

Indirect Inference Estimation of Spatial Autoregressions

Yong Bao*
ybao@purdue.edu
Department of Economics
Purdue University

Xiaotian Liu
liu2004@purdue.edu
Department of Economics
Purdue University

Lihong Yang
lihong.yang@ruc.edu.cn
School of Economics, Nanjing Audit University, China
National Academy of Development and Strategy, Renmin University of China

Abstract

The ordinary least squares (OLS) estimator for spatial autoregressions may be consistent as pointed out by Lee (2002), provided that each spatial unit is influenced aggregately by a significant portion of the total units. This paper presents a unified asymptotic distribution result of the properly recentered OLS estimator and proposes a new estimator that is based on the indirect inference (II) procedure. The resulting estimator can always be used regardless of the degree of aggregate influence on each spatial unit from other units and is consistent and asymptotically normal. The new estimator does not rely on distributional assumptions and is robust to unknown heteroscedasticity. Its good finite-sample performance, in comparison with existing estimators that are also robust to heteroscedasticity, is demonstrated by a Monte-Carlo study.

Key Words: Spatial Autoregression; OLS; Indirect Inference

JEL Classifications: C21, C10, C13

*Corresponding author

Indirect Inference Estimation of Spatial Autoregressions

Abstract

The ordinary least squares (OLS) estimator for spatial autoregressions may be consistent as pointed out by Lee (2002), provided that each spatial unit is influenced aggregately by a significant portion of the total units. This paper presents a unified asymptotic distribution result of the properly recentered OLS estimator and proposes a new estimator that is based on the indirect inference (II) procedure. The resulting estimator can always be used regardless of the degree of aggregate influence on each spatial unit from other units and is consistent and asymptotically normal. The new estimator does not rely on distributional assumptions and is robust to unknown heteroscedasticity. Its good finite-sample performance, in comparison with existing estimators that are also robust to heteroscedasticity, is demonstrated by a Monte-Carlo study.

Key Words: Spatial Autoregression; OLS; Indirect Inference.

JEL Classifications: C21, C10, C13

1 Introduction

Spatial autoregressions (SAR) have attracted lots of attention from both practical and theoretical sides in economics and other disciplines of social sciences since the classical work of Cliff and Ord (1981). Its popularity is mainly due to its parsimonious representation of the cross-sectional correlation by a weight matrix. Correlation in spatial data arises naturally due to competition, copycatting, spillover, aggregation, to name just a few. In these contexts, space embodied in the weight matrix can be defined in terms of not only geographical distance but also economic distance.

The spatial autoregression model extends autocorrelation in time series to the spatial dimension in the sense that in the structural equation a spatially “lagged” dependent variable is included as a regressor. In time series, the autoregression model can be estimated consistently by the ordinary least squares (OLS), but for the spatial autoregression, the OLS is usually regarded as an inconsistent method. Robinson (2008) provided an excellent discussion of the intuition behind this. Estimation strategies like the maximum likelihood (ML), quasi maximum likelihood (QML), instrument variables (IV), and generalized method of moments (GMM) have

been proposed in the literature. The ML is the most efficient, but it imposes stringent distributional assumptions on the data generating process, whereas both the ML and QML rule out heteroscedasticity in the error term. Further, the (Q)ML involves calculating the determinant or eigenvalues of a matrix that is of the same size as the sample, and thus many researchers dismiss its use in moderately large samples and advocate the more flexible IV and GMM estimators that may incur less computational burden and are also robust to heteroscedasticity.

Lee (2002) overturned the traditional wisdom regarding the OLS estimator in spatial autoregressions when there are exogenous regressors included. He showed that while the OLS estimator is inconsistent for spatial autoregressions with a sparse weight matrix, it can be consistent when spatial units may have small spatial impacts on other units but each unit may be influenced aggregately by a significant portion of the total units. For the special case of the so-called pure SAR model, namely, when there is no other exogenous regressor, Lee (2002) demonstrated that regardless of the structure of the weight matrix, the OLS estimator is always inconsistent.

In practice, one may have limited knowledge to judge whether a spatial unit is influenced by a significant portion of the total units. When a researcher is constructing the weight matrix, she may know the number of neighboring units for each unit, but she may be unable to tell whether the number of neighbors is significant in finite samples. Thus, this poses a challenge for practitioners regarding the usefulness of the OLS estimator: it may or may not be consistent, depending on the degree of aggregate influence on each unit from other units, which may hardly manifest itself in finite samples. This paper carefully analyzes the asymptotic distribution theory for the OLS estimator. A unified asymptotic distribution result for the recentered OLS estimator is presented under different regimes for the spatial weight matrix. Given the asymptotic result for the recentered OLS estimator, a new estimator based on the indirect inference (II) procedure is proposed.

Kyriacou et al. (2017) novelly used the II procedure to correct the inconsistency of the OLS estimator in the pure SAR model under homoscedasticity. Even though they provided promising simulation results under some mild heteroscedasticity, they have yet to show rigorously how to construct a consistent estimator with no restrictions on the form of heteroscedasticity. In contrast, this paper considers the SAR model with exogenous regressors and it adds to the existing spatial literature with unknown heteroscedasticity.¹

¹Recent literature on dealing with heteroscedasticity in the spatial framework includes Kelejian and Prucha (2010), Badinger and Egger (2011), Liu and Yang (2015), Jin and Lee (2019), among others. An essential idea in this strand of literature is to use some moment conditions that are robust to unknown heteroscedasticity.

Note that the problem of inconsistency (of the OLS estimator) is solely due to the presence of the endogenous spatially lagged variable. Once the spatial autoregression parameter is estimated consistently, the OLS procedure can be used to estimate the remaining parameters, though the asymptotic variance needs to be modified accordingly to take into account the uncertainty in the estimated spatial autoregression parameter.

The plan of this paper is as follows. In the next section, the asymptotic behavior of the OLS estimator is discussed under different spatial scenarios. The II estimator that aims to correct the possible inconsistency of the OLS estimator is defined and its asymptotic distribution is derived. A very important message from this section is that regardless of the degree of aggregate influence on each spatial unit from other units, the II procedure can always be used and the resulting estimator is consistent and asymptotically normal. Section 3 discusses the special case of pure SAR. Section 4 provides Monte Carlo evidence of the effectiveness of the II estimation strategy. It shows that the II estimator possesses good finite-sample performance relative to other consistent estimators that are also robust to unknown heteroscedasticity and that the II method may be favored for the purpose of hypothesis testing, especially for testing the spatial autoregression parameter. Section 5 concludes. Technical details, including some useful lemmas and proofs of the results presented in Sections 2 and 3, are collected in the appendix.

Throughout, K is used to denote a positive constant on different occasions, arbitrarily large but bounded, that does not depend on the sample size n and whose value may vary in different contexts. \mathbf{I}_n is the identity matrix of dimension n and $\mathbf{1}_n$ is an $n \times 1$ vector of ones. For an $n \times 1$ vector \mathbf{a}_n , $a_{i,n}$ denotes its i -th element and for an $n \times n$ matrix \mathbf{A}_n , $a_{ij,n}$ denotes its ij -th element. $\|\mathbf{A}_n\|_\infty = \max_{1 \leq i < n} \sum_{j=1}^n |a_{ij,n}|$ and $\|\mathbf{A}_n\|_1 = \max_{1 \leq j < n} \sum_{i=1}^n |a_{ij,n}|$ are the maximum row sum norm and maximum column sum norm, respectively. A sequence of matrices $\{\mathbf{A}_n\}$ is uniformly bounded in row sum if $\|\mathbf{A}_n\|_\infty \leq K$, and is uniformly bounded in column sum if $\|\mathbf{A}_n\|_1 \leq K$. tr and \odot are matrix trace and Hadamard product operators, respectively. $\text{Dg}(\mathbf{a}_n)$ denotes a square diagonal matrix with the vector \mathbf{a}_n spanning the main diagonal, $\text{dg}(\mathbf{A}_n)$ is an $n \times 1$ column vector that collects in order the diagonal elements of the square matrix \mathbf{A}_n , and $\text{Dg}(\mathbf{A}_n) = \text{Dg}(\text{dg}(\mathbf{A}_n))$. The subscript 0 is used to signify the true parameter value.

2 Main Results

Consider the SAR model

$$\mathbf{y}_n = \lambda \mathbf{W}_n \mathbf{y}_n + \mathbf{X}_n \boldsymbol{\beta} + \mathbf{u}_n = \mathbf{Z}_n \boldsymbol{\theta} + \mathbf{u}_n, \quad (1)$$

where n is the total number of cross-sectional units, $\mathbf{Z}_n = (\mathbf{W}_n \mathbf{y}_n, \mathbf{X}_n)$, $\boldsymbol{\theta} = (\lambda, \boldsymbol{\beta}')'$, \mathbf{y}_n is an $n \times 1$ vector, collecting observations on the dependent variable, \mathbf{X}_n is an $n \times k$ matrix of observations on k exogenous nonstochastic regressors with coefficient vector $\boldsymbol{\beta}$, \mathbf{W}_n is an $n \times n$ matrix of spatial weights with zero diagonals, λ is the spatial autoregression coefficient, and \mathbf{u}_n is an n -dimensional vector of error terms.

For the ease of presentation, the following matrix notation is introduced in this paper:

$$\begin{aligned} \mathbf{S}_n(\lambda) &= \mathbf{I}_n - \lambda \mathbf{W}_n, & \mathbf{G}_n(\lambda) &= \mathbf{W}_n \mathbf{S}_n^{-1}(\lambda), & \mathbf{M}_n &= \mathbf{I}_n - \mathbf{X}_n (\mathbf{X}_n' \mathbf{X}_n)^{-1} \mathbf{X}_n', \\ \mathbf{D}_n(\lambda) &= \text{Dg}(\mathbf{M}_n \mathbf{G}_n(\lambda)), & \mathbf{E}_n(\lambda) &= \mathbf{M}_n \mathbf{G}_n(\lambda) - \mathbf{D}_n(\lambda). \end{aligned} \quad (2)$$

When a matrix is presented without its argument λ , it means that it is evaluated at the parameter value λ_0 . Namely, $\mathbf{S}_n = \mathbf{S}_n(\lambda_0)$, $\mathbf{G}_n = \mathbf{G}_n(\lambda_0)$, $\mathbf{D}_n = \mathbf{D}_n(\lambda_0)$, and $\mathbf{E}_n = \mathbf{E}_n(\lambda_0)$.

If λ is known (equal to its true value), the model becomes a standard linear model with $\mathbf{S}_n \mathbf{y}_n$ being the dependent variable; otherwise, $\mathbf{W}_n \mathbf{y}_n$ appears on the right-hand side of (1) as a spatially lagged or weighted variable. The OLS estimator of $\boldsymbol{\theta}_0$ is $\hat{\boldsymbol{\theta}}_n = (\mathbf{Z}_n' \mathbf{Z}_n)^{-1} \mathbf{Z}_n' \mathbf{y}_n$, which may be inconsistent. Since the inconsistency of $\hat{\boldsymbol{\theta}}_n$ is solely due to the endogenous $\mathbf{W}_n \mathbf{y}_n$, the properties of $\hat{\lambda}_n$ (the first element of $\hat{\boldsymbol{\theta}}_n$) are discussed first in this paper. Once a consistent estimator of λ_0 is available, one can show that a consistent estimator of $\boldsymbol{\beta}_0$ follows immediately (see Theorem 4 to be introduced).²

Let $r_n = r_{1n} + r_{2n}$, $r_{1n} = \mathbf{u}_n' \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n$, $r_{2n} = \boldsymbol{\beta}_0' \mathbf{X}_n' \mathbf{G}_n' \mathbf{M}_n \mathbf{u}_n$, $d_n = d_{1n} + d_{2n} + d_{3n}$, $d_{1n} = \mathbf{u}_n' \mathbf{G}_n' \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n$, $d_{2n} = 2\boldsymbol{\beta}_0' \mathbf{X}_n' \mathbf{G}_n' \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n$, and $d_{3n} = \boldsymbol{\beta}_0' \mathbf{X}_n' \mathbf{G}_n' \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \boldsymbol{\beta}_0$. By using the partitioned regression formula and substituting $\mathbf{W}_n \mathbf{y}_n = \mathbf{G}_n \mathbf{X}_n \boldsymbol{\beta}_0 + \mathbf{G}_n \mathbf{u}_n$, one may put

$$\hat{\lambda}_n - \lambda_0 = \frac{\mathbf{y}_n' \mathbf{W}_n' \mathbf{M}_n \mathbf{u}_n}{\mathbf{y}_n' \mathbf{W}_n' \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n} = \frac{r_n}{d_n} = \frac{r_{1n} + r_{2n}}{d_{1n} + d_{2n} + d_{3n}}. \quad (3)$$

²It should be pointed out when $\lim_{n \rightarrow \infty} \sqrt{n}/h_n = 0$ (and $\lim_{n \rightarrow \infty} h_n/n = 0$, where h_n^{-1} is the order of magnitude of elements of \mathbf{W}_n), $\hat{\boldsymbol{\theta}}_n$ is consistent, as shown in Lee (2002), and thus one may not need to seek a consistent estimator of λ_0 separately and then use it to construct a consistent estimator of $\boldsymbol{\beta}_0$. In practice, one may not know *a priori* the rate of h_n , but the II estimator to be introduced is always consistent regardless of the rate of h_n .

The following assumptions are made throughout this paper.

Assumption 1. (i) $\forall i \neq j$, $w_{ij,n} = O(h_n^{-1})$, where the rate sequence $\{h_n\}$ is uniformly bounded away from zero and $\lim_{n \rightarrow \infty} h_n/n = 0$; (ii) The sequence $\{\mathbf{W}_n\}$ is uniformly bounded in row and column sums; (iii) $w_{ii,n} = 0$.

Assumption 2. (i) \mathbf{S}_n^{-1} exists; (ii) The sequence $\{\mathbf{S}_n^{-1}\}$ is uniformly bounded in row and column sums.

Assumption 3. The error terms $\{u_{i,n}\}$ in $\mathbf{u}_n = (u_{1,n}, \dots, u_{n,n})'$ have following properties: (i) $E(u_{i,n}) = 0$; (ii) $E(|u_{i,n}|^{4+\delta}) < \infty$ for some positive constant δ ; (iii) $u_{i,n}$ and $u_{j,n}$ are independent for any $i \neq j$.

Assumption 4. λ_0 is contained in a compact parameter space Λ . For any admissible $\lambda \in \Lambda$, $\{\mathbf{S}_n^{-1}(\lambda)\}$ is uniformly bounded in row and column sums.

Assumption 5. (i) The elements of \mathbf{X}_n are uniformly bounded constants for all n ; (ii) The probability limit $\mathbf{\Gamma}$ of $n^{-1}\mathbf{Z}'_n\mathbf{Z}_n$ exists and is nonsingular.

Assumption 6. (i) $\lim_{n \rightarrow \infty} n^{-1}\text{Var}(r_n)$ exists and is nonzero; (ii) $\lim_{n \rightarrow \infty} n^{-1}\text{Var}(\mathbf{u}'_n\mathbf{E}_n\mathbf{u}_n + \beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\mathbf{u}_n)$ exists and is nonzero.

Intuitions and related discussions for Assumptions 1–5 are provided in Kelejian and Prucha (1999) and Lee (2001, 2002, 2004). Assumption 1.(i) follows naturally when \mathbf{W}_n is row- or column-normalized, as is typically the case. Assumption 1.(ii) and Assumption 2 limit the degree of dependency among the spatial units and is originated by Kelejian and Prucha (1999). Given Assumption 2, the equilibrium solution of \mathbf{y}_n is $\mathbf{y}_n = \mathbf{S}_n^{-1}\mathbf{X}_n\beta_0 + \mathbf{S}_n^{-1}\mathbf{u}_n$. Under Assumption 3, let $\sigma_{i,n}^2 = E(u_{i,n}^2)$, $\mathbf{\Sigma}_n = \text{Dg}(\sigma_{1,n}^2, \dots, \sigma_{n,n}^2)$, $\mathbf{\Sigma}_n^{(3)} = \text{Dg}(E(u_{1,n}^3), \dots, E(u_{n,n}^3))$, and $\mathbf{\Sigma}_n^{(4)} = \text{Dg}(E(u_{1,n}^4) - 3\sigma_{1,n}^4, \dots, E(u_{n,n}^4) - 3\sigma_{n,n}^4)$. Lee (2002) emphasized that Assumption 5.(ii) is related to an identification condition for estimation in the least squares and IV frameworks. It rules out possible multicollinearities among \mathbf{X}_n and $\mathbf{G}_n\mathbf{X}_n\beta_0$ for large n and implies that the limit of d_{3n}/n is bounded away from zero.³ Assumption 6 ensures that the asymptotic variances of the (properly recentered) OLS estimator and the resulting II estimator are positive.⁴

³Multicollinearities can happen, for example, when $\mathbf{X}_n = \mathbf{1}_n$, namely, it consists of only a constant term. Lee (2004) showed that under homoscedasticity, however, the QML works in violation of this condition. Since the II estimator to be discussed in this paper is to correct the possible inconsistency of the OLS estimator, Assumption 5.(ii) is maintained.

⁴The asymptotic variances are given by $\lim_{n \rightarrow \infty} n^{-1}\text{Var}(r_n)/[\text{plim}(d_n/n)]^2 = \lim_{n \rightarrow \infty} [\text{Var}(r_n)/n]/[E(d_n)/n]^2$ and $\lim_{n \rightarrow \infty} n^{-1}\text{Var}(\mathbf{u}'_n\mathbf{E}_n\mathbf{u}_n + \beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\mathbf{u}_n)/[\text{plim}(d_n/n)]^2 =$

2.1 The Asymptotic Behavior of the OLS estimator

From Lemma 6 in the appendix, r_{1n} and d_{1n} are both $O_P(n/h_n)$, r_{2n} and d_{2n} are both $O_P(\sqrt{n})$, and $d_{3n} = O(n)$. The asymptotic properties of the OLS estimator crucially depend on the magnitude of h_n .

When h_n is bounded, the OLS estimator cannot be consistent, since now

$$\hat{\lambda}_n - \lambda_0 = \frac{\frac{1}{n}r_{1n}}{\frac{1}{n}d_{1n} + \frac{1}{n}d_{3n}} + o_P(1),$$

but the probability limit of the numerator $n^{-1}r_{1n}$ is typically nonzero.

If $h_n \rightarrow \infty$, d_{3n} dominates the denominator in (3), then

$$\hat{\lambda}_n - \lambda_0 = \frac{\frac{1}{h_n} \left(\frac{h_n}{n} r_{1n} \right) + \frac{1}{\sqrt{n}} \left(\frac{1}{\sqrt{n}} r_{2n} \right)}{\frac{1}{n} d_{3n}} + o_P(1) \xrightarrow{p} 0,$$

indicating that $\hat{\lambda}_n$ is consistent as long as $h_n \rightarrow \infty$.

More interestingly, the behavior of $\hat{\lambda}_n$ depends on how fast h_n diverges to infinity. If h_n tends to infinity at a rate slower than \sqrt{n} , then r_{1n} dominates r_{2n} in the numerator, and

$$h_n(\hat{\lambda}_n - \lambda_0) = \frac{\frac{h_n}{n} r_{1n}}{\frac{1}{n} d_{3n}} + o_P(1), \quad (4)$$

which implies that $\hat{\lambda}_n$ converges at the slower rate h_n , but it does not converge to λ_0 at rate h_n ,

as $\text{plim}_{n \rightarrow \infty} (h_n/n)r_{1n} = \lim_{n \rightarrow \infty} (h_n/n)\text{tr}(\Sigma_n \mathbf{M}_n \mathbf{G}_n)$ is typically nonzero.

If $h_n = O(\sqrt{n})$, r_{1n} and r_{2n} are of the same order in the numerator, so

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) = \frac{\frac{1}{\sqrt{n}}r_{1n} + \frac{1}{\sqrt{n}}r_{2n}}{\frac{1}{n}d_{3n}} + o_P(1), \quad (5)$$

indicating that $\hat{\lambda}_n$ converges at rate \sqrt{n} , but at this rate it does not converge to λ_0 , as in general

$\lim_{n \rightarrow \infty} [\text{Var}(\mathbf{u}'_n \mathbf{E}_n \mathbf{u}_n + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{u}_n)/n] / [\text{E}(d_n)/n]^2$, respectively, for the (properly recentered) OLS estimator and the resulting II estimator. Their explicit expressions are given respectively in Theorems 1 and 2 to be introduced. Assumption 5.(ii) implies that $\text{plim}_{n \rightarrow \infty} (d_n/n) = \text{plim}_{n \rightarrow \infty} (\mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n/n)$ exists and is nonzero. It can be shown (see the appendix) that $\text{Var}(r_n) = \text{Var}(\mathbf{u}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n) + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \Sigma_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0 + 2\text{Cov}(\beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \Sigma_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0, \mathbf{u}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n)$, where the covariance term disappears under normality, and $\text{Var}(\mathbf{u}'_n \mathbf{E}_n \mathbf{u}_n + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{u}_n) = \text{Var}(\mathbf{u}'_n \mathbf{E}_n \mathbf{u}_n) + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \Sigma_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0$. When h_n diverges, $\beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \Sigma_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0$ is the dominating term in $\text{Var}(r_n)$ as well as $\text{Var}(\mathbf{u}'_n \mathbf{E}_n \mathbf{u}_n + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{u}_n)$. Then the usual condition that $\text{plim}_{n \rightarrow \infty} n^{-1} \mathbf{Z}'_n \Sigma_n \mathbf{Z}_n$ exists and is nonsingular is sufficient for Assumption 6 to hold. When h_n is bounded, a more precise characterization of a sufficient condition is not immediately obvious. Essentially, it requires, in addition to the existence and nonsingularity of $\text{plim}_{n \rightarrow \infty} n^{-1} \mathbf{Z}'_n \Sigma_n \mathbf{Z}_n$, the existence of $\lim_{n \rightarrow \infty} n^{-1} \text{Var}(\mathbf{u}'_n \mathbf{E}_n \mathbf{u}_n)$ and $\lim_{n \rightarrow \infty} n^{-1} \text{Var}(\mathbf{u}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n)$, where $\text{Var}(\mathbf{u}'_n \mathbf{E}_n \mathbf{u}_n) = O(n/h_n)$ and $\text{Var}(\mathbf{u}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n) = O(n/h_n)$.

$\text{plim}_{n \rightarrow \infty} n^{-1/2} r_{1n} = \lim_{n \rightarrow \infty} n^{-1/2} \text{tr}(\boldsymbol{\Sigma}_n \mathbf{M}_n \mathbf{G}_n)$ is nonzero.

If h_n tends to infinity at a rate faster than \sqrt{n} (and yet smaller than n), namely, $\lim_{n \rightarrow \infty} \sqrt{n}/h_n = 0$ (and $\lim_{n \rightarrow \infty} h_n/n = 0$), then in the numerator r_{2n} dominates r_{1n} ,

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) = \frac{\frac{1}{\sqrt{n}} r_{2n}}{\frac{1}{n} d_{3n}} + o_P(1), \quad (6)$$

indicating that $\hat{\lambda}_n$ converges to λ_0 at rate \sqrt{n} and is asymptotically normal if one applies a central limit theorem to $n^{-1/2} r_{2n}$.

One sees that the asymptotic behavior of $\hat{\lambda}_n$ depends on the magnitude of h_n , which may be unknown in practice. The following theorem shows that if one can properly recenter $\hat{\lambda}_n$, then a unified asymptotic distribution result follows.

Theorem 1. *Under Assumptions 1–6, the OLS estimator $\hat{\lambda}_n$ of λ_0 in the model (1) has the following asymptotic distribution,*

$$\sqrt{n} \left(\hat{\lambda}_n - \lambda_0 - \frac{\text{tr}(\boldsymbol{\Sigma}_n \mathbf{M}_n \mathbf{G}_n)}{\mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n} \right) \xrightarrow{d} \text{N}(0, v), \quad (7)$$

where

$$v = \lim_{n \rightarrow \infty} \frac{n \text{Var}(r_n)}{[\text{tr}(\boldsymbol{\Sigma}_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n) + \boldsymbol{\beta}'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \boldsymbol{\beta}_0]^2}$$

with

$$\begin{aligned} \text{Var}(r_n) &= \text{tr}(\boldsymbol{\Sigma}_n^{(4)} \odot \mathbf{M}_n \mathbf{G}_n \odot \mathbf{M}_n \mathbf{G}_n) + \text{tr}[\boldsymbol{\Sigma}_n \mathbf{M}_n \mathbf{G}_n \boldsymbol{\Sigma}_n (\mathbf{M}_n \mathbf{G}_n + \mathbf{G}'_n \mathbf{M}_n)] \\ &+ \boldsymbol{\beta}'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \boldsymbol{\Sigma}_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \boldsymbol{\beta}_0 + 2 \boldsymbol{\beta}'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \text{dg}(\boldsymbol{\Sigma}_n^{(3)} \odot \mathbf{M}_n \mathbf{G}_n). \end{aligned}$$

Remark 1. *The recentering term $\text{tr}(\boldsymbol{\Sigma}_n \mathbf{M}_n \mathbf{G}_n)/\mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n$ is in fact $\text{E}(r_n)/d_n$. One could have recentered $\hat{\lambda}_n - \lambda_0$ by $\text{E}(r_n)/\text{E}(d_n)$. (This is the approach taken by Kyriacou et al. (2017) when dealing with the special case of pure SAR model.) But then by following a similar expansion as in the proof of Theorem 1, one can find that the asymptotic variance of the resulting recentered estimator is much more complicated, involving the variances of r_n and d_n as well as their covariance.*

Remark 2. *When $\lim_{n \rightarrow \infty} \sqrt{n}/h_n = 0$, the recentering term $\text{tr}(\boldsymbol{\Sigma}_n \mathbf{M}_n \mathbf{G}_n)/\mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n$ is in fact $o_P(n^{-1/2})$ and the asymptotic distribution of $\sqrt{n}(\hat{\lambda}_n - \lambda_0)$ centers at zero, so one does*

not need to recenter $\hat{\lambda}_n - \lambda_0$ by $\text{tr}(\boldsymbol{\Sigma}_n \mathbf{M}_n \mathbf{G}_n) / \mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n$. (Nor does one need to recenter it by $\mathbf{y}'_n \mathbf{S}'_n \mathbf{M}_n \mathbf{D}_n \mathbf{M}_n \mathbf{S}_n \mathbf{y}_n / \mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n$, which is also $o_P(n^{-1/2})$, as in Theorem 2.)

Remark 3. When h_n diverges, one may single out dominating terms in $\text{Var}(r_n)$ and $\text{E}(d_n)$ so that a finer expression of $v = \lim_{n \rightarrow \infty} [\text{Var}(r_n)/n] / [\text{E}(d_n)/n]^2$ can be presented. For example, under divergent h_n , one can replace $\text{Var}(r_n)$ with $\text{Var}(r_{2n}) = \boldsymbol{\beta}'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \boldsymbol{\Sigma}_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \boldsymbol{\beta}_0$ and replace $\text{E}(d_n)$ with its dominating term $\boldsymbol{\beta}'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \boldsymbol{\beta}_0$. When h_n is bounded, however, such replacements are not available and one needs to keep all the term in $\text{Var}(r_n)$ and $\text{E}(d_n)$. Moreover, since $\text{Var}(r_n)$ depends on higher-order moments of u , namely, $\boldsymbol{\Sigma}_n^{(3)}$ and $\boldsymbol{\Sigma}_n^{(4)}$, then, with bounded h_n , the asymptotic variance v of the recentered OLS estimator depends on them, too.

Remark 4. In view of Remark 3, under divergent h_n , if further $\boldsymbol{\Sigma}_n = \sigma_0^2 \mathbf{I}_n$, namely, under homoscedasticity with $\sigma_0^2 = \text{E}(u_{i,n}^2)$, then v corresponds to the top left element of $\sigma_0^2 \boldsymbol{\Gamma}^{-1}$. This, together with Remark 2, is in line with the observation in Lee (2002) that the (consistent, uncentered) OLS estimator (when $\lim_{n \rightarrow \infty} \sqrt{n}/h_n = 0$) has the same limiting distribution of the optimal IV estimator and under normality it has the same limiting distribution of the ML estimator. It also implies that under other cases of divergent h_n (slower than or equal to rate \sqrt{n}), as long as the OLS estimator is properly recentered, it achieves the same limiting distribution.

Remark 5. Theorem 1 gives a unified representation of the asymptotic distribution of the properly recentered OLS estimator, regardless of the possibly unknown magnitude of h_n . Further, it facilitates the construction of the indirect inference estimator to be introduced that corrects the inconsistency, when present, of the OLS estimator.

2.2 The Indirect Inference Estimator

One can see from Theorem 1 in the previous subsection that the OLS estimator $\hat{\lambda}_n$ may have an asymptotic bias. Yet a direct feasible bias-correction procedure is not possible, since the bias itself depends on unknown parameters, including λ_0 , which may not be consistently estimated by the OLS estimator $\hat{\lambda}_n$. Following Phillips (2012) and Kyriacou et al. (2017), one may define the binding function that involves the bias of $\hat{\lambda}_n$. Unfortunately, the resulting binding function then involves the unknown $\boldsymbol{\Sigma}_n$, which appears in the recentering quantity for the OLS estimator in (7). The strategy in this paper is to replace $\text{tr}(\boldsymbol{\Sigma}_n \mathbf{M}_n \mathbf{G}_n) = \text{E}(\mathbf{u}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n)$ with a term that is of the same order of $\mathbf{u}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n$ and at the same time does not directly involve $\boldsymbol{\Sigma}_n$.

Recall that $D_n = \text{Dg}(M_n G_n)$. Then under a general form of unknown heteroscedasticity, $\text{tr}(\Sigma_n M_n G_n) = \text{tr}(\Sigma_n D_n) = E(\mathbf{u}'_n D_n \mathbf{u}_n)$, where $\mathbf{u}_n = \mathbf{S}_n \mathbf{y}_n - \mathbf{X}_n \beta_0$. If λ_0 is known, then β_0 can be consistently estimated by $\tilde{\beta}_n = \tilde{\beta}_n(\lambda_0) = (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{S}_n \mathbf{y}_n$. Let $\tilde{\mathbf{u}}_n = \tilde{\mathbf{u}}_n(\lambda_0) = \mathbf{S}_n \mathbf{y}_n - \mathbf{X}_n \tilde{\beta}_n = M_n \mathbf{S}_n \mathbf{y}_n$. Now one may be able to replace $E(\mathbf{u}'_n D_n \mathbf{u}_n)$ with $\tilde{\mathbf{u}}'_n D_n \tilde{\mathbf{u}}_n = \mathbf{y}'_n \mathbf{S}'_n M_n D_n M_n \mathbf{S}_n \mathbf{y}_n$ and use $\mathbf{y}'_n \mathbf{S}'_n M_n D_n M_n \mathbf{S}_n \mathbf{y}_n / \mathbf{y}'_n \mathbf{W}'_n M_n \mathbf{W}_n \mathbf{y}_n$ as the recentering quantity.

Theorem 2. *Under Assumptions 1–6, the OLS estimator $\hat{\lambda}_n$ of λ_0 in the model (1) has the following asymptotic distribution:*

$$\sqrt{n} \left(\hat{\lambda}_n - \lambda_0 - \frac{\mathbf{y}'_n \mathbf{S}'_n M_n D_n M_n \mathbf{S}_n \mathbf{y}_n}{\mathbf{y}'_n \mathbf{W}'_n M_n \mathbf{W}_n \mathbf{y}_n} \right) \xrightarrow{d} N(0, \eta), \quad (8)$$

where

$$\eta = \lim_{n \rightarrow \infty} \frac{n \{ \text{tr}[\Sigma_n \mathbf{E}_n \Sigma_n (\mathbf{E}_n + \mathbf{E}'_n)] + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n M_n \Sigma_n M_n \mathbf{G}_n \mathbf{X}_n \beta_0 \}}{[\text{tr}(\Sigma_n \mathbf{G}'_n M_n \mathbf{G}_n) + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n M_n \mathbf{G}_n \mathbf{X}_n \beta_0]^2}.$$

Remark 6. *When one replaces $\text{tr}(\Sigma_n M_n \mathbf{G}_n)$ that involves the unknown Σ_n and appears in the recentering term of the OLS estimator, the asymptotic variance η of the newly recentered estimator no longer involves $\Sigma_n^{(3)}$ and $\Sigma_n^{(4)}$. This stands in contrast to the asymptotic variance v (see Remark 3). So replacing $\text{tr}(\Sigma_n M_n \mathbf{G}_n)$ with $\mathbf{y}'_n \mathbf{S}'_n M_n D_n M_n \mathbf{S}_n \mathbf{y}_n$ facilitates not only the construction of the indirect inference estimator to be introduced but also the inference procedure.*

Given Theorem (2) and the observed sample data \mathbf{y}_n and \mathbf{X}_n , one can always define the sample binding function. Recall that $\mathbf{S}_n(\lambda) = \mathbf{I}_n - \lambda \mathbf{W}_n$ and $D_n(\lambda) = \text{Dg}(M_n \mathbf{G}_n(\lambda)) = \text{Dg}(M_n \mathbf{W}_n \mathbf{S}_n^{-1}(\lambda))$ are functions of the parameter λ (as well as \mathbf{X}_n). So the binding function can be defined as

$$b_n(\lambda) = \lambda + \frac{\mathbf{y}'_n \mathbf{S}'_n(\lambda) M_n D_n(\lambda) M_n \mathbf{S}_n(\lambda) \mathbf{y}_n}{\mathbf{y}'_n \mathbf{W}'_n M_n \mathbf{W}_n \mathbf{y}_n} \quad (9)$$

and the II estimator inverts this binding function:

$$\hat{\lambda}_n^{II} = b_n^{-1}(\hat{\lambda}_n). \quad (10)$$

Intuitively, the II estimator defined as such tries to match $\hat{\lambda}_n$ from the observed data to its expected

tation, at least approximately. Typically, the expectation may be approximated, to an arbitrary degree of accuracy, via the method of simulations, as in the original spirit of Gouriéroux et al. (1993) and Smith (1993). But in simulations one needs to make some distributional assumption to generate the pseudo error term. So instead one may use some analytical approximation as in Phillips (2012), Kyriacou et al. (2017), and this paper.

In the definition of the II estimator as in (10), it is implicitly assumed that the binding function $b_n(\lambda)$ is invertible. Note that $b_n(\lambda)$ is a function of λ and the sample data \mathbf{y}_n and \mathbf{X}_n and thus it is random.⁵

Assumption 7. *For all $\lambda \in \Lambda$, the binding function (9) is monotonic in λ with probability 1 and when h_n is bounded, $b'_n(\lambda_0) \xrightarrow{a.s.} b_0 \neq 0$, where*

$$b'_n(\lambda_0) = 1 + \frac{\mathbf{y}'_n \mathbf{S}'_n \mathbf{M}_n \text{Dg}(\mathbf{M}_n \mathbf{G}_n^2) \mathbf{M}_n \mathbf{S}_n \mathbf{y}_n - 2 \mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{D}_n \mathbf{M}_n \mathbf{S}_n \mathbf{y}_n}{\mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n}. \quad (11)$$

One can see from the expression (A.10) in the appendix that the derivative of the binding function with respect to λ is

$$b'_n(\lambda) = 1 + O_P\left(\frac{1}{h_n}\right).$$

For divergent h_n , one can show that the $O_P(h_n^{-1})$ term converges almost surely to zero for all $\lambda \in \Lambda$. So in large samples, Assumption 7 is more likely to hold. Assumption 7 lists the conditions under which the II estimator exists and is consistent. It would be desirable if one could lay down some primitive conditions on the data matrix \mathbf{X}_n , the weight matrix \mathbf{W}_n , and the parameter space Λ so that Assumption 7 would be satisfied. Given the sample data, one may plot the binding function against λ to verify numerically the validity of this assumption. Simulations as in Gospodinov et al. (2017) may also help to establish this assumption's credibility.

Theorem 3. *For the model (1), under Assumptions 1–7, the II estimator $\hat{\lambda}_n^{II}$ of λ_0 , defined as in (10) that is based on the binding function (9), has the following asymptotic distribution:*

$$\sqrt{n}(\hat{\lambda}_n^{II} - \lambda_0) \xrightarrow{d} \text{N}\left(0, \frac{\eta}{b_0^2}\right). \quad (12)$$

⁵The use of observed, endogenous but non-simulated, variables within the binding function does not appear to be common. An interesting example is Gospodinov et al. (2017), where the authors used observed data within the binding function to hedge against misspecification bias. In their set-up of the autoregressive distributed lag model with a latent scalar predictor under the presence of measurement error, a similar technical difficulty exists regarding the invertibility condition of their binding function and they resorted to simulations to approximate the binding function and then the invertibility condition is numerically verified based on the approximated binding function.

Remark 7. Since when h_n diverges, $b'_n(\lambda_0)$ converges almost surely to 1, the asymptotic distribution of the II estimator is identical to that of the properly recentered OLS estimator. Further, under divergent h_n , $\beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \Sigma_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0$ dominates $\text{Var}(\mathbf{u}'_n \mathbf{E}_n \mathbf{u}_n + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{u}_n)$ and $\beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0$ dominates $\text{E}(d_n)$, so $\eta = v + o(1)$. In light of Remark 4, this means that under homoscedasticity, the II estimator is as efficient as the optimal IV estimator and can be as efficient as the ML estimator if the spatial data is normally distributed.

Remark 8. Theorem 3 shows that regardless of the magnitude of h_n , which researchers may not know in practice, one can always apply the II procedure after the OLS estimation is done. At worst, when $\lim_{n \rightarrow \infty} \sqrt{n}/h_n = 0$, this procedure is redundant (see Remark 2), but still the resulting II estimator has exactly the same asymptotic distribution as the consistent OLS estimator (since $\eta = v + o(1)$, see Remark 7). Otherwise, the II procedure provides a correction to the inconsistent OLS estimator.

Once the spatial autoregression parameter λ is consistently estimated by $\hat{\lambda}_n^{II}$, one can estimate the parameter vector β by

$$\hat{\beta}_n^{II} = (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \hat{\mathbf{S}}_n \mathbf{y}_n, \quad (13)$$

where $\hat{\mathbf{S}}_n = \mathbf{S}_n(\hat{\lambda}_n^{II}) = \mathbf{S}_n - (\hat{\lambda}_n^{II} - \lambda_0) \mathbf{W}_n$.

Theorem 4. For the model (1), under Assumptions 1–7, the OLS estimator of β_0 defined as in (13) has the asymptotic distribution

$$\sqrt{n}(\hat{\beta}_n^{II} - \beta_0) \xrightarrow{d} \text{N}(0, \mathbf{V}),$$

and jointly,

$$\sqrt{n} \begin{pmatrix} \hat{\lambda}_n^{II} - \lambda_0 \\ \hat{\beta}_n^{II} - \beta_0 \end{pmatrix} \xrightarrow{d} \text{N} \left(\mathbf{0}, \begin{pmatrix} \frac{\eta}{b_0^2} & \gamma' \\ \gamma & \mathbf{V} \end{pmatrix} \right),$$

where \mathbf{V} , assumed to exist and be positive definite, is given by (A.14) and γ is given by (A.15), respectively, in the appendix.

Remark 9. One can see (from the appendix) that the expression of \mathbf{V} contains the traditional OLS variance term under heteroscedasticity, $\lim_{n \rightarrow \infty} n(\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \Sigma_n \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1}$, as well as terms that signal the additional uncertainty introduced by $\hat{\lambda}_n^{II}$ in the definition of $\hat{\beta}_n^{II}$ as in (13).

In practice, in order to make asymptotically valid inference from the II estimation strategy,

one needs to estimate η/b_0^2 , \mathbf{V} , and γ in Theorems 3 and 4. As η/b_0^2 , \mathbf{V} , and γ involve the unknown \mathbf{G}_n , \mathbf{S}_n , β_0 , and Σ_n , one may replace \mathbf{G}_n by $\hat{\mathbf{G}}_n = \mathbf{G}_n(\hat{\lambda}_n^{II})$, \mathbf{D}_n by $\hat{\mathbf{D}}_n = \text{Dg}(\mathbf{M}_n \hat{\mathbf{G}}_n)$, \mathbf{E}_n by $\hat{\mathbf{E}}_n = \mathbf{M}_n \hat{\mathbf{G}}_n - \hat{\mathbf{D}}_n$, \mathbf{S}_n by $\hat{\mathbf{S}}_n = \mathbf{S}_n(\hat{\lambda}_n^{II})$, β_0 by $\hat{\beta}_n^{II}$, and Σ_n by $\text{Dg}(\hat{u}_{1,n}^2, \dots, \hat{u}_{n,n}^2)$, where $\hat{u}_{i,n}$'s are the sample residuals from the II estimation.⁶

3 The Special Case of Pure SAR

It is worthwhile to discuss the case when there is no \mathbf{X}_n , namely, the so-called pure SAR model

$$\mathbf{y}_n = \lambda \mathbf{W}_n \mathbf{y}_n + \mathbf{u}_n. \quad (14)$$

This case is of special interest since there is no IV available. On the other hand, the QML estimator is not consistent under heteroscedasticity. Kyriacou et al. (2017) were the first to explore the possibility of using the II procedure to correct the inconsistency of the OLS estimator under some mild form of heteroscedasticity and their results were quite promising. In this paper, no restrictions are imposed on the form of the unknown heteroscedasticity.

Given the expansion (3), one can see obviously that

$$\text{plim}_{n \rightarrow \infty} (\hat{\lambda}_n - \lambda_0) = \frac{\text{plim}_{n \rightarrow \infty} \frac{h_n}{n} r_{1n}}{\text{plim}_{n \rightarrow \infty} \frac{h_n}{n} d_{1n}} = \frac{\lim_{n \rightarrow \infty} \frac{h_n}{n} \text{tr}(\Sigma_n \mathbf{G}_n)}{\lim_{n \rightarrow \infty} \frac{h_n}{n} \text{tr}(\Sigma_n \mathbf{G}_n' \mathbf{G}_n)} \neq 0,$$

regardless of the magnitude of h_n . Proceeding similarly as before, as long as $h_n = o(n)$, one can write $r_{1n} - \mathbf{E}(r_{1n}) = O_P(\sqrt{n/h_n})$, $d_{1n} - \mathbf{E}(d_{1n}) = O_P(\sqrt{n/h_n})$, and $\mathbf{E}(d_{1n}) = O(n/h_n)$. (See Lemma 6 in the appendix.) Then, by a Nagar-type (Nagar, 1959) expansion,

$$\begin{aligned} \sqrt{\frac{n}{h_n}} \left(\hat{\lambda}_n - \lambda_0 - \frac{\mathbf{E}(r_{1n})}{d_{1n}} \right) &= \sqrt{\frac{n}{h_n}} \left(\frac{r_{1n} - \mathbf{E}(r_{1n})}{\mathbf{E}(d_{1n}) + d_{1n} - \mathbf{E}(d_{1n})} \right) \\ &= \sqrt{\frac{n}{h_n}} \frac{r_{1n} - \mathbf{E}(r_{1n})}{\mathbf{E}(d_{1n})} + o_P(1). \end{aligned} \quad (15)$$

Assumption 6 needs to be modified accordingly to ensure the asymptotic variance of the properly recentered $\hat{\lambda}_n$ exists and is positive. Now let $\mathbf{D}_n = \text{Dg}(\mathbf{G}_n)$ and $\mathbf{E}_n = \mathbf{G}_n - \mathbf{D}_n$.

⁶This follows similarly from the proof of Proposition 2 of Lin and Lee (2010).

Assumption 6'. (i)

$$v = \lim_{n \rightarrow \infty} \frac{n\{\text{tr}(\boldsymbol{\Sigma}_n^{(4)} \odot \mathbf{G}_n \odot \mathbf{G}_n) + \text{tr}[\boldsymbol{\Sigma}_n \mathbf{G}_n \boldsymbol{\Sigma}_n (\mathbf{G}_n + \mathbf{G}'_n)]\}}{h_n [\text{tr}(\boldsymbol{\Sigma}_n \mathbf{G}'_n \mathbf{G}_n)]^2}$$

exists and is positive; (ii)

$$\eta = \lim_{n \rightarrow \infty} \frac{n \text{tr}[\boldsymbol{\Sigma}_n \mathbf{E}_n \boldsymbol{\Sigma}_n (\mathbf{E}_n + \mathbf{E}'_n)]}{h_n [\text{tr}(\boldsymbol{\Sigma}_n \mathbf{G}'_n \mathbf{G}_n)]^2}$$

exists and is positive.

Corollary 1. Under Assumptions 1–4 and 6', the OLS estimator $\hat{\lambda}_n$ of λ in the model (14) has the following asymptotic distribution,

$$\sqrt{\frac{n}{h_n}} \left(\hat{\lambda}_n - \lambda_0 - \frac{\text{tr}(\boldsymbol{\Sigma}_n \mathbf{G}_n)}{\mathbf{y}'_n \mathbf{W}'_n \mathbf{W}_n \mathbf{y}_n} \right) \xrightarrow{d} \text{N}(0, v), \quad (16)$$

where v is defined as in Assumption 6' (i).

Corollary 2. Under Assumptions 1–4 and 6', the OLS estimator $\hat{\lambda}_n$ of λ in the model (14) has the following asymptotic distribution:

$$\sqrt{\frac{n}{h_n}} \left(\hat{\lambda}_n - \lambda_0 - \frac{\mathbf{y}'_n \mathbf{S}'_n \mathbf{D}_n \mathbf{S}_n \mathbf{y}_n}{\mathbf{y}'_n \mathbf{W}'_n \mathbf{W}_n \mathbf{y}_n} \right) \xrightarrow{d} \text{N}(0, \eta), \quad (17)$$

where η is defined as in Assumption 6' (ii).

Let the sample binding function be

$$b_n(\lambda) = \lambda + \frac{\mathbf{y}'_n \mathbf{S}'_n(\lambda) \mathbf{D}_n(\lambda) \mathbf{S}_n(\lambda) \mathbf{y}_n}{\mathbf{y}'_n \mathbf{W}'_n \mathbf{W}_n \mathbf{y}_n}. \quad (18)$$

Accordingly, Assumption 7 is modified as follows.

Assumption 7'. For all λ in Λ , the binding function (18) is monotonic in λ with probability 1 and $b'_n(\lambda_0) \xrightarrow{a.s.} b_0 \neq 0$, where

$$b'_n(\lambda_0) = 1 + \frac{\mathbf{y}'_n \mathbf{S}'_n \text{Dg}(\mathbf{G}_n^2) \mathbf{S}_n \mathbf{y}_n - 2 \mathbf{y}'_n \mathbf{W}'_n \mathbf{D}_n \mathbf{S}_n \mathbf{y}_n}{\mathbf{y}'_n \mathbf{W}'_n \mathbf{W}_n \mathbf{y}_n}. \quad (19)$$

Corollary 3. For the model (14), under Assumptions 1–4, 6', and 7', the II estimator $\hat{\lambda}_n^{II}$ of λ

that is based on the binding function (18) has the following asymptotic distribution:

$$\sqrt{\frac{n}{h_n}} \left(\hat{\lambda}_n^{II} - \lambda_0 \right) \xrightarrow{d} N \left(0, \frac{\eta}{b_0^2} \right), \quad (20)$$

where η is defined as in Assumption 6' (ii).

Remark 10. If $\Sigma_n = \sigma_0^2 \mathbf{I}_n$ (namely, under homoscedasticity) and ones uses $E(r_{1n})/E(d_{1n}) = \text{tr}(\mathbf{G}_n)/\text{tr}(\mathbf{G}'_n \mathbf{G}_n)$ as the recentering term, which fortunately does not involve the unknown variance parameter σ_0^2 , then the binding function, its derivative, and the asymptotic distribution of the resulting II estimator can be modified accordingly as in Kyriacou et al. (2017). In contrast to the SAR model with \mathbf{X}_n , the recentered OLS estimator and the II estimator in the pure SAR model have convergence rate $\sqrt{n/h_n}$.

Remark 11. One can see that $b'_n(\lambda_0)$ does not converge almost surely to 1 when h_n is divergent as in the case when \mathbf{X} is present. This is because $\mathbf{y}'_n \mathbf{W}'_n \mathbf{W}_n \mathbf{y}_n = O_P(n/h_n)$ for the pure case and $\mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n = O_P(n)$ when \mathbf{X} is present, whereas both the numerators in the second terms on the right-hand sides of (19) and (11) are $O_P(n/h_n)$.

Remark 12. Admittedly, the convergence rate in (20) depends on h_n . However, this does not prevent one from using the II estimator if one is interested in estimating the pure SAR model, since the binding function (18) does not involve h_n . For inference purpose, as η defined as in Assumption 6' (ii) has the scaling factor n/h_n , once $\hat{\lambda}_n^{II}$ and the sample residuals $\hat{u}_{i,n}$ are available, the standard error of $\hat{\lambda}_n^{II}$ can be calculated as $\text{tr}[\hat{\Sigma}_n \hat{\mathbf{E}}_n \hat{\Sigma}_n (\hat{\mathbf{E}}_n + \hat{\mathbf{E}}'_n)] / [\text{tr}(\hat{\Sigma}_n \hat{\mathbf{G}}'_n \hat{\mathbf{G}}_n)]^2$, where $\hat{\mathbf{G}}_n = \mathbf{G}_n(\hat{\lambda}_n^{II})$, $\hat{\mathbf{E}}_n = \mathbf{E}_n(\hat{\lambda}_n^{II})$, and $\hat{\Sigma}_n = \text{Dg}(\hat{u}_{1,n}^2, \dots, \hat{u}_{n,n}^2)$.

4 Monte Carlo Evidence

In this section, Monte Carlo simulations are provided to demonstrate the performance of $\hat{\lambda}_n^{II}$ (as well as $\hat{\beta}_n^{II}$ in (13)) in finite samples, in comparison with consistent estimators that are also robust under heteroscedasticity: the optimal robust GMM estimator of Lin and Lee (2010) and the modified QML (MQML) estimator of Liu and Yang (2015). For the optimal robust GMM estimator of Lin and Lee (2010), $\mathbf{u}'_n \Sigma_n^{-1} (\mathbf{G}_n(\lambda) - \text{Dg}(\mathbf{G}_n(\lambda)) \mathbf{u}_n$ and $\mathbf{u}'_n \Sigma_n^{-1} (\mathbf{G}_n(\lambda) \mathbf{X}_n \boldsymbol{\beta}, \mathbf{X}_n)$ are used as the optimal moment conditions, see Debarsy et al. (2015). They involve λ (appearing in $\mathbf{G}_n(\lambda)$) and $\boldsymbol{\beta}$ as well as the covariance matrix Σ_n . For λ and $\boldsymbol{\beta}$, an initial estimation is constructed from the simple 2SLS with $\mathbf{W}_n \mathbf{X}_n$ and \mathbf{X}_n as IV's. One may assume a model for Σ_n

and then estimate the assumed model so as to construct the moment conditions. In this section, two choices are made regarding this: one is to use $\mathbf{u}'_n(\mathbf{G}_n - \text{Dg}(\mathbf{G}_n))\mathbf{u}_n$ and $\mathbf{u}'_n(\mathbf{G}_n\mathbf{X}_n\boldsymbol{\beta}, \mathbf{X}_n)$ as the moment conditions and the other is to use $\mathbf{u}'_n\boldsymbol{\Sigma}_n^{-1}(\mathbf{G}_n - \text{Dg}(\mathbf{G}_n))\mathbf{u}_n$ and $\mathbf{u}'_n\boldsymbol{\Sigma}_n^{-1}(\mathbf{G}_n\mathbf{X}_n\boldsymbol{\beta}, \mathbf{X}_n)$ with the true $\boldsymbol{\Sigma}_n$ (known in simulations) plugged in. The two resulting estimators are denoted by GMM and GMM($\boldsymbol{\Sigma}_n$), respectively, in Tables 1–4. One would expect that in practice the performance of the optimal robust GMM estimator with an estimated $\boldsymbol{\Sigma}_n$ appearing in the moment conditions would stand between the two.

In each experiment, for each estimator, reported are the bias and root mean squared error (RMSE) from 1,000 Monte Carlo simulations. The empirical rejection probabilities of the relevant t tests for testing each parameter equal to its true value at 5% are also reported, denoted by $P(5\%)$, where the asymptotic variances for λ_n^{II} and $\hat{\boldsymbol{\beta}}_n^{II}$, as discussed in the last paragraph of Section 2, are estimated with the unknowns replaced by their estimates based on the II procedure.⁷

For the purpose of comparison, the experimental design in Lin and Lee (2010) is followed closely. The spatial scenario under a group interaction weight matrix of Case (1991) is considered. The exogenous variables include a constant term and two independently distributed random variables following $N(3, 1)$ and $U(-1, 2)$, respectively. The size of each group is determined by a $U(3, 20)$ random variable. The error terms follow a zero-mean normal distribution with variances varying across groups. Two variance structures (V1 and V2) are considered. V1: for each group, if the group size is greater than 10, then the error variance is the same as the group size; otherwise, the variance is the inverse of the square of the group size. V2: for each group, the error variance is the inverse of the group size. Two sets of parameter configurations are used: $\boldsymbol{\theta}_0 = (\lambda_0, 0.8, 0.2, 1.5)'$ and $\boldsymbol{\theta}_0 = (\lambda_0, 0.2, 0.2, 0.1)'$, named P1 and P2, respectively. Different degrees of spatial autocorrelation are considered: $\lambda_0 = 0.2, 0.6, 0.9$. Results are reported in Tables 1–4.

⁷Neither Lin and Lee (2010) nor Liu and Yang (2015) reported how the inference procedures based on their estimators would perform in finite samples.

TABLE 1: Estimation of SAR with variance structure VI and parameter configuration P1

θ_0	MQML				GMM				GMM(Σ_n)				II			
	$R = 100$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$	$P(5\%)$	Bias	RMSE	$P(5\%)$	$P(5\%)$	Bias	RMSE	$P(5\%)$	$P(5\%)$
0.2	-0.009	0.067	0.073	3.6%	-0.015	0.073	0.073	5.8%	0.000	0.010	0.010	4.9%	-0.009	0.067	0.067	3.8%
0.8	0.034	0.367	0.377	4.5%	0.050	0.377	0.377	5.4%	0.001	0.036	0.036	5.6%	0.033	0.367	0.367	4.7%
0.2	-0.002	0.099	0.099	5.0%	-0.002	0.099	0.099	5.1%	0.000	0.008	0.008	5.5%	-0.002	0.099	0.099	5.0%
1.5	-0.004	0.088	0.088	4.9%	-0.005	0.088	0.088	5.1%	0.000	0.007	0.007	4.1%	-0.004	0.088	0.088	4.9%
0.6	-0.005	0.035	0.039	3.0%	-0.008	0.039	0.039	7.2%	0.000	0.005	0.005	5.3%	-0.005	0.035	0.035	4.1%
0.8	0.032	0.366	0.378	3.8%	0.047	0.378	0.378	5.9%	0.001	0.037	0.037	5.6%	0.031	0.367	0.367	4.7%
0.2	-0.001	0.102	0.102	5.5%	-0.001	0.102	0.102	5.5%	0.000	0.008	0.008	5.7%	-0.001	0.102	0.102	5.5%
1.5	0.004	0.089	0.089	4.9%	0.004	0.089	0.089	4.9%	0.000	0.007	0.007	4.6%	0.004	0.089	0.089	4.9%
0.9	-0.001	0.009	0.01	0.3%	-0.002	0.01	0.01	7.7%	0.000	0.001	0.001	6.1%	-0.001	0.009	0.009	4.3%
0.8	0.022	0.367	0.378	4.0%	0.037	0.378	0.378	6.0%	0.003	0.040	0.040	6.5%	0.02	0.367	0.367	5.6%
0.2	-0.002	0.102	0.102	4.4%	-0.002	0.102	0.102	4.4%	0.000	0.008	0.008	5.3%	-0.002	0.102	0.102	4.4%
1.5	0.000	0.093	0.094	5.9%	0.000	0.094	0.094	6.0%	0.000	0.007	0.007	4.7%	0.00	0.094	0.094	6.1%
0.2	0.000	0.047	0.051	5.6%	-0.003	0.051	0.051	8.0%	0.000	0.007	0.007	5.3%	0.000	0.047	0.047	6.1%
0.8	-0.005	0.255	0.261	5.1%	0.003	0.261	0.261	5.8%	0.001	0.026	0.026	6.9%	-0.005	0.255	0.255	5.2%
0.2	0.002	0.071	0.071	5.8%	0.001	0.071	0.071	5.7%	0.000	0.006	0.006	5.4%	0.002	0.071	0.071	5.8%
1.5	0.003	0.061	0.061	3.3%	0.002	0.061	0.061	3.4%	0.000	0.005	0.005	5.4%	0.003	0.061	0.061	3.3%
0.6	-0.003	0.025	0.028	2.2%	-0.005	0.028	0.028	8.6%	0.000	0.004	0.004	5.1%	-0.003	0.025	0.025	5.0%
0.8	0.021	0.256	0.264	4.1%	0.030	0.264	0.264	6.0%	0.000	0.026	0.026	5.2%	0.020	0.256	0.256	4.7%
0.2	-0.001	0.069	0.069	4.5%	-0.001	0.069	0.069	4.6%	0.000	0.006	0.006	5.7%	-0.001	0.069	0.069	4.5%
1.5	0.000	0.062	0.062	4.2%	-0.001	0.062	0.062	4.3%	0.000	0.005	0.005	4.9%	0.000	0.062	0.062	4.2%
0.9	-0.001	0.006	0.007	0.7%	-0.001	0.007	0.007	8.3%	0.000	0.001	0.001	6.0%	-0.001	0.006	0.006	4.7%
0.8	0.022	0.262	0.269	4.2%	0.030	0.269	0.269	6.7%	0.000	0.026	0.026	4.5%	0.021	0.262	0.262	5.7%
0.2	-0.001	0.072	0.072	5.5%	-0.001	0.072	0.072	5.3%	0.000	0.005	0.005	4.8%	-0.001	0.072	0.072	5.5%
1.5	0.000	0.063	0.063	3.8%	0.000	0.063	0.063	3.9%	0.000	0.005	0.005	4.4%	0.000	0.063	0.063	3.8%

Notes: The weight matrix is a social interaction matrix $\mathbf{W}_n = \mathbf{I}_R \otimes [(\mathbf{1}_{m_j} \mathbf{1}'_{m_j} - \mathbf{I}_{m_j}) / (m_j - 1)]$, $m_j \sim \text{IID } U(3, 20)$, $j = 1, \dots, R$. Reported for each estimator are the bias, RMSE, and the empirical size of the t test for each parameter equal to its true value at 5% from 1,000 simulations. The exogenous regressors are $(1, x_{i1}, x_{i2})$, $x_{i1} \sim \text{IID } N(3, 1)$, $x_{i2} \sim \text{IID } U(-1, 2)$, and x_{i1} is independent of x_{i2} . The group size m_j follows IID $U(3, 20)$. The error terms follow a zero-mean independent normal distribution. The variance structure is such that if $m_j > 10$, the error variance in the j -th group is m_j and otherwise is $1/m_j$, $j = 1, \dots, R$.

TABLE 2: Estimation of SAR with variance structure V1 and parameter configuration P2

θ_0	$R = 100$	MQML			GMM			GMM(Σ_n)			II		
		Bias	RMSE	P(5%)	Bias	RMSE	P(5%)	Bias	RMSE	P(5%)	Bias	RMSE	P(5%)
0.2		-0.010	0.076	5.5%	-0.014	0.078	5.9%	-0.003	0.049	5.4%	-0.009	0.076	5.7%
0.2		0.007	0.335	5.7%	0.013	0.336	6.0%	0.005	0.056	5.8%	0.007	0.335	5.8%
0.2		0.001	0.102	5.7%	0.001	0.102	5.7%	-0.001	0.008	6.3%	0.001	0.102	5.7%
0.1		0.000	0.095	5.9%	0.000	0.096	5.9%	0.000	0.007	6.0%	0.000	0.095	5.9%
0.6		-0.005	0.039	2.8%	-0.007	0.040	5.2%	-0.002	0.026	4.8%	-0.004	0.039	4.6%
0.2		-0.004	0.330	4.1%	0.002	0.331	4.3%	0.005	0.058	5.3%	-0.004	0.330	4.2%
0.2		0.005	0.100	4.3%	0.005	0.100	4.4%	0.000	0.008	6.4%	0.005	0.100	4.3%
0.1		0.000	0.091	5.6%	0.000	0.091	5.7%	0.000	0.007	5.3%	0.000	0.091	5.6%
0.9		-0.001	0.010	0.9%	-0.002	0.010	4.6%	0.000	0.006	3.9%	-0.001	0.010	5.0%
0.2		0.013	0.335	5.2%	0.018	0.336	6.1%	0.001	0.055	4.1%	0.012	0.335	5.7%
0.2		-0.001	0.102	5.2%	-0.001	0.102	5.4%	0.000	0.008	7.3%	-0.001	0.102	5.2%
0.1		0.002	0.089	5.1%	0.002	0.089	5.1%	0.000	0.007	5.9%	0.002	0.089	5.1%
0.2	$R = 200$	-0.007	0.054	5.4%	-0.009	0.055	5.9%	-0.005	0.034	5.6%	-0.007	0.054	5.9%
0.2		0.013	0.231	4.7%	0.016	0.232	4.7%	0.006	0.040	6.0%	0.013	0.231	4.7%
0.2		-0.002	0.071	4.3%	-0.002	0.071	4.4%	0.000	0.005	5.1%	-0.002	0.071	4.3%
0.1		0.000	0.064	4.7%	0.000	0.064	4.7%	0.000	0.005	5.1%	0.000	0.064	4.7%
0.6		-0.003	0.027	2.8%	-0.004	0.027	5.3%	-0.001	0.017	4.1%	-0.003	0.027	4.9%
0.2		-0.001	0.233	5.2%	0.001	0.233	5.4%	0.002	0.039	4.4%	-0.002	0.233	5.2%
0.2		0.003	0.071	4.7%	0.003	0.071	4.7%	0.000	0.005	5.1%	0.003	0.071	4.7%
0.1		0.000	0.063	5.5%	0.000	0.063	5.5%	0.000	0.005	5.8%	0.000	0.063	5.5%
0.9		-0.001	0.007	1.3%	-0.001	0.007	5.7%	-0.001	0.005	4.9%	-0.001	0.007	5.3%
0.2		0.012	0.232	3.4%	0.015	0.233	4.2%	0.004	0.039	5.0%	0.012	0.232	4.1%
0.2		-0.002	0.072	5.0%	-0.002	0.072	5.1%	0.000	0.006	5.5%	-0.002	0.072	5.0%
0.1		0.002	0.063	5.0%	0.002	0.063	5.0%	0.000	0.005	6.7%	0.002	0.063	5.0%

Notes: The weight matrix is a social interaction matrix $\mathbf{W}_n = \mathbf{I}_R \otimes [(\mathbf{1}_{m_j} \mathbf{1}'_{m_j} - \mathbf{I}_{m_j}) / (m_j - 1)]$, $m_j \sim \text{IID } U(3, 20)$, $j = 1, \dots, R$. Reported for each estimator are the bias, RMSE, and the empirical size of the t test for each parameter equal to its true value at 5% from 1,000 simulations. The exogenous regressors are $(1, x_{i1}, x_{i2})$, $x_{i1} \sim \text{IID } N(3, 1)$, $x_{i2} \sim \text{IID } U(-1, 2)$, and x_{i1} is independent of x_{i2} . The group size m_j follows $\text{IID } U(3, 20)$. The error terms follow a zero-mean independent normal distribution. The variance structure is such that if $m_j > 10$, the error variance in the j -th group is m_j and otherwise is $1/m_j^2$, $j = 1, \dots, R$.

TABLE 3: Estimation of SAR with variance structure V_2 and parameter configuration P_1

θ_0	MQML				GMM				GMM(Σ_n)				II			
	$R = 100$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$
0.2	0.000	0.014	0.014	4.6%	0.000	0.014	5.2%	-0.001	0.013	4.9%	0.000	0.014	5.0%	0.000	0.014	5.0%
0.8	0.001	0.047	0.047	3.6%	0.000	0.047	3.7%	0.003	0.042	3.7%	0.001	0.046	3.7%	0.001	0.046	3.7%
0.2	0.000	0.008	0.008	4.1%	0.000	0.008	4.1%	0.000	0.008	5.1%	0.000	0.008	4.2%	0.000	0.008	4.2%
1.5	0.000	0.008	0.008	5.3%	0.000	0.008	5.4%	0.000	0.007	4.6%	0.000	0.008	5.3%	0.000	0.008	5.3%
0.6	0.000	0.008	0.008	4.3%	0.000	0.008	4.9%	-0.001	0.007	4.3%	0.000	0.008	4.6%	0.000	0.008	4.6%
0.8	0.001	0.049	0.049	4.5%	0.002	0.049	5.7%	0.002	0.043	4.3%	0.001	0.049	5.5%	0.001	0.049	5.5%
0.2	0.000	0.008	0.008	5.3%	0.000	0.008	5.1%	0.000	0.008	4.9%	0.000	0.008	5.2%	0.000	0.008	5.2%
1.5	0.000	0.008	0.008	5.3%	0.000	0.008	5.4%	0.000	0.007	4.7%	0.000	0.008	5.3%	0.000	0.008	5.3%
0.9	0.000	0.002	0.002	5.1%	0.000	0.002	7.8%	0.000	0.002	5.9%	0.000	0.002	6.8%	0.000	0.002	6.8%
0.8	0.002	0.050	0.050	4.1%	0.002	0.050	5.8%	0.002	0.045	4.9%	0.002	0.049	4.6%	0.002	0.049	4.6%
0.2	0.000	0.008	0.008	5.4%	0.000	0.008	5.4%	0.000	0.008	5.6%	0.000	0.008	5.4%	0.000	0.008	5.4%
1.5	0.000	0.008	0.008	5.2%	0.000	0.008	5.2%	0.000	0.007	4.7%	0.000	0.008	5.0%	0.000	0.008	5.0%
0.2	0.000	0.010	0.010	3.8%	0.000	0.010	4.7%	0.000	0.009	3.8%	0.000	0.010	4.4%	0.000	0.010	4.4%
0.8	0.001	0.033	0.033	4.6%	0.001	0.033	4.1%	0.001	0.030	2.6%	0.001	0.033	4.6%	0.001	0.033	4.6%
0.2	0.000	0.006	0.006	4.6%	0.000	0.006	4.6%	0.000	0.005	4.7%	0.000	0.006	4.6%	0.000	0.006	4.6%
1.5	0.000	0.005	0.005	4.9%	0.000	0.005	4.6%	0.000	0.005	5.5%	0.000	0.005	4.9%	0.000	0.005	4.9%
0.6	0.000	0.005	0.005	3.9%	0.000	0.006	5.4%	0.000	0.005	4.5%	0.000	0.005	4.6%	0.000	0.005	4.6%
0.8	0.000	0.034	0.034	3.7%	-0.001	0.034	5.0%	0.000	0.030	3.8%	0.000	0.034	3.9%	0.000	0.034	3.9%
0.2	0.000	0.006	0.006	5.9%	0.000	0.006	6.1%	0.000	0.006	5.1%	0.000	0.006	5.9%	0.000	0.006	5.9%
1.5	0.000	0.006	0.006	5.5%	0.000	0.006	5.4%	0.000	0.005	4.8%	0.000	0.006	5.6%	0.000	0.006	5.6%
0.9	0.000	0.001	0.001	3.9%	0.000	0.001	5.9%	0.000	0.001	5.4%	0.000	0.001	5.0%	0.000	0.001	5.0%
0.8	-0.001	0.035	0.035	4.3%	-0.002	0.035	5.4%	0.000	0.031	5.8%	-0.001	0.034	5.0%	0.000	0.034	5.0%
0.2	0.000	0.006	0.006	3.8%	0.000	0.006	3.9%	0.000	0.005	4.4%	0.000	0.006	3.8%	0.000	0.006	3.8%
1.5	0.000	0.006	0.006	5.1%	0.000	0.006	5.2%	0.000	0.005	4.6%	0.000	0.006	5.5%	0.000	0.006	5.5%

Notes: The weight matrix is a social interaction matrix $\mathbf{W}_n = \mathbf{I}_R \otimes [(\mathbf{1}_{m_j} \mathbf{1}'_{m_j} - \mathbf{I}_{m_j}) / (m_j - 1)]$, $m_j \sim \text{IID } U(3, 20)$, $j = 1, \dots, R$. Reported for each estimator are the bias, RMSE, and the empirical size of the t test for each parameter equal to its true value at 5% from 1,000 simulations. The exogenous regressors are $(1, x_{i1}, x_{i2})$, $x_{i1} \sim \text{IID } N(3, 1)$, $x_{i2} \sim \text{IID } U(-1, 2)$, and x_{i1} is independent of x_{i2} . The group size m_j follows IID $U(3, 20)$. The error terms follow a zero-mean independent normal distribution. The variance structure is such that the error variance in the j -th group is $1/m_j$, $j = 1, \dots, R$.

TABLE 4: Estimation of SAR with variance structure V_2 and parameter configuration P_2

θ_0	$R = 100$	MQML			GMM			GMM(Σ_n)			II		
		Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$
0.2		-0.005	0.058	5.2%	-0.004	0.080	19.0%	-0.005	0.051	6.4%	-0.005	0.058	5.9%
0.2		0.006	0.066	5.0%	0.006	0.086	15.4%	0.006	0.058	6.4%	0.006	0.066	6.7%
0.2		0.000	0.008	4.8%	0.000	0.008	4.9%	0.000	0.008	5.2%	0.000	0.008	4.8%
0.1		0.000	0.008	4.6%	0.000	0.008	4.6%	0.000	0.007	4.8%	0.000	0.008	4.6%
0.6		-0.003	0.029	1.4%	-0.001	0.042	19.7%	-0.003	0.027	6.5%	-0.003	0.029	5.5%
0.2		0.007	0.067	1.5%	0.003	0.091	17.9%	0.007	0.061	6.2%	0.006	0.067	5.2%
0.2		0.000	0.008	3.7%	0.000	0.008	3.6%	0.000	0.008	4.2%	0.000	0.008	3.8%
0.1		0.000	0.007	4.4%	0.000	0.007	4.0%	0.000	0.007	5.6%	0.000	0.007	4.4%
0.9		-0.001	0.008	0.4%	-0.001	0.011	19.2%	-0.001	0.007	5.4%	-0.001	0.007	5.8%
0.2		0.009	0.067	0.4%	0.008	0.093	17.8%	0.009	0.059	4.4%	0.007	0.065	5.0%
0.2		0.000	0.009	4.8%	0.000	0.009	5.7%	0.000	0.008	6.1%	0.000	0.009	5.1%
0.1		0.000	0.008	5.5%	0.000	0.008	5.7%	0.000	0.007	5.6%	0.000	0.008	5.6%
0.2	$R = 200$	-0.003	0.039	5.4%	-0.004	0.058	20.1%	-0.004	0.035	5.9%	-0.003	0.039	6.0%
0.2		0.005	0.045	4.9%	0.006	0.064	17.1%	0.005	0.041	6.4%	0.004	0.045	5.7%
0.2		-0.001	0.006	5.3%	-0.001	0.006	5.6%	0.000	0.006	5.4%	-0.001	0.006	5.3%
0.1		0.000	0.005	5.8%	0.000	0.005	5.7%	0.000	0.005	5.3%	0.000	0.005	5.8%
0.6		-0.002	0.020	0.7%	-0.001	0.029	16.8%	-0.003	0.018	4.8%	-0.001	0.020	4.6%
0.2		0.003	0.045	1.1%	0.002	0.062	15.6%	0.005	0.040	4.3%	0.002	0.044	4.5%
0.2		0.000	0.006	5.0%	0.000	0.006	5.0%	0.000	0.005	4.9%	0.000	0.006	5.0%
0.1		0.000	0.005	5.3%	0.000	0.005	5.2%	0.000	0.005	5.2%	0.000	0.005	5.1%
0.9		0.000	0.005	0.1%	0.000	0.008	17.3%	0.000	0.004	5.4%	0.000	0.005	4.7%
0.2		0.002	0.045	0.2%	0.002	0.064	15.9%	0.003	0.039	3.7%	0.002	0.044	3.9%
0.2		0.000	0.006	4.4%	0.000	0.006	5.3%	0.000	0.006	4.6%	0.000	0.006	4.7%
0.1		0.000	0.005	5.8%	0.000	0.005	5.8%	0.000	0.005	5.1%	0.000	0.005	5.9%

Notes: The weight matrix is a social interaction matrix $\mathbf{W}_n = \mathbf{I}_R \otimes [(\mathbf{1}_{m_j} \mathbf{1}'_{m_j} - \mathbf{I}_{m_j}) / (m_j - 1)]$, $m_j \sim \text{IID } U(3, 20)$, $j = 1, \dots, R$. Reported for each estimator are the bias, RMSE, and the empirical size of the t test for each parameter equal to its true value at 5% from 1,000 simulations. The exogenous regressors are $(1, x_{i1}, x_{i2})$, $x_{i1} \sim \text{IID } N(3, 1)$, $x_{i2} \sim \text{IID } U(-1, 2)$, and x_{i1} is independent of x_{i2} . The group size m_j follows IID $U(3, 20)$. The error terms follow a zero-mean independent normal distribution. The variance structure is such that the error variance in the j -th group is $1/m_j$, $j = 1, \dots, R$.

Some interesting observations arise. Firstly, all the consistent estimators deliver almost unbiased results across all the experimental configurations, though the estimated intercept term associated with \mathbf{X}_n is relatively more biased. Secondly, among the consistent estimators, the optimal robust GMM using the true $\boldsymbol{\Sigma}_n$ usually achieves the smallest RMSE. The other three estimators have very similar performance in terms of RMSE. Thirdly, for the purpose of hypothesis testing, it appears that the II-based procedure is as good as the one based on (the infeasible) GMM($\boldsymbol{\Sigma}_n$), with the empirical rejection rates matching very closely the nominal size. The MQML of Liu and Yang (2015) tends to deliver under-sized t -test regarding the spatial autoregression parameter λ when its value is relatively high. For example, in Table 4, one sees the rejection rates of 0.4% and 0.1%, under $R = 100$ and $R = 200$, respectively, for testing λ equal to its true value when $\lambda_0 = 0.9$ from a 5% t -test based on Liu and Yang (2015). This under-size problem, when the degree of spatial correlation is high, also carries over to the t test associated with the intercept parameter. The GMM estimator, when one is unsure of the error variance structure and uses $\mathbf{u}'_n(\mathbf{G}_n - \text{Dg}(\mathbf{G}_n)\mathbf{u}_n$ and $\mathbf{u}'_n(\mathbf{G}_n\mathbf{X}_n\boldsymbol{\beta}, \mathbf{X}_n)$ as the moment conditions, delivers very disappointing size performance in Table 4 when testing either the spatial autoregression parameter λ or the intercept parameter: the rejection rates approach around 20% at the 5% nominal size.

Given the simulated data, it is worthwhile to look at a plot of the binding function to check whether the binding function is monotonic, as required by Assumption 7. Figure 1 is drawn for 1,000 simulated data sets under variance structure V1 and parameter configuration P1 with $R = 100$.⁸ Recall that for a given $\boldsymbol{\theta}_0 = (\lambda_0, \boldsymbol{\beta}'_0)'$ and the exogenous \mathbf{X}_n , the data generating process generates the observable data \mathbf{y}_n and the binding function $\lambda + \mathbf{y}'_n\mathbf{S}'_n(\lambda)\mathbf{M}_n\mathbf{D}_n(\lambda)\mathbf{M}_n\mathbf{S}_n(\lambda)\mathbf{y}_n/\mathbf{y}'_n\mathbf{W}'_n\mathbf{M}_n\mathbf{W}_n\mathbf{y}_n$ is a function of λ . Figure 1 clearly illustrates that the binding function is a monotonic function of λ and thus the monotonicity condition in Assumption 7 is numerically valid. Figures drawn for the simulated data under other configurations display similar patterns and are omitted.

One may wonder about the performance of the proposed II estimator under homoscedasticity, relative to the QML estimator of Lee (2004) and the best GMM estimator of Lee (2007).⁹ Tables 5–6 report the Monte Carlo results under parameter configurations P1 and P2, but now the error term is simulated as a standard normal random variable. The exogenous variables are simulated

⁸Each sub-figure contains 1,000 lines, one for each of the simulated data set.

⁹The authors thank a referee for suggesting this comparison. Since one no longer needs to concentrate out the nuisance matrix $\boldsymbol{\Sigma}_n$, the II procedure needs to be modified, see the appendix.

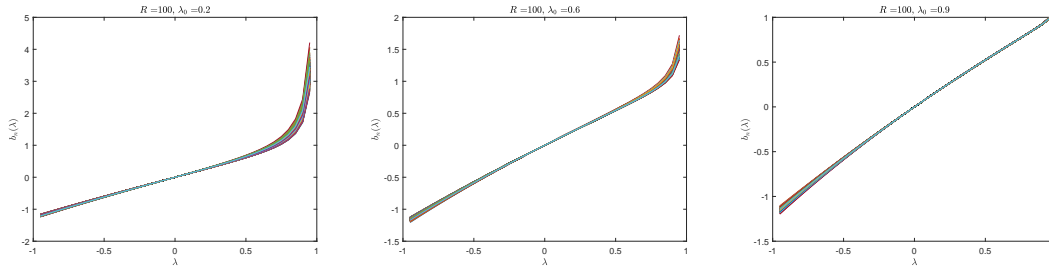


FIGURE 1: $b_n(\lambda)$ under variance structure V1 and parameter configuration P1, $R = 100$

the same as before. From Tables 5–6, one observes that the II estimator is slightly better than the best GMM estimator of Lee (2007), usually delivering smaller finite-sample bias and lower RMSE. Both methods have good finite-sample size performance in terms of the 5% t test. The finite-sample performance of the QML, on the other hand, is quite different from what the asymptotic theory predicts. Its bias is more severe than the other two and it also gives higher RMSE. Moreover, its size performance is very poor.

The simulation results suggest that the II estimator could be used at least as a complement to other consistent estimators proposed in the literature that are robust to unknown heteroscedasticity. The II method may be favored for the purpose of hypothesis testing, especially for testing the spatial autoregression parameter. It also has very good finite-sample performance under homoscedasticity.

5 Concluding Remarks

Lee (2002) challenged the traditional wisdom that the OLS estimator is biased and inconsistent in spatial autoregressions and showed that it may be consistent under some special circumstances if there are exogenous regressors included. This paper examines thoroughly the asymptotic behavior of the OLS estimator under different specifications of the degree of aggregate influence on each unit from other units and provides a unified asymptotic distribution result of the re-centered OLS estimator. Based on this, an indirect inference estimator, which is consistent and asymptotically normal, is introduced. The new estimator is relatively easy to calculate, does not rely on distributional assumptions on the data, and is robust to heteroscedasticity. Monte Carlo experiments in this paper show the good finite-sample performance of the II estimator in comparison with other consistent estimators that are robust to unknown heteroscedasticity.

TABLE 5: Estimation of SAR under homoscedasticity with parameter configuration P1

θ_0	$R = 100$	QML			GMM			II		
		Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$
0.2		0.014	0.043	19.0%	-0.007	0.040	7.5%	-0.003	0.037	5.6%
0.8		-0.032	0.146	18.5%	0.025	0.144	6.7%	0.013	0.138	6.3%
0.2		-0.002	0.029	15.9%	-0.002	0.029	6.1%	-0.002	0.029	6.1%
1.5		-0.001	0.026	17.7%	-0.002	0.026	6.1%	-0.001	0.026	5.9%
0.6		0.026	0.032	34.1%	-0.004	0.020	5.4%	-0.002	0.019	5.2%
0.8		-0.137	0.188	26.5%	0.022	0.137	5.4%	0.010	0.131	4.7%
0.2		-0.001	0.029	17.3%	-0.001	0.029	5.1%	0.000	0.029	5.1%
1.5		-0.005	0.026	15.3%	0.000	0.026	4.5%	0.001	0.026	4.4%
0.9		0.011	0.011	51.3%	-0.001	0.005	5.2%	0.000	0.005	4.1%
0.8		-0.213	0.248	39.9%	0.022	0.140	5.5%	0.009	0.134	5.0%
0.2		-0.002	0.028	14.5%	-0.001	0.028	4.3%	-0.001	0.028	3.9%
1.5		-0.011	0.029	19.8%	0.001	0.027	5.1%	0.001	0.027	4.7%
0.2	$R = 200$	0.017	0.032	23.0%	-0.003	0.026	5.6%	-0.001	0.025	4.3%
0.8		-0.044	0.103	18.6%	0.009	0.094	5.0%	0.003	0.090	3.6%
0.2		-0.001	0.020	15.7%	-0.001	0.020	4.1%	-0.001	0.020	4.2%
1.5		0.001	0.017	14.1%	0.001	0.017	4.6%	0.001	0.017	4.9%
0.6		0.028	0.030	62.1%	-0.002	0.014	5.3%	-0.001	0.013	4.2%
0.8		-0.145	0.172	47.1%	0.009	0.096	4.8%	0.003	0.093	3.8%
0.2		0.000	0.021	18.4%	0.000	0.021	4.5%	0.000	0.021	4.5%
1.5		-0.005	0.019	18.1%	0.000	0.018	5.3%	0.001	0.018	4.8%
0.9		0.011	0.011	85.1%	0.000	0.004	6.3%	0.000	0.003	4.1%
0.8		-0.220	0.237	68.7%	0.009	0.097	5.9%	0.004	0.093	5.0%
0.2		-0.002	0.021	17.7%	0.000	0.021	4.5%	0.000	0.021	4.4%
1.5		-0.014	0.023	25.0%	-0.002	0.019	4.6%	-0.001	0.019	4.5%

Notes: The weight matrix is a social interaction matrix $\mathbf{W}_n = \mathbf{I}_R \otimes [(\mathbf{1}_{m_j} \mathbf{1}'_{m_j} - \mathbf{I}_{m_j}) / (m_j - 1)]$, $m_j \sim \text{IID } U(3, 20)$, $j = 1, \dots, R$. Reported for each estimator are the bias, RMSE, and the empirical size of the t test for each parameter equal to its true value at 5% from 1,000 simulations. The exogenous regressors are $(1, x_{i1}, x_{i2})'$, $x_{i1} \sim \text{IID } N(3, 1)$, $x_{i2} \sim \text{IID } U(-1, 2)$, and x_{i1} is independent of x_{i2} . The group size m_j follows $\text{IID } U(3, 20)$. The error terms follow a standard normal distribution.

TABLE 6: Estimation of SAR under homoscedasticity with parameter conguration P2

θ_0	$R = 100$	QML			GMM			II		
		Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$
0.2		0.041	0.078	13.3%	-0.007	0.056	5.7%	-0.004	0.055	5.6%
0.2		-0.037	0.121	14.7%	0.017	0.112	5.3%	0.011	0.109	5.1%
0.2		-0.002	0.029	16.2%	-0.004	0.029	5.9%	-0.002	0.029	6.0%
0.1		0.000	0.026	18.0%	-0.001	0.026	6.6%	0.000	0.026	6.4%
0.6		0.072	0.076	29.7%	-0.004	0.029	4.6%	-0.002	0.028	4.3%
0.2		-0.149	0.181	25.9%	0.007	0.108	5.3%	0.001	0.107	5.6%
0.2		-0.001	0.029	16.4%	-0.001	0.029	5.2%	0.001	0.029	5.2%
0.1		0.000	0.026	18.2%	0.000	0.027	6.6%	0.001	0.027	6.2%
0.9		0.026	0.027	48.5%	-0.001	0.008	5.3%	-0.001	0.008	5.3%
0.2		-0.214	0.235	38.4%	0.010	0.110	5.3%	0.004	0.109	4.9%
0.2		-0.003	0.028	15.8%	0.000	0.028	4.0%	0.001	0.028	4.2%
0.1		-0.002	0.025	15.2%	0.000	0.026	5.1%	0.001	0.026	4.9%
0.2	$R = 200$	0.046	0.064	23.3%	-0.004	0.038	4.6%	-0.003	0.038	4.5%
0.2		-0.046	0.093	18.7%	0.009	0.079	5.3%	0.005	0.078	4.7%
0.2		-0.001	0.021	17.5%	-0.002	0.021	6.0%	-0.001	0.021	5.6%
0.1		0.000	0.018	15.8%	0.000	0.018	4.3%	0.000	0.018	4.0%
0.6		0.074	0.076	70.7%	-0.002	0.020	4.8%	-0.001	0.020	4.6%
0.2		-0.152	0.170	49.3%	0.004	0.081	6.3%	0.000	0.079	5.9%
0.2		-0.002	0.021	17.6%	-0.001	0.021	5.0%	0.001	0.021	5.0%
0.1		0.000	0.019	17.8%	0.000	0.019	6.1%	0.001	0.019	6.4%
0.9		0.027	0.027	95.2%	-0.001	0.005	4.6%	0.000	0.005	4.1%
0.2		-0.218	0.229	72.6%	0.005	0.081	6.0%	0.001	0.079	5.2%
0.2		-0.003	0.020	15.3%	0.000	0.020	4.8%	0.001	0.020	4.8%
0.1		-0.001	0.018	15.0%	0.000	0.018	4.2%	0.001	0.018	3.9%

Notes: The weight matrix is a social interaction matrix $\mathbf{W}_n = \mathbf{I}_R \otimes [(\mathbf{1}_{m_j} \mathbf{1}'_{m_j} - \mathbf{I}_{m_j}) / (m_j - 1)]$, $m_j \sim \text{IID } U(3, 20)$, $j = 1, \dots, R$. Reported for each estimator are the bias, RMSE, and the empirical size of the t test for each parameter equal to its true value at 5% from 1,000 simulations. The exogenous regressors are $(1, x_{i1}, x_{i2})'$, $x_{i1} \sim \text{IID } N(3, 1)$, $x_{i2} \sim \text{IID } U(-1, 2)$, and x_{i1} is independent of x_{i2} . The group size m_j follows $\text{IID } U(3, 20)$. The error terms follow a standard normal distribution.

In this paper, no attempt is made to conduct some comparison of the asymptotic variances of the GMM and II estimators. The II estimator in this paper may be interpreted as an estimator that uses one moment condition, namely, by matching the OLS estimator $\hat{\lambda}_n$ with its approximate analytical expectation. In contrast, the GMM estimator in Lin and Lee (2010) is based on a set of exact expectations of bilinear and quadratic forms in \mathbf{u}_n . The OLS estimator itself is based on an incorrect moment condition, namely, exogeneity of $\mathbf{W}_n \mathbf{y}_n$. It is not clear whether correcting an incorrect moment condition is as efficient as using a set of correct moment conditions. A fruitful strategy is perhaps to design a combined estimator.¹⁰ Another possible extension is to consider the more general higher-order SARAR (spatial autoregressive model with spatial autoregressive disturbances) with heteroscedastic innovations as in Badinger and Egger (2011) and Jin and Lee (2019). In this more general setup, the II procedure may be implemented as follows. One can first derive the approximate analytical expectation of the OLS estimator of the parameters in the SAR part, taken the parameters in the disturbance part as given, and thus design a “corrected” SAR estimator. Then based on the residuals that arise from the “corrected” SAR estimator, one can derive approximate analytical expectation of the OLS estimator of the parameters in the disturbance part. In the end, one can jointly estimate all the parameters in the SAR and disturbance parts by using the two sets of approximate analytical expectations. Some preliminary simulations show very promising results of this approach. Rigorous treatments of these extensions are left for future studies.

Appendix

The appendix first collects several lemmas that are useful for deriving the main results. Some of these results (without proofs) were either derived or presented in different ways, see Kelejian and Prucha (1999, 2001) and Lee (2001, 2002, 2004). The proofs of Theorems 1–4 in Section 2 and Corollary 2 in Section 3 are provided and those of Corollaries 1 and 3, which follow similarly, are skipped.

Lemma 1. *If $\{\mathbf{A}_n\}$ and $\{\mathbf{B}_n\}$ are uniformly bounded in row and column sums, then so are $\{\mathbf{A}_n + \mathbf{B}_n\}$ and $\{\mathbf{A}_n \mathbf{B}_n\}$.*

¹⁰Very recently, Zhang and Yu (2018) and Lam and Souza (2019) proposed combining spatial weight matrices in recognition of possible misspecification of the weight matrix and Chen et al. (2019) suggested combining a conservative GMM estimator based on valid moment conditions and an aggressive GMM estimator based on both valid and possibly misspecified moment conditions.

Lemma 2. Suppose $\{\mathbf{A}_n\}$ has its elements of order $O(h_n^{-1})$. If $\{\mathbf{B}_n\}$ is uniformly bounded in column sum, then the elements of $\mathbf{A}_n\mathbf{B}_n$ are $O(h_n^{-1})$; if $\{\mathbf{B}_n\}$ is uniformly bounded in row sum, then the elements of $\mathbf{B}_n\mathbf{A}_n$ are $O(h_n^{-1})$. In either case, $\text{tr}(\mathbf{A}_n\mathbf{B}_n) = O(n/h_n)$.

Lemma 3. For a product involving (powers of) \mathbf{G}_n and \mathbf{G}'_n , $\prod_{l=1}^m (\mathbf{G}_n^{i_1} \mathbf{G}_n'^{i_2})^{j_l}$, where $i_1, i_2 \geq 0$, $i_1 i_2 > 0$, $j_l \geq 0$, $\prod_{l=1}^m j_l > 0$, i_1, i_2, j_l all being integers, under Assumptions 1 and 2, its elements are of order $O(h_n^{-1})$ and its trace is of order $O(n/h_n)$.

Proof. Under Assumptions 1 and 2, from Lemma 2, the elements of $\mathbf{G}_n = \mathbf{W}_n \mathbf{S}_n^{-1}$ are $O(h_n^{-1})$ and $\text{tr}(\mathbf{G}_n) = O(n/h_n)$. From Lemma 1, $\mathbf{G}_n = \mathbf{W}_n \mathbf{S}_n^{-1}$ is uniformly bounded in row and column sums. By successive using of Lemma 1, $\mathbf{G}_n^{i_1-1}$ is uniformly bounded in row and column sums, and through Lemma 2, the elements of $\mathbf{G}_n^{i_1} = \mathbf{G}_n \mathbf{G}_n^{i_1-1}$ are $O(h_n^{-1})$ and $\text{tr}(\mathbf{G}_n^{i_1}) = O(n/h_n)$. Similarly, such a claim applies to $\mathbf{G}_n'^{i_2}$, which is also uniformly bounded in row and column sums. Then $\mathbf{G}_n^{i_1} \mathbf{G}_n'^{i_2}$ is uniformly bounded in row and column sums with its elements being $O(h_n^{-1})$ and $\text{tr}(\mathbf{G}_n^{i_1} \mathbf{G}_n'^{i_2}) = O(n/h_n)$. Proceeding similarly, one can see that the product $\prod_{l=1}^m (\mathbf{G}_n^{i_1} \mathbf{G}_n'^{i_2})^{j_l}$ shares these properties too. \square

Lemma 4. For the sequence $\{\mathbf{u}_n\}$ with the elements following Assumption 3, let \mathbf{A}_n and \mathbf{B}_n be nonrandom, then

$$\mathbb{E}(\mathbf{u}'_n \mathbf{A}_n \mathbf{u}_n) = \text{tr}(\boldsymbol{\Sigma}_n \mathbf{A}_n), \quad (\text{A.1})$$

$$\mathbb{E}(\mathbf{u}_n \mathbf{u}'_n \mathbf{A}_n \mathbf{u}_n) = \text{dg}(\boldsymbol{\Sigma}_n^{(3)} \odot \mathbf{A}_n), \quad (\text{A.2})$$

$$\begin{aligned} \mathbb{E}(\mathbf{u}'_n \mathbf{A}_n \mathbf{u}_n \mathbf{u}'_n \mathbf{B}_n \mathbf{u}_n) &= \text{tr}(\boldsymbol{\Sigma}_n^{(4)} \odot \mathbf{A}_n \odot \mathbf{B}_n) + \text{tr}(\boldsymbol{\Sigma}_n \mathbf{A}_n) \text{tr}(\boldsymbol{\Sigma}_n \mathbf{B}_n) \\ &\quad + \text{tr}[\boldsymbol{\Sigma}_n \mathbf{A}_n \boldsymbol{\Sigma}_n (\mathbf{B}_n + \mathbf{B}'_n)]. \end{aligned} \quad (\text{A.3})$$

Lemma 5. Under Assumptions 1–5,

$$\begin{aligned} \mathbf{u}'_n \mathbf{G}_n \mathbf{u}_n &= O_P(n/h_n), \quad \mathbf{u}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n = O_P(n/h_n), \\ \mathbf{u}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n &= O_P(n/h_n), \quad \boldsymbol{\beta}'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{u}_n = O_P(\sqrt{n}), \\ \boldsymbol{\beta}'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n &= O_P(\sqrt{n}), \quad \boldsymbol{\beta}'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \boldsymbol{\beta}_0 = O(n). \end{aligned}$$

Proof. From (A.1), $\mathbb{E}(\mathbf{u}'_n \mathbf{G}_n \mathbf{u}_n) = \text{tr}(\boldsymbol{\Sigma}_n \mathbf{G}_n)$. Under Assumption 3, $\boldsymbol{\Sigma}_n$ is uniformly bounded, so $\text{tr}(\boldsymbol{\Sigma}_n \mathbf{G}_n) \leq K \text{tr}(\mathbf{G}_n) = O(n/h_n)$ from Lemma 3. Lee (2004) shows that, under Assumption 5, $\lim_{n \rightarrow \infty} n^{-1} (\mathbf{X}_n, \mathbf{G}_n \mathbf{X}_n \boldsymbol{\beta}_0)' (\mathbf{X}_n, \mathbf{G}_n \mathbf{X}_n \boldsymbol{\beta}_0)$ is nonsingular if and only if both the limits of

$n^{-1}\mathbf{X}'_n\mathbf{X}_n$ and $n^{-1}\beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{X}_n\beta_0$ are nonsingular, indicating that $\mathbf{X}'_n\mathbf{X}_n = O(n)$ and $\beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{X}_n\beta_0 = O(n)$. Also from Lee (2004), \mathbf{M}_n is uniformly bounded in row and column sums, and then $E(\mathbf{u}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{u}_n) = \text{tr}(\boldsymbol{\Sigma}_n\mathbf{M}_n\mathbf{G}_n) \leq K\text{tr}(\mathbf{M}_n\mathbf{G}_n) = O(n/h_n)$ from Lemmas 2 and 3. Similarly, one can show $E(\mathbf{u}'_n\mathbf{G}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{u}_n) = O(n/h_n)$. As for $\beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\mathbf{u}_n$, its expectation is zero and its variance is given by $\beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\boldsymbol{\Sigma}_n\mathbf{M}_n\mathbf{G}_n\mathbf{X}_n\beta_0$, which is bounded by $K\beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{X}_n\beta_0 = O(n)$. Then it follows that $\beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\mathbf{u}_n = O_P(\sqrt{n})$. Similarly, $\beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{u}_n = O_P(\sqrt{n})$. Note that $\mathbf{G}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{G}'_n\mathbf{M}_n\mathbf{G}_n$ is uniformly bounded in row and column sums through Lemmas 1 and 2. Given that the elements of \mathbf{X}_n are uniformly bounded, one has $\beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{G}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{X}_n\beta_0 = O(n)$. Then it follows that $\beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{u}_n = O_P(\sqrt{n})$. \square

Lemma 6. *Under Assumptions 1–5, $E(r_n) = O(n/h_n)$, $\text{Var}(r_n) = O(n)$, $E(d_n) = O(n)$, and $\text{Var}(d_n) = O(n)$. When there is no \mathbf{X} , then $E(r_{1n}) = O(n/h_n)$, $\text{Var}(r_{1n}) = O(n/h_n)$, $E(d_{1n}) = O(n/h_n)$, and $\text{Var}(d_{1n}) = O(n/h_n)$.*

Proof. Given Lemmas 4 and 5,

$$E(r_n) = \text{tr}(\boldsymbol{\Sigma}_n\mathbf{M}_n\mathbf{G}_n) = O(n/h_n) \quad (\text{A.4})$$

and

$$E(d_n) = \text{tr}(\boldsymbol{\Sigma}_n\mathbf{G}'_n\mathbf{M}_n\mathbf{G}_n) + \beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{X}_n\beta_0 = O(n) \quad (\text{A.5})$$

are obvious. Using Lemma 4,

$$\begin{aligned} \text{Var}(r_n) &= \text{Var}(\mathbf{u}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{u}_n) + \text{Var}(\beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\mathbf{u}_n) \\ &\quad + 2\text{Cov}(\mathbf{u}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{u}_n, \beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\mathbf{u}_n) \\ &= E[(\mathbf{u}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{u}_n)^2] - [E(\mathbf{u}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{u}_n)]^2 + E(\mathbf{u}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{X}_n\beta_0\beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\mathbf{u}_n) \\ &\quad + 2\beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_nE(\mathbf{u}_n\mathbf{u}'_n\mathbf{M}_n\mathbf{G}_n\mathbf{u}_n) \\ &= \text{tr}(\boldsymbol{\Sigma}_n^{(4)} \odot \mathbf{M}_n\mathbf{G}_n \odot \mathbf{M}_n\mathbf{G}_n) + \text{tr}[\boldsymbol{\Sigma}_n\mathbf{M}_n\mathbf{G}_n\boldsymbol{\Sigma}_n(\mathbf{M}_n\mathbf{G}_n + \mathbf{G}'_n\mathbf{M}_n)] \\ &\quad + \beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\boldsymbol{\Sigma}_n\mathbf{M}_n\mathbf{G}_n\mathbf{X}_n\beta_0 + 2\beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\text{dg}(\boldsymbol{\Sigma}_n^{(3)} \odot \mathbf{M}_n\mathbf{G}_n), \end{aligned} \quad (\text{A.6})$$

where in view of Lemma 3 and 5, $\beta'_0\mathbf{X}'_n\mathbf{G}'_n\mathbf{M}_n\boldsymbol{\Sigma}_n\mathbf{M}_n\mathbf{G}_n\mathbf{X}_n\beta_0 = O(n)$ is the leading term.

Similarly,

$$\begin{aligned}
\text{Var}(d_n) &= \text{E}[(\mathbf{u}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n)^2] - [\text{E}(\mathbf{u}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n)]^2 \\
&\quad + 4(\mathbf{u}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \boldsymbol{\beta}_0 \boldsymbol{\beta}'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n) \\
&\quad + 4\boldsymbol{\beta}'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \text{E}(\mathbf{u}_n \mathbf{u}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n) \\
&= \text{tr}(\boldsymbol{\Sigma}_n^{(4)} \odot \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \odot \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n) + 2\text{tr}[\boldsymbol{\Sigma}_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \boldsymbol{\Sigma}_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n] \\
&\quad + 4\boldsymbol{\beta}'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \boldsymbol{\Sigma}_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \boldsymbol{\beta}_0 \\
&\quad + 4\boldsymbol{\beta}'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \text{dg}(\boldsymbol{\Sigma}_n^{(3)} \odot \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n), \tag{A.7}
\end{aligned}$$

in which $\boldsymbol{\beta}'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \boldsymbol{\Sigma}_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \boldsymbol{\beta}_0 = O(n)$ is the leading term. For the case when there is no \mathbf{X} , the results are obvious. \square

Lemma 7. *Suppose $\{\mathbf{A}_n\}$ is a sequence of matrices with uniformly bounded row and column sums. Let $\{\mathbf{b}_n\}$ be a sequence of constants with uniformly bounded elements and $\sup n^{-1} \sum_{i=1}^n |b_{i,n}|^{2+\eta_1} < \infty$ for some $\eta_1 > 0$. For the sequence $\{\mathbf{u}_n\}$ that satisfies Assumption 3, let $Q_n = \mathbf{b}'_n \mathbf{u}_n + \mathbf{u}'_n \mathbf{A}_n \mathbf{u}_n$. Then*

$$\frac{Q_n - \text{E}(Q_n)}{\sqrt{\text{Var}(Q_n)}} \xrightarrow{d} \text{N}(0, 1). \tag{A.8}$$

Proof of Theorem 1

By a Nagar-type (Nagar, 1959) expansion,

$$\begin{aligned}
\sqrt{n} \left(\hat{\lambda}_n - \lambda_0 - \frac{\text{tr}(\boldsymbol{\Sigma}_n \mathbf{M}_n \mathbf{G}_n)}{\mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n} \right) &= \sqrt{n} \left(\frac{r_{1n} - \text{E}(r_{1n}) + r_{2n}}{d_n} \right) \\
&= \sqrt{n} \left(\frac{r_{1n} - \text{E}(r_{1n}) + r_{2n}}{\text{E}(d_n) + d_n - \text{E}(d_n)} \right) \\
&= \sqrt{n} \left(\frac{r_{1n} - \text{E}(r_{1n}) + r_{2n}}{\text{E}(d_n)} \right) \left(1 + \frac{d_n - \text{E}(d_n)}{\text{E}(d_n)} \right)^{-1} \\
&= \sqrt{n} \left(\frac{r_{1n} - \text{E}(r_{1n}) + r_{2n}}{\text{E}(d_n)} \right) + o_P(1)
\end{aligned}$$

where, in light of the proof of Lemma 6, $r_{1n} - \text{E}(r_{1n}) = O_P(\sqrt{n/h_n})$, $r_{2n} = O(\sqrt{n})$, and $\text{E}(d_n) = O(n)$. So when h_n diverges,

$$\sqrt{n} \left(\hat{\lambda}_n - \lambda_0 - \frac{\text{tr}(\boldsymbol{\Sigma}_n \mathbf{M}_n \mathbf{G}_n)}{\mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n} \right) = \frac{\sqrt{n} r_{2n}}{\text{E}(d_n)} + o_P(1),$$

and one can apply Lemma 7 to $r_{2n} = \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{u}_n$; when h_n is bounded, one can apply Lemma 7 to $r_n = r_{1n} + r_{2n} = \mathbf{u}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{u}_n + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{u}_n$. From Lemma 6, one sees that $\text{Var}(r_{2n}) = \text{Var}(r_n) + o(n)$ when h_n diverges. Therefore, regardless of h_n ,

$$\sqrt{n} \left(\hat{\lambda}_n - \lambda_0 - \frac{\text{tr}(\boldsymbol{\Sigma}_n \mathbf{M}_n \mathbf{G}_n)}{\mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n} \right) \xrightarrow{d} \text{N} \left(0, \lim_{n \rightarrow \infty} \frac{n \text{Var}(r_n)}{[\text{E}(d_n)]^2} \right).$$

□

Proof of Theorem 2

Note that

$$\begin{aligned} \tilde{\mathbf{u}}'_n \mathbf{D}_n \tilde{\mathbf{u}}_n &= \mathbf{u}'_n \mathbf{D}_n \mathbf{u}_n + (\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)' \mathbf{X}'_n \mathbf{X}_n (\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) - 2(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)' \mathbf{X}'_n \mathbf{D}_n \mathbf{u}_n \\ &= \mathbf{u}'_n \mathbf{D}_n \mathbf{u}_n + O_P(1) + O_P(n^{-1/2}) O_P(\sqrt{n/h_n^2}), \end{aligned}$$

in view of $\mathbf{X}'_n \mathbf{X}_n = O(n)$, $\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0 = O_P(n^{-1/2})$ and $\text{Var}(\mathbf{X}'_n \mathbf{D}_n \mathbf{u}_n) = \mathbf{X}'_n \mathbf{D}_n \boldsymbol{\Sigma}_n \mathbf{D}_n \mathbf{X}_n = O(n/h_n^2)$. Then

$$\begin{aligned} &\sqrt{n} \left(\hat{\lambda}_n - \lambda_0 - \frac{\mathbf{y}'_n \mathbf{S}'_n \mathbf{M}_n \mathbf{D}_n \mathbf{M}_n \mathbf{S}_n \mathbf{y}_n}{\mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n} \right) \\ &= \sqrt{n} \left(\hat{\lambda}_n - \lambda_0 - \frac{\mathbf{u}'_n \mathbf{D}_n \mathbf{u}_n}{d_n} + \frac{\mathbf{u}'_n \mathbf{D}_n \mathbf{u}_n - \tilde{\mathbf{u}}'_n \mathbf{D}_n \tilde{\mathbf{u}}_n}{d_n} \right) \\ &= \sqrt{n} \left(\frac{r_n - \mathbf{u}'_n \mathbf{D}_n \mathbf{u}_n}{d_n} \right) + o_P(1) \\ &= \sqrt{n} \left(\frac{r_n - \mathbf{u}'_n \mathbf{D}_n \mathbf{u}_n}{\text{E}(d_n)} \right) + o_P(1). \end{aligned}$$

It follows from Lemma 7 that

$$\sqrt{n} \left(\hat{\lambda}_n - \lambda_0 - \frac{\mathbf{y}'_n \mathbf{S}'_n \mathbf{M}_n \mathbf{D}_n \mathbf{M}_n \mathbf{S}_n \mathbf{y}_n}{\mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n} \right) \xrightarrow{d} \text{N}(0, \eta),$$

where

$$\begin{aligned} \eta &= \lim_{n \rightarrow \infty} n \frac{\text{Var}(r_n - \mathbf{u}'_n \mathbf{D}_n \mathbf{u}_n)}{[\text{E}(d_n)]^2} \\ &= \lim_{n \rightarrow \infty} n \frac{\text{Var}(\mathbf{u}'_n \mathbf{E}_n \mathbf{u}_n + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{u}_n)}{[\text{E}(d_n)]^2} \\ &= \lim_{n \rightarrow \infty} \frac{n \{ \text{tr}[\boldsymbol{\Sigma}_n \mathbf{E}_n \boldsymbol{\Sigma}_n (\mathbf{E}_n + \mathbf{E}'_n)] + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \boldsymbol{\Sigma}_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0 \}}{[\text{E}(d_n)]^2} \end{aligned} \quad (\text{A.9})$$

by using Lemma 4 and the fact that $\text{dg}(\mathbf{E}_n) = \mathbf{0}$. \square

Proof of Theorem 3

One can apply the extended delta method as in Phillips (2012) to derive the asymptotic distribution of $\sqrt{n}(\hat{\lambda}_n^{II} - \lambda_0)$. For this purpose, one needs to check a technical condition, namely, the $\partial b_n^{-1}(\lambda)/\partial \lambda$ should be asymptotically locally equicontinuous at λ_0 : for given $\zeta > 0$, if $s_n \rightarrow \infty$ and $s_n/\sqrt{n} \rightarrow 0$,

$$\sup_{s_n|\lambda - \lambda_0| < \zeta} \left| \frac{b_n^{-1'}(\lambda) - b_n^{-1'}(\lambda_0)}{b_n^{-1'}(\lambda_0)} \right| = \sup_{s_n|\lambda - \lambda_0| < \zeta} \left| \frac{b_n'(\lambda_0) - b_n'(\lambda)}{b_n'(\lambda)} \right| \xrightarrow{a.s.} 0.$$

The first derivative of the binding function is

$$\begin{aligned} & b_n'(\lambda) \\ &= 1 + \frac{\mathbf{y}_n' \mathbf{S}_n'(\lambda) \mathbf{M}_n \text{Dg}(\mathbf{M}_n \mathbf{G}_n^2(\lambda)) \mathbf{M}_n \mathbf{S}_n(\lambda) \mathbf{y}_n - 2 \mathbf{y}_n' \mathbf{W}_n' \mathbf{M}_n \text{Dg}(\mathbf{M}_n \mathbf{G}_n(\lambda)) \mathbf{M}_n \mathbf{S}_n(\lambda) \mathbf{y}_n}{\mathbf{y}_n' \mathbf{W}_n' \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n}. \end{aligned} \tag{A.10}$$

By substituting $\mathbf{y}_n = \mathbf{S}_n^{-1} \mathbf{X}_n \boldsymbol{\beta}_0 + \mathbf{S}_n^{-1} \mathbf{u}_n$ and using Lemmas 2–4, one can see that the second term in (A.10) converges almost surely to a bounded constant for all $\lambda \in \Lambda$. In a similar way, one can show

$$\begin{aligned} & b_n''(\lambda) \\ &= \frac{2 \mathbf{y}_n' \mathbf{S}_n'(\lambda) \mathbf{M}_n \text{Dg}(\mathbf{M}_n \mathbf{G}_n^3(\lambda)) \mathbf{M}_n \mathbf{S}_n(\lambda) \mathbf{y}_n + 2 \mathbf{y}_n' \mathbf{W}_n' \mathbf{M}_n \text{Dg}(\mathbf{M}_n \mathbf{G}_n(\lambda)) \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n}{\mathbf{y}_n' \mathbf{W}_n' \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n} \end{aligned}$$

also converges almost surely to a bounded constant. Thus, for some λ^* that lies between λ_0 and λ ,

$$\left| \frac{b_n'(\lambda_0) - b_n'(\lambda)}{b_n'(\lambda)} \right| = |\lambda - \lambda_0| \left| \frac{b_n''(\lambda^*)}{b_n'(\lambda)} \right| < \frac{\zeta}{s_n} \left| \frac{b_n''(\lambda^*)}{b_n'(\lambda)} \right| \xrightarrow{a.s.} 0.$$

With all these results, (12) follows immediately from Theorem 1 of Phillips (2012) and Theorem 2 in this paper. \square

Proof of Theorem 4

Upon substitution,

$$\begin{aligned}
\hat{\beta}_n^{II} &= (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \hat{\mathbf{S}}_n \mathbf{S}_n^{-1} \mathbf{X}_n \beta_0 + (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \hat{\mathbf{S}}_n \mathbf{S}_n^{-1} \mathbf{u}_n \\
&= (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \left[\mathbf{S}_n - (\hat{\lambda}_n^{II} - \lambda_0) \mathbf{W}_n \right] \mathbf{S}_n^{-1} \mathbf{X}_n \beta_0 \\
&\quad + (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \left[\mathbf{S}_n - (\hat{\lambda}_n^{II} - \lambda_0) \mathbf{W}_n \right] \mathbf{S}_n^{-1} \mathbf{u}_n \\
&= \beta_0 + (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{u}_n - (\hat{\lambda}_n^{II} - \lambda_0) \left[(\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{G}_n \mathbf{X}_n \beta_0 + (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{G}_n \mathbf{u}_n \right],
\end{aligned}$$

where $\hat{\lambda}_n^{II} - \lambda_0 = O_p(n^{-1/2})$ (from Theorem 3). Further, $(\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{u}_n = O_p(n^{-1/2})$, $(\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{G}_n \mathbf{X}_n \beta_0 = O(1)$, and $(\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{G}_n \mathbf{u}_n = O_p(n^{-1/2})$ (from Lemma 5). Note that if one expands $\hat{\lambda}_n^{II} - \lambda_0 = b_n^{-1}(\hat{\lambda}_n) - b_n^{-1}(b_n(\lambda_0))$,

$$\sqrt{n}(\hat{\lambda}_n^{II} - \lambda_0) = \frac{\sqrt{n}(\hat{\lambda}_n - b_n(\lambda_0))}{b'_n(\lambda_0)} + O_P(n^{-1/2}). \quad (\text{A.11})$$

From the proof of Theorem 2,

$$\hat{\lambda}_n - b_n(\lambda_0) = \frac{r_n - \mathbf{u}'_n \mathbf{D}_n \mathbf{u}_n}{\mathbb{E}(d_n)} + o_P(n^{-1/2}) = \frac{\mathbf{u}'_n \mathbf{E}_n \mathbf{u}_n + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{u}_n}{\mathbb{E}(d_n)} + o_P(n^{-1/2}). \quad (\text{A.12})$$

Given the above results, one has

$$\begin{aligned}
&\sqrt{n}(\hat{\beta}_n^{II} - \beta_0) \\
&= \sqrt{n}(\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{u}_n - \sqrt{n}(\hat{\lambda}_n^{II} - \lambda_0)(\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{G}_n \mathbf{X}_n \beta_0 + o_P(1) \\
&= \sqrt{n}(\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{u}_n - \left[\frac{\sqrt{n}(\hat{\lambda}_n - b_n(\lambda_0))}{b'_n(\lambda_0)} \right] (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{G}_n \mathbf{X}_n \beta_0 + o_P(1) \\
&= \sqrt{n}(\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{u}_n \\
&\quad - \frac{\sqrt{n}(\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{G}_n \mathbf{X}_n \beta_0}{b'_n(\lambda_0)} \frac{\mathbf{u}'_n \mathbf{E}_n \mathbf{u}_n + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{u}_n}{\mathbb{E}(d_n)} + o_P(1). \quad (\text{A.13})
\end{aligned}$$

One can check that Lemma 7 can be applied to each element or any nonstochastic linear combination of elements of $\sqrt{n}(\hat{\beta}_n^{II} - \beta_0)$ under such a representation. So $\sqrt{n}(\hat{\beta}_n^{II} - \beta_0)$ converges

to a normal distribution, with the asymptotic covariance matrix,

$$\begin{aligned}
\mathbf{V} = & \lim_{n \rightarrow \infty} \left\{ n(\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \boldsymbol{\Sigma}_n \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} \right. \\
& + \frac{\eta}{b_0^2} (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{G}_n \mathbf{X}_n \beta_0 \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} \\
& - \frac{n(\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{G}_n \mathbf{X}_n \beta_0 \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \boldsymbol{\Sigma}_n \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1}}{b_0 [\text{tr}(\boldsymbol{\Sigma}_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n) + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0]} \\
& \left. - \frac{n(\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \boldsymbol{\Sigma}_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0 \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1}}{b_0 [\text{tr}(\boldsymbol{\Sigma}_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n) + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0]} \right\}. \quad (\text{A.14})
\end{aligned}$$

The asymptotic covariance between $\hat{\beta}_n^{II}$ and $\hat{\lambda}_n^{II}$ follows from the expansion of $\sqrt{n}(\hat{\lambda}_n^{II} - \lambda_0)$ (see (A.11) and (A.12)) and that of $\sqrt{n}(\hat{\beta}_n^{II} - \beta_0)$ (see (A.13)), given by

$$\boldsymbol{\gamma} = \lim_{n \rightarrow \infty} \left[\frac{n}{b_0 \mathbb{E}(d_n)} (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \boldsymbol{\Sigma}_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0 - \frac{\eta}{b_0^2} (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{G}_n \mathbf{X}_n \beta_0 \right]. \quad (\text{A.15})$$

□

Proof of Corollary 2

By substitution and using the Nagar (1959) expansion,

$$\begin{aligned}
\sqrt{\frac{n}{h_n}} \left(\hat{\lambda}_n - \lambda_0 - \frac{\mathbf{y}'_n \mathbf{S}'_n \mathbf{D}_n \mathbf{S}_n \mathbf{y}_n}{\mathbf{y}'_n \mathbf{W}'_n \mathbf{W}_n \mathbf{y}_n} \right) &= \sqrt{\frac{n}{h_n}} \left(\frac{r_{1n}}{d_{1n}} - \frac{\mathbf{u}'_n \mathbf{D}_n \mathbf{u}_n}{d_{1n}} \right) \\
&= \sqrt{\frac{n}{h_n}} \left(\frac{\mathbf{u}'_n \mathbf{E}_n \mathbf{u}_n}{d_{1n}} \right) \\
&= \sqrt{\frac{n}{h_n}} \frac{\mathbf{u}'_n \mathbf{E}_n \mathbf{u}_n}{\mathbb{E}(d_{1n})} + o_P(1),
\end{aligned}$$

where $\mathbb{E}(\mathbf{u}'_n \mathbf{E}_n \mathbf{u}_n) = 0$, $\text{Var}(\mathbf{u}'_n \mathbf{E}_n \mathbf{u}_n) = \text{tr}[\boldsymbol{\Sigma}_n \mathbf{E}_n \boldsymbol{\Sigma}_n (\mathbf{E}_n + \mathbf{E}'_n)] = O(n/h_n)$, and the asymptotic distribution follows when one applies Lemma 7 to the quadratic form $\mathbf{u}'_n \mathbf{E}_n \mathbf{u}_n$. □

The Case of Homoscedastic Error Term

If $\boldsymbol{\Sigma}_n = \sigma^2 \mathbf{I}_n$ (and further $\boldsymbol{\Sigma}_n^{(j)} = \mu_j \mathbf{I}_n$, $\mu_j = \mathbb{E}(u_{i,n}^j)$, $j = 3, 4$), then the recentering term in (7) becomes $\sigma^2 \text{tr}(\mathbf{M}_n \mathbf{G}_n) / \mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n$. To make the II procedure feasible, one may replace $\sigma^2 \text{tr}(\mathbf{M}_n \mathbf{G}_n) / \mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n$ by $n^{-1} \mathbf{y}'_n \mathbf{S}'_n \mathbf{M}_n \mathbf{S}_n \mathbf{y}_n \text{tr}(\mathbf{M}_n \mathbf{G}_n) / \mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n$. Similar

to the proof of Theorem 2, one can put

$$\sqrt{n} \left(\hat{\lambda}_n - \lambda_0 - \frac{n^{-1} \mathbf{y}'_n \mathbf{S}'_n \mathbf{M}_n \mathbf{S}_n \mathbf{y}_n \text{tr}(\mathbf{M}_n \mathbf{G}_n)}{\mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n} \right) = \sqrt{n} \left(\frac{r_n^* - \mathbb{E}(r_n^*)}{\mathbb{E}(d_n)} \right) + o_P(1) \xrightarrow{d} \mathbf{N}(0, \omega),$$

where $\omega = \lim_{n \rightarrow \infty} n \text{Var}(r_n^*) / \{\mathbb{E}(d_n)\}^2$, $r_n^* = \mathbf{u}'_n \mathbf{M}_n \mathbf{G}_n^* \mathbf{u}_n + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{u}_n$, $\mathbf{G}_n^* = \mathbf{G}_n - n^{-1} \text{tr}(\mathbf{M}_n \mathbf{G}_n)$. In particular,

$$\begin{aligned} \text{Var}(r_n^*) &= \mu_4 \text{tr}(\mathbf{M}_n \mathbf{G}_n^* \odot \mathbf{M}_n \mathbf{G}_n^*) + \sigma^2 \text{tr}[\mathbf{M}_n \mathbf{G}_n^* (\mathbf{M}_n \mathbf{G}_n^* + \mathbf{G}_n^{*'} \mathbf{M}_n)] \\ &\quad + \sigma^2 \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0 + 2\mu_3 \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \text{dg}(\mathbf{M}_n \mathbf{G}_n^*). \end{aligned}$$

Define the binding function as $b_n(\lambda) = \lambda + n^{-1} \mathbf{y}'_n \mathbf{S}'_n(\lambda) \mathbf{M}_n \mathbf{S}_n(\lambda) \mathbf{y}_n \text{tr}(\mathbf{M}_n \mathbf{G}_n(\lambda)) / \mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n$ and $b_0 = 1 + \text{plim}_{n \rightarrow \infty} (n \mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n)^{-1} [\mathbf{y}'_n \mathbf{S}'_n \mathbf{M}_n \mathbf{S}_n \mathbf{y}_n \text{tr}(\mathbf{M}_n \mathbf{G}_n^2) - 2 \mathbf{y}'_n \mathbf{W}'_n \mathbf{M}_n \mathbf{S}_n \mathbf{y}_n \text{tr}(\mathbf{M}_n \mathbf{G}_n)]$. The asymptotic distribution of $(\hat{\lambda}_n^{II}, \hat{\beta}_n^{II})'$ resulting from this binding function follows similarly from the proofs of Theorems 3 and 4, given by

$$\sqrt{n} \begin{pmatrix} \hat{\lambda}_n^{II} - \lambda_0 \\ \hat{\beta}_n^{II} - \beta_0 \end{pmatrix} \xrightarrow{d} \mathbf{N} \left(\mathbf{0}, \begin{pmatrix} \frac{\omega}{b_0^2} & \gamma' \\ \gamma & \mathbf{V} \end{pmatrix} \right),$$

where

$$\begin{aligned} \mathbf{V} &= \lim_{n \rightarrow \infty} \left\{ n \sigma^2 (\mathbf{X}'_n \mathbf{X}_n)^{-1} + \frac{\omega}{b_0^2} (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{G}_n \mathbf{X}_n \beta_0 \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} \right. \\ &\quad - \frac{n (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{G}_n \mathbf{X}_n \beta_0 [\sigma^2 \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n + \mu_3 \text{dg}(\mathbf{M}_n \mathbf{G}_n^*)'] \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1}}{b_0 [\sigma^2 \text{tr}(\mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n) + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0]} \\ &\quad \left. - \frac{n (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n [\sigma^2 \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0 + \mu_3 \text{dg}(\mathbf{M}_n \mathbf{G}_n^*)] \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1}}{b_0 [\sigma^2 \text{tr}(\mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n) + \beta'_0 \mathbf{X}'_n \mathbf{G}'_n \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0]} \right\} \end{aligned}$$

and

$$\gamma = \lim_{n \rightarrow \infty} \left\{ \frac{n}{b_0 \mathbb{E}(d_n)} (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n [\sigma^2 \mathbf{M}_n \mathbf{G}_n \mathbf{X}_n \beta_0 + \mu_3 \text{dg}(\mathbf{M}_n \mathbf{G}_n^*)] - \frac{\omega}{b_0^2} (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \mathbf{G}_n \mathbf{X}_n \beta_0 \right\}.$$

Acknowledgements

The authors are grateful to three anonymous referees, seminar participants at Huazhong Agricultural University, Huazhong University of Science and Technology, Nanjing Audit University, Shandong University, South China Normal University, and University of California, Riverside

and conference participants at the 2019 CES conference (Dalian) for their helpful comments. Jeff Ello from the Krannert Computing Center at Purdue University kindly created a virtual machine from a computer cluster to facilitate the simulations conducted in an early version of this paper. Lihong Yang thanks support from the National Natural Science Foundation of China (Grant number 71573269). The authors are responsible for all remaining errors.

References

- Badinger, H. and Egger, P. (2011). ‘Estimation of higher-order spatial autoregressive cross-section models with heteroscedastic disturbances’, *Papers in Regional Science*, Vol. 90, pp. 213–235.
- Case, A. C. (1991). ‘Spatial patterns in household demand’, *Econometrica*, Vol. 59, pp. 953–965.
- Case, A. C. and Katz, L. F. (1991). ‘The company you keep: The effects of family and neighborhood on disadvantaged youths’, Technical report, National Bureau of Economic Research.
- Cheng, X., Liao, Z. and Shi, R. (2019). ‘On uniform asymptotic risk of averaging GMM estimators’, *Quantitative Economics*, forthcoming.
- Cliff, A. D. and Ord, J. K. (1981). *Spatial Processes: Models and Applications*, Pion Ltd, London.
- Debarsy, N., Jin, F. and Lee, L.F. (2015). ‘Large sample properties of the matrix exponential spatial specification with an application to FDI’, *Journal of Econometrics*, Vol. 188, pp.1–21.
- Gospodinov, N., Komunjer, I. and Ng, S. (2017) ‘Simulated minimum distance estimation of dynamic models with errors-in-variables’, *Journal of Econometrics*, Vol. 200, pp. 181–193.
- Gouriéroux, C., Monfort, A. and Renault, E. (1993). ‘Indirect inference’, *Journal of Applied Econometrics*, Vol. 8, pp. S85–S118.
- Jin, F. and Lee, L. F. (2019), ‘GEL estimation and tests of spatial autoregressive models’, *Journal of Econometrics*, Vol. 208, pp. 585–612.
- Kelejian, H. H. and Prucha, I. R. (1999). ‘A generalized moments estimator for the autoregressive parameter in a spatial model’, *International Economic Review*, Vol. 40, pp. 509–533.
- Kelejian, H. H. and Prucha, I. R. (2001). ‘On the asymptotic distribution of the Moran I test statistic with applications’, *Journal of Econometrics*, Vol. 104, pp. 219–257.
- Kelejian, H. H. and Prucha, I. R. (2010). ‘Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances’, *Journal of Econometrics*, Vol. 157, pp. 53–67.
- Kyriacou, M., Phillips, P. C. B. and Rossi, F. (2017). ‘Indirect inference in spatial autoregression’, *The Econometrics Journal*, Vol. 20, pp. 168–189.
- Lam, C. and Souza, P. C. L. (2019). ‘Estimation and selection of spatial weight matrix in a spatial lag model’, *Journal of Business and Economic Statistics*, forthcoming.
- Lee, L. F. (2001). ‘Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models I: Spatial autoregressive processes’, Manuscript, Ohio State University.

- Lee, L. F. (2002). ‘Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models’, *Econometric Theory*, Vol. 18, pp. 252–277.
- Lee, L. F. (2004). ‘Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models’, *Econometrica*, Vol. 72, pp. 1899–1925.
- Lee, L.F. (2007). ‘GMM and 2SLS estimation of mixed regressive, spatial autoregressive models’, *Journal of Econometrics* Vol. 137, pp. 489–514.
- Lin, X. and Lee, L. F. (2010). ‘GMM estimation of spatial autoregressive models with unknown heteroskedasticity’, *Journal of Econometrics*, Vol. 157, pp. 34–52.
- Liu, S. F. and Yang, Z. (2015). ‘Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and nonnormality’, *Regional Science and Urban Economics*, Vol. 52, pp. 50–70.
- Nagar, A. L. (1959). ‘The bias and moment matrix of the general k -class estimators of the parameters in simultaneous equations’, *Econometrica*, Vol. 27, pp. 575–595.
- Phillips, P. C. B. (2012). ‘Folklore theorems, implicit maps, and indirect inference’, *Econometrica*, Vol. 80, pp. 425–454.
- Robinson, P. M. (2008). ‘Correlation testing in time series, spatial and cross-sectional data’, *Journal of Econometrics*, Vol. 147, pp. 5–16.
- Smith Jr., A. A. (1993). ‘Estimating nonlinear time-series models using simulated vector autoregressions’, *Journal of Applied Econometrics*, Vol. 8, pp. S63–S84.
- Zhang, X. and Yu, J. (2018). ‘Spatial weights matrix election and model averaging for spatial autoregressive models’, *Journal of Econometrics*, Vol. 203, pp. 1–18.