.383(0.149){0.067}	0.716(0.580){0.312}	0.625(0.410){0.242}	0.568(0.330){0.182}
$\sigma_{v,o}$	C2SLS	0.243(0.059)[0.616]	0.180(0.032)[0.465]	0.141(0.020)[0.362]	0.411(0.174)[0.964]	0.328(0.108)[0.744]	0.283(0.080)[0.653]
	OGMM	0.236(0.056)[0.602]	0.176(0.031)[0.457]	0.138(0.019)[0.356]	0.398(0.165)[0.942]	0.312(0.098)[0.731]	0.277(0.077)[0.646]
	BGMM	0.236(0.056)[0.602]	0.176(0.031)[0.457]	0.138(0.019)[0.356]	0.398(0.165)[0.944]	0.312(0.098)[0.732]	0.277(0.077)[0.645]

Note: (i) $(\sigma_{u,o}, \sigma_{v,o}) = \left(\sqrt{3}, \sqrt{3}\right)$ in Case I, and $(\sigma_{u,o}, \sigma_{v,o}) = (2, 2)$ in Case II".

(ii) In each cell, the black font represents "SD(MSE)[IDR]", and the blue font represents "SD(MSE){Zero − Rate}".

(continued on the next page)

Table 2 (continued from the previous page)

$\lambda_o = 0.5$	$\sigma_{u,o} > \sigma_{v,o}$	Case III: $v_{in} \sim \mathcal{U}(-3,3)$			Case IV: v_{in} is mixed		
		$n = 100$	$n = 225$	$n = 400$	$n = 100$	$n = 225$	$n = 400$
λ_o	C2SLS	0.190(0.036)[0.468]	0.125(0.016)[0.321]	0.094(0.009)[0.235]	0.227(0.052)[0.544]	0.147(0.022)[0.363]	0.111(0.012)[0.280]
	OGMM	0.104(0.011)[0.259]	0.066(0.004)[0.171]	0.049(0.002)[0.125]	0.109(0.012)[0.271]	0.069(0.005)[0.176]	0.051(0.003)[0.130]
	BGMM	0.099(0.010)[0.253]	0.065(0.004)[0.167]	0.049(0.002)[0.123]	0.103(0.011)[0.258]	0.067(0.005)[0.173]	0.050(0.003)[0.128]
$\beta_{0,o}$	C2SLS	0.920(0.961)[2.421]	0.652(0.455)[1.680]	0.467(0.228)[1.067]	1.240(1.700)[3.131]	0.917(0.910)[2.452]	0.740(0.577)[1.785]
	OGMM	0.844(0.835)[2.260]	0.609(0.403)[1.503]	0.441(0.205)[0.982]	1.073(1.363)[2.848]	0.842(0.792)[2.393]	0.682(0.498)[1.538]
	BGMM	0.839(0.836)[2.244]	0.607(0.403)[1.492]	0.440(0.206)[0.985]	1.064(1.365)[2.821]	0.839(0.794)[2.389]	0.681(0.498)[1.534]
$\beta_{1,o}$	C2SLS	0.216(0.047)[0.556]	0.144(0.021)[0.360]	0.106(0.011)[0.272]	0.259(0.067)[0.652]	0.171(0.029)[0.439]	0.128(0.017)[0.335]
	OGMM	0.216(0.047)[0.555]	0.143(0.020)[0.361]	0.106(0.011)[0.272]	0.257(0.066)[0.645]	0.169(0.029)[0.433]	0.127(0.016)[0.331]
	BGMM	0.214(0.046)[0.555]	0.142(0.020)[0.358]	0.105(0.011)[0.271]	0.255(0.065)[0.644]	0.169(0.028)[0.431]	0.127(0.016)[0.331]
$\beta_{2,o}$	C2SLS	0.220(0.049)[0.560]	0.145(0.021)[0.372]	0.108(0.012)[0.278]	0.264(0.070)[0.664]	0.172(0.030)[0.442]	0.131(0.017)[0.338]
	OGMM	0.216(0.047)[0.547]	0.142(0.020)[0.366]	0.107(0.011)[0.273]	0.258(0.067)[0.654]	0.169(0.029)[0.437]	0.136(0.018)[0.332]
	BGMM	0.215(0.046)[0.544]	0.142(0.020)[0.364]	0.107(0.011)[0.273]	0.257(0.066)[0.655]	0.168(0.028)[0.433]	0.128(0.016)[0.333]
$\sigma_{u,o}$	C2SLS	0.950(1.084)[2.629]	0.696(0.534)[1.648]	0.496(0.263)[1.077]	1.283(1.938)[3.384]	1.030(1.176)[3.182]	0.856(0.775)[1.842]
		0.572(0.327){0.206}	0.475(0.230){0.081}	0.400(0.166){0.025}	0.777(0.608){0.236}	0.656(0.431){0.128}	0.594(0.354){0.071}
	OGMM	0.949(1.085)[2.620]	0.691(0.525)[1.639]	0.495(0.261)[1.071]	1.260(1.888)[3.362]	1.017(1.150)[3.172]	0.836(0.744)[1.834]
		0.584(0.341){0.203}	0.474(0.229){0.079}	0.395(0.162){0.026}	0.764(0.585){0.229}	0.644(0.415){0.126}	0.564(0.320){0.071}
	BGMM	0.943(1.066)[2.618]	0.690(0.522)[1.626]	0.494(0.260)[1.072]	1.256(1.867)[3.369]	1.015(1.145)[3.175]	0.835(0.740)[1.822]
		0.568(0.322){0.202}	0.472(0.227){0.078}	0.394(0.161){0.026}	0.749(0.563){0.229}	0.638(0.408){0.126}	0.565(0.320){0.071}
$\sigma_{v,o}$	C2SLS	0.269(0.073)[0.678]	0.193(0.037)[0.507]	0.148(0.022)[0.379]	0.448(0.201)[1.077]	0.351(0.123)[0.839]	0.292(0.085)[0.691]
	OGMM	0.261(0.068)[0.663]	0.189(0.036)[0.500]	0.146(0.021)[0.373]	0.428(0.184)[1.045]	0.339(0.115)[0.830]	0.281(0.079)[0.684]
	BGMM	0.260(0.067)[0.664]	0.189(0.036)[0.501]	0.146(0.021)[0.372]	0.428(0.185)[1.047]	0.339(0.115)[0.829]	0.279(0.078)[0.683]

Note: (i) $(\sigma_{u,o}, \sigma_{v,o}) = \left(2, \sqrt{3}\right)$ in Case I, and $(\sigma_{u,o}, \sigma_{v,o}) = (2.5, 2)$ in Case II".

(ii) In each cell, the black font represents "SD(MSE)[IDR]", and the blue font represents "SD(MSE){Zero − Rate}".

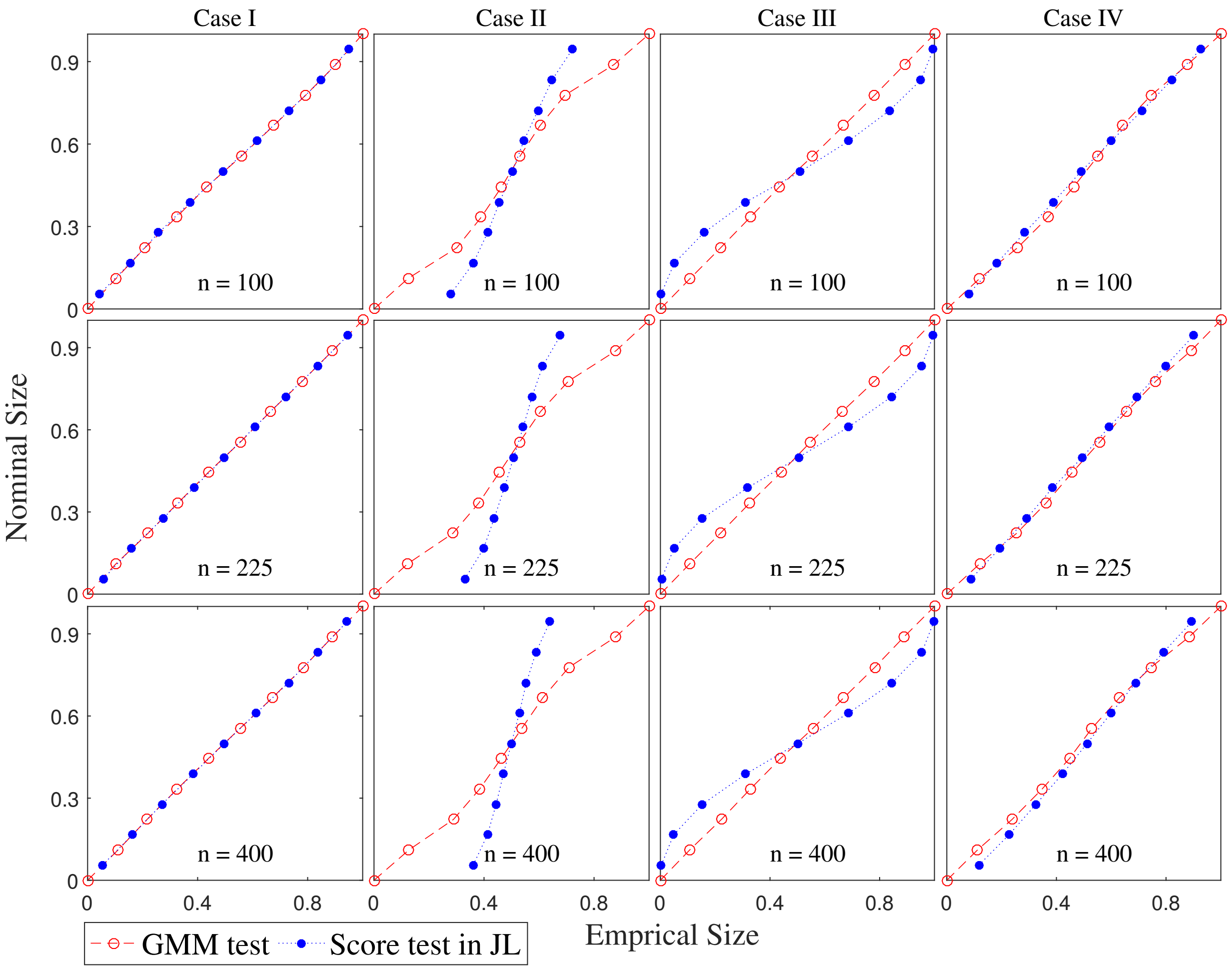

Figure 1: Plots of Nominal and Empirical sizes for DGP-I.

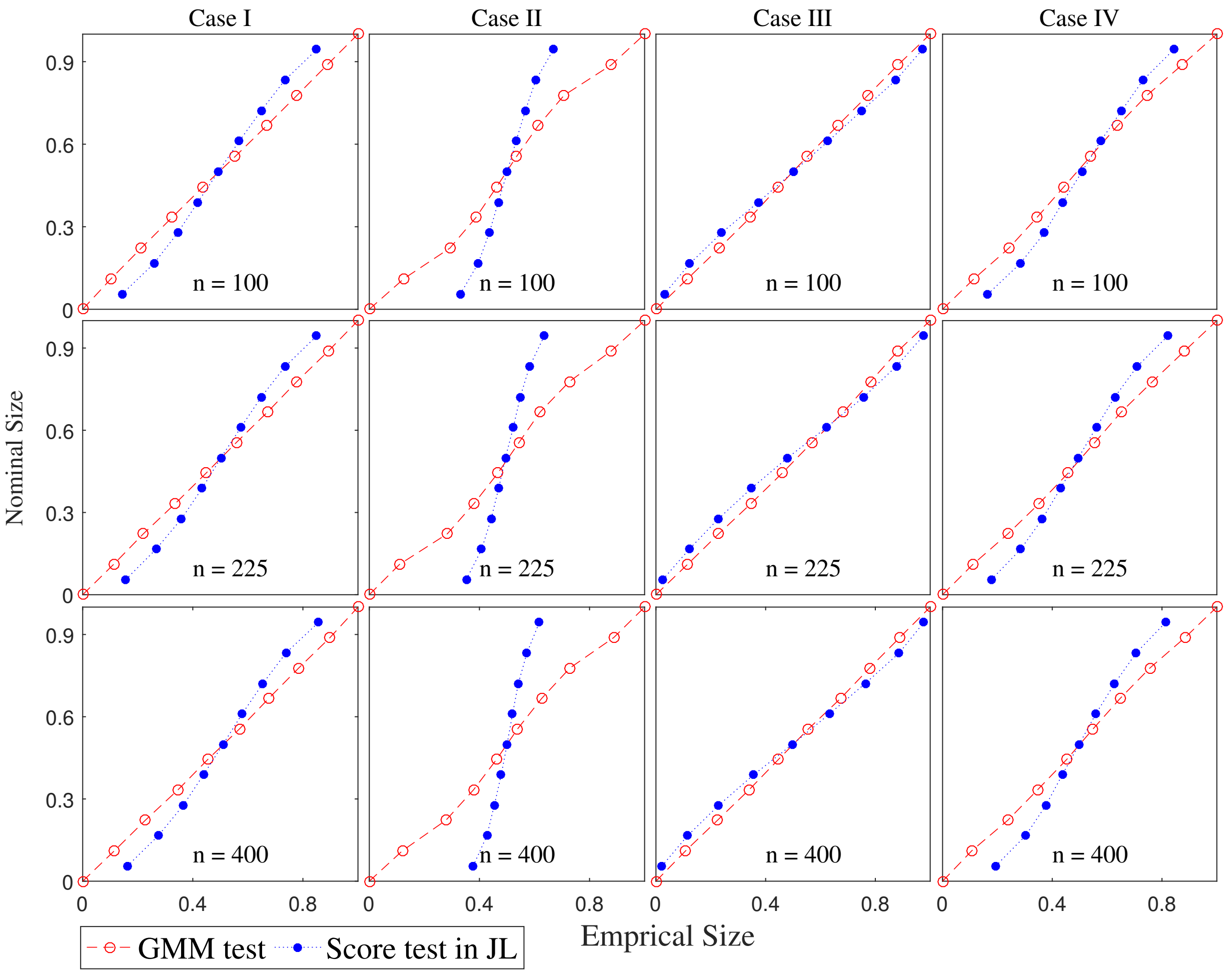

Figure 2: Plots of Nominal and Empirical sizes for DGP-II.

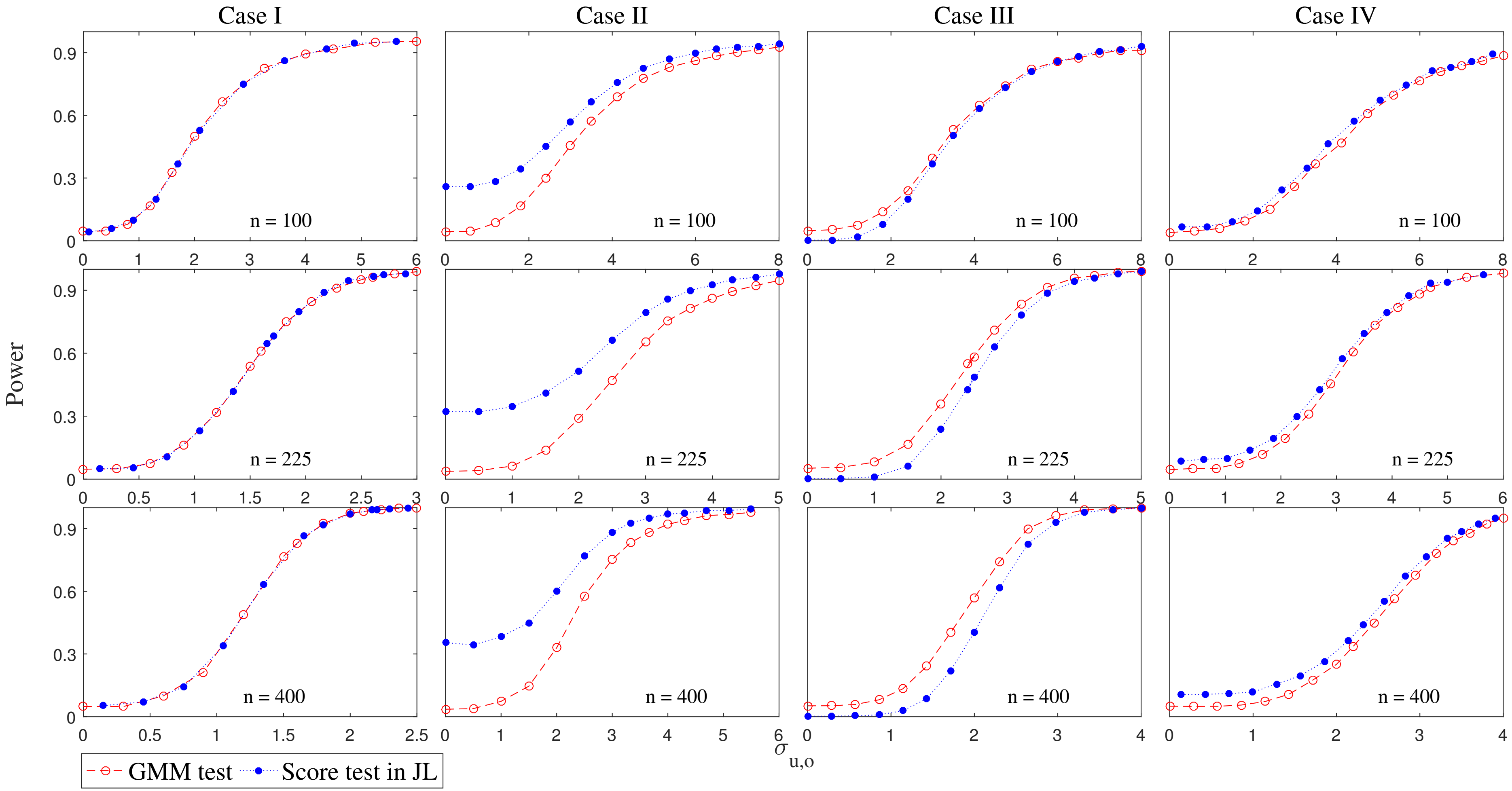

Figure 3: Empirical powers for DGP-I.

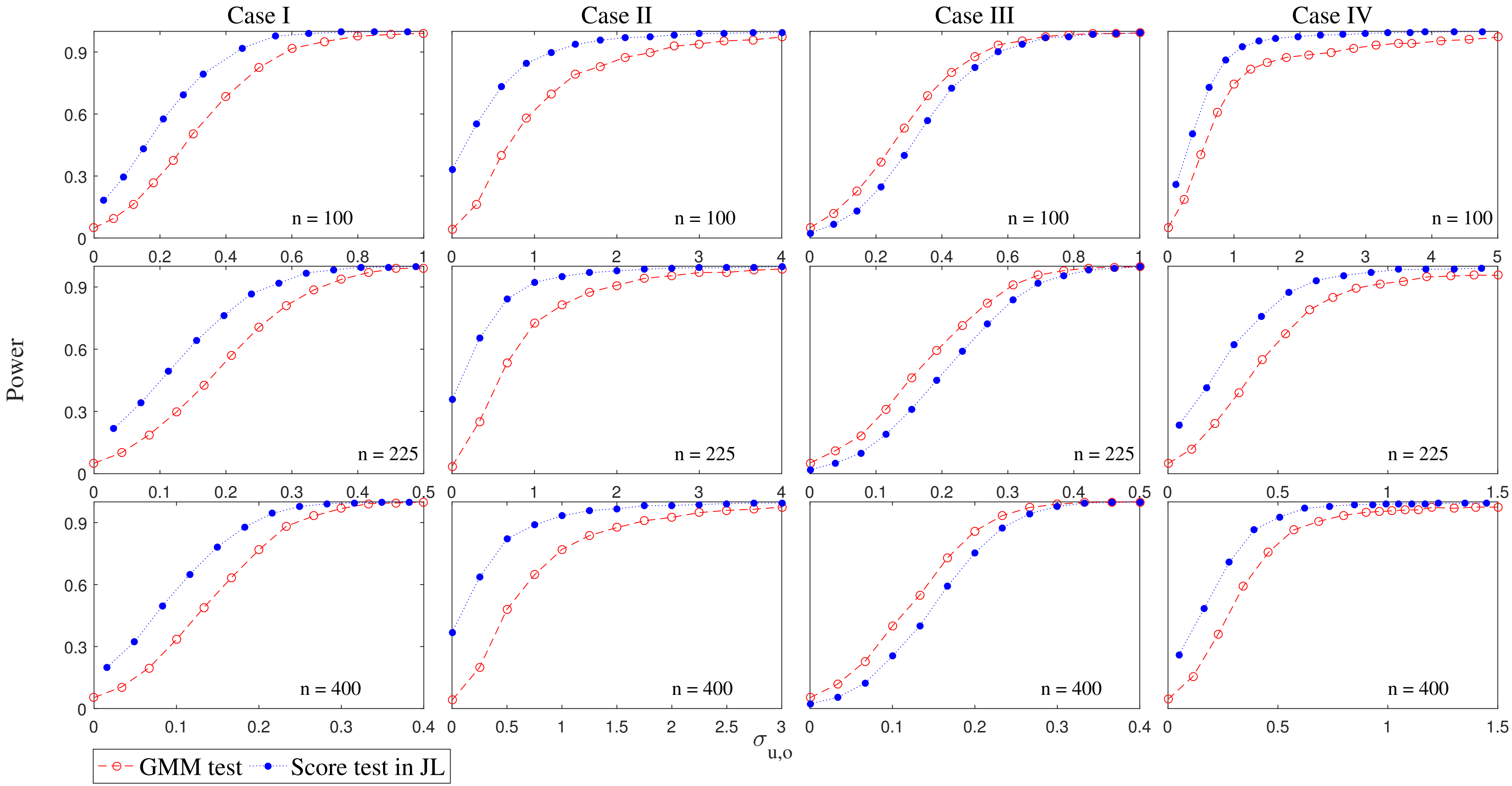

Figure 4: Empirical powers for DGP-II.

References

Adetutu, M., Glass, A., Kenjegalieva, K., Sickles, R., 2015. The effects of efficiency and TFP growth on pollution in Europe: a multistage spatial analysis. Journal of Productivity Analysis 43, 307–326.

Aigner, D., Lovell, C.A.K., Schmidt, P., 1977. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 6, 21–37.

Arbia, G., Bera, A.K., Doğan, O., Taşpınar, S., 2020. Testing impact measures in spatial autoregressive models. International Regional Science Review 43, 40–75.

Areal, F.J., Balcombe, K., Tiffin, R., 2012. Integrating spatial dependence into stochastic frontier analysis. Australian Journal of Agricultural and Resource Economics 56, 521–541.

Baltagi, B.H., 2011. Spatial Panels, in: Ullah, A., Giles, D.E.A. (Eds.), Handbook of empirical economics and finance. Chapman and Hall, Taylor & Francis Group, Boca Raton, Florida, pp. 435–454.

Baltagi, B.H., 2013. Econometric Analysis of Panel Data, the 5th edn. John Wiley & Sons, Inc., Chichester, Unites Kingdom.

Baptista, R., 2000. Do innovations diffuse faster within geographical clusters? International Journal of Industrial Organization 18, 515–535.

Carvalho, A., 2018. Efficiency spillovers in Bayesian stochastic frontier models: application to electricity distribution in New Zealand. Spatial Economic Analysis 13, 171–190.

Cliff, A., Ord, J., 1973. Spatial autocorrelation. Pion, London.

Cliff, A., Ord, J., 1981. Spatial process: models and applications. Pion, London.

Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press, Oxford.

Deng, M.Y., Wang, M., 2022. Artificial regression test diagnostics for impact measures in spatial models. Economics Letters 217.

Druska, V., Horrace, W.C., 2004. Generalized moments estimation for spatial panel data: Indonesian rice farming. American Journal of Agricultural Economics 86, 185–198.

Engle, R.F., McFadden, D.L. (Eds.), 1994. Handbook of econometrics. Vol. IV. North-Holland Publishing Co., Amsterdam.

Fried, H.O., Lovell, C.K., Schmidt, S.S., 2008. The Measurement of Productive Efficiency and Productivity Growth. Oxford University Press, New York.

Fusco, E., Vidoli, F., 2013. Spatial stochastic frontier models: controlling spatial global and local heterogeneity. International Review of Applied Economics 190, 679–694.

Galli, F., 2023. A spatial stochastic frontier model including both frontier and error-based spatial cross-sectional dependence. Spatial Economic Analysis 18, 239–258.

Glass, A., Kenjegalieva, K., Paez-Farrell, J., 2013. Productivity growth decomposition using a spatial autoregressive frontier model. Economics Letters 119, 291–295.

Glass, A., Kenjegalieva, K., Sickles, R.C., 2014. Estimating efficiency spillovers with state level evidence formanufacturing in the US. Economics Letters 123, 154–159.

Glass, A.J., Kenjegalieva, K., Sickles, R.C., 2016. A spatial autoregressive stochastic frontier model for panel data with asymmetric efficiency spillovers. Journal of Econometrics 190, 289–300.

Greene, W.H., 1990. A Gamma-distributed stochastic frontier model. Journal of Econometrics 46, 141–163.

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

Jenisha, N., Prucha, I.R., 2009. Central limit theorems and uniform laws of large numbers for arrays of random fields. Journal of Econometrics 150, 86–98.

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

Jin, F., Lee, L.f., 2020. Asymptotic properties of a spatial autoregressive stochastic frontier model. Journal of Spatial Econometrics 1.

Jin, F., Wang, Y., 2021. Gmm estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. Econometric Reviews 41, 652–674.

Jin, F., Wang, Y., 2023. Consistent non-Gaussian pseudo maximum likelihood estimations of spatial autoregressive models. Econometric Theory 0, 1–39.

Kelejian, H., Piras, G., 2017. Spatial econometrics. Academic Press.

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

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

Kopp, R.J., Mullahy, J., 1990. Moment-based estimation and testing of stochastic frontier models. Journal of Econometrics 46, 165–183.

Kutlu, L., 2012. U.S. banking efficiency, 1984–1995. Economics Letters 117, 53–56.

Kutlu, L., 2017. A constrained state space approach for estimating firm efficiency. Economics Letters 152, 54–56.

Kutlu, L., 2018. Estimating efficiency in a spatial autoregressive stochastic frontier model. Economics Letters 163, 155–157.

Kutlu, L., 2022. Spatial stochastic frontier model with endogenous weighting matrix. Empirical Economics 63, 1947–1968.

Kutlu, L., 2023. Calculating efficiency for spatial autoregressive stochastic frontier model. Economics Letters 225.

Kutlu, L., Tran, K.C., Tsionas, M.G., 2020. A spatial stochastic frontier model with endogenous frontier and environmental variables. European Journal of Operational Research 286, 389–399.

Lahiri, S., 1996. On inconsistency of estimators based on spatial data under infill asymptotics. Sankhyā 58, 403–417.

Lai, H.p., Tran, K.C., 2022. Persistent and transient inefficiency in a spatial autoregressive panel stochastic frontier model. Journal of Productivity Analysis 58, 1–13.

Lee, L.f., 2001. Generalized method of moments estimation of spatial autoregressive processes. Manuscript, Department of Economics, OSU, Columbus, OH.

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

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

Lee, L.f., 2007b. The method of elimination and substitution in the GMM estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics 140, 155–189.

Lee, L.f., Yang, C., Yu, J., 2022. QML and efficient GMM estimation of spatial autoregressive models with dominant (popular) units. Journal of Business & Economic Statistics 41, 550–562.

LeSage, J.P., Pace, R.K., 2009. Introduction to spatial econometrics. CRC Press, Taylor & Francis Group, Boca Raton.

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

Liu, T., Lee, L.f., 2019. A likelihood ratio test for spatial model selection. Journal of Econometrics 213, 434–458.

Liu, X., Lee, L.f., Bollinger, C.R., 2010. An efficient GMM estimator of spatial autoregressive models. Journal of Econometrics 159, 303–309.

Meeusen, W., van den Broeck, J., 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review 18, 435–444.

Olson, A.J., Schmidt, P., Waldman, D.M., 1980. A Monte Carlo study of estimators of stochastic frontier production functions. Journal of Econometrics 13, 67–82.

Pavlyuk, D., 2011. Application of the spatial stochastic frontier model for analysis of a regional tourism sector. Transp Telecommun 12, 28–38.

Schmidt, A.M., B., A.R., Helfand, S.M., Fonseca, T.C.O., 2009. Spatial stochastic frontier models: accounting for unobserved local determinants of inefficiency. Journal of Productivity Analysis 31, 101–112.

Schmidt, P., 1986. Frontier production functions. Econometric Reviews 4, 289–328.

Stevenson, R., 1980. Likelihood functions for generalized stochastic frontier estimation. Journal of Econometrics 13, 57–66.

Su, L., 2012. Semiparametric GMM estimation of spatial autoregressive models. Journal of Econometrics 167, 543–560.

Taşpınar, S., Doğan, O., Vijverberg, W.P.M., 2018. GMM inference in spatial autoregressive models. Econometric Reviews 37, 931–954.

Tran, K.C., Tsionas, M.G., Prokhorov, A.B., 2023. Semiparametric estimation of spatial autoregressive smooth-coefficientpanel stochastic frontier models. European Journal of Operational Research 304, 1189–1199.

Tsionas, E.G., Michaelides, P.G., 2016. A spatial stochastic frontier model with spillovers: evidence from Italian regions. Scottish Journal of Political Economy 63, 243–257.

Vidoli, F., Cardillob, C., Fuscoc, E., Canello, J., 2016. Spatial nonstationarity in the stochastic frontier model: an application to the italian wine industry. Regional Science and Urban Economics 61, 153–164.

Waldman, D.M., 1982. A stationary point for the stochastic frontier likelihood. Journal of Econometrics 18, 275–279.

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

Xu, X., Lee, L.f., 2015. Maximum likelihood estimation of a spatial autoregressive Tobit model. Journal of Econometrics 188, 164–280.

Xu, X., Lee, L.f., 2018. Sieve maximum likelihood estimation of the spatial autoregressive Tobit model. Journal of Econometrics 203, 96–112.