Article

Indirect Inference Estimation of Spatial Autoregressions

Yong Bao [1,*], Xiaotian Liu [1] and Lihong Yang [2,3]

[1] Department of Economics, Purdue University, West Lafayette, IN 47907, USA; liu2004@purdue.edu
[2] School of Economics, Nanjing Audit University, Nanjing 211815, China; lihong.yang@ruc.edu.cn
[3] National Academy of Development and Strategy, Renmin University of China, Beijing 100872, China
[*] Correspondence: ybao@purdue.edu

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.

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 method 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]

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 structure 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, which 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. Some useful lemmas are collected in Appendix A and proofs of the results presented in Sections 2 and 3 are given in Appendix B.

Throughout this paper, 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. 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 a_n, $a_{i,n}$ denotes its i-th element and for an $n \times n$ matrix A_n, $a_{ij,n}$ denotes its ij-th element. $||A_n||_\infty = \max_{1 \le i < n} \sum_{j=1}^n |a_{ij,n}|$ and $||A_n||_1 = \max_{1 \le j < n} \sum_{i=1}^n |a_{ij,n}|$ are the maximum row sum norm and maximum column sum norm, respectively. A sequence of matrices $\{A_n\}$ is uniformly bounded in row sum if $||A_n||_\infty \le K$, and is uniformly bounded in column sum if $||A_n||_1 \le K$. tr and $\odot$ are matrix trace and Hadamard product operators, respectively. $\mathrm{Dg}(a_n)$ denotes a square diagonal matrix with the vector a_n spanning the main diagonal, $\mathrm{dg}(A_n)$ is an $n \times 1$ column vector that collects in order

---

[1] Recent 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.

the diagonal elements of the square matrix A_n, and $\mathrm{Dg}(A_n) = \mathrm{Dg}(\mathrm{dg}(A_n))$. The subscript 0 is used to signify the true parameter value.

2. Main Results

Consider the SAR model

$$y_n = \lambda W_n y_n + X_n \beta + u_n = Z_n \theta + u_n, \tag{1}$$

where n is the total number of cross-sectional units, $Z_n = (W_n y_n, X_n)$, $\theta = (\lambda, \beta')'$, y_n is an $n \times 1$ vector, collecting observations on the dependent variable, X_n is an $n \times k$ matrix of observations on k exogenous nonstochastic regressors with coefficient vector β, W_n is an $n \times n$ matrix of spatial weights with zero diagonals, λ is the spatial autoregression coefficient, and u_n is an n-dimensional vector of error terms.

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

$$S_n(\lambda) = I_n - \lambda W_n, \qquad G_n(\lambda) = W_n S_n^{-1}(\lambda), \qquad M_n = I_n - X_n(X_n' X_n)^{-1} X_n',$$
$$D_n(\lambda) = \mathrm{Dg}(M_n G_n(\lambda)), \qquad E_n(\lambda) = M_n G_n(\lambda) - D_n(\lambda). \tag{2}$$

When a matrix is presented without its argument λ, it means that it is evaluated at the parameter value λ_0. Namely, $S_n = S_n(\lambda_0)$, $G_n = G_n(\lambda_0)$, $D_n = D_n(\lambda_0)$, and $E_n = E_n(\lambda_0)$.

If λ is known (equal to its true value), the model becomes a standard linear model with $S_n y_n$ being the dependent variable; otherwise, $W_n y_n$ appears on the right-hand side of (1) as a spatially lagged or weighted variable. The OLS estimator of θ_0 is $\hat{\theta}_n = (Z_n' Z_n)^{-1} Z_n' y_n$, which may be inconsistent. Since the inconsistency of $\hat{\theta}_n$ is solely due to the endogenous $W_n y_n$, the properties of $\hat{\lambda}_n$ (the first element of $\hat{\theta}_n$) are discussed first in this paper. Once a consistent estimator of λ_0 is available, one can show that a consistent estimator of β_0 follows immediately (see Theorem 4 to be introduced).[2]

Let $r_n = r_{1n} + r_{2n}$, $r_{1n} = u_n' M_n G_n u_n$, $r_{2n} = \beta_0' X_n' G_n' M_n u_n$, $d_n = d_{1n} + d_{2n} + d_{3n}$, $d_{1n} = u_n' G_n' M_n G_n u_n$, $d_{2n} = 2\beta_0' X_n' G_n' M_n G_n u_n$, and $d_{3n} = \beta_0' X_n' G_n' M_n G_n X_n \beta_0$. By using the partitioned regression formula and substituting $W_n y_n = G_n X_n \beta_0 + G_n u_n$, one may put

$$\hat{\lambda}_n - \lambda_0 = \frac{y_n' W_n' M_n u_n}{y_n' W_n' M_n W_n y_n} = \frac{r_n}{d_n} = \frac{r_{1n} + r_{2n}}{d_{1n} + d_{2n} + d_{3n}}. \tag{3}$$

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 $\lim\limits_{n \to \infty} h_n / n = 0$; (ii) the sequence $\{W_n\}$ is is uniformly bounded in row and column sums; (iii) $w_{ii,n} = 0$.*

Assumption 2. *(i) S_n^{-1} exists; (ii) the sequence $\{S_n^{-1}\}$ is uniformly bounded in row and column sums.*

Assumption 3. *The error terms $\{u_{i,n}\}$ in $u_n = (u_{1,n}, \cdots, u_{n,n})'$ have following properties: (i) $\mathrm{E}(u_{i,n}) = 0$; (ii) $\mathrm{E}(|u_{i,n}|^{4+\delta}) < \infty$ 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 $\lambda \in \Lambda$, $\{S_n^{-1}(\lambda)\}$ is uniformly bounded in row and column sums.*

---

[2]　It should be pointed out when $\lim\limits_{n \to \infty} \sqrt{n}/h_n = 0$ (and $\lim\limits_{n \to \infty} h_n/n = 0$, where h_n^{-1} is the order of magnitude of elements of W_n), $\hat{\theta}_n$ 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 β_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.

Assumption 5. *(i) The elements of X_n are uniformly bounded constants for all n; (ii) the probability limit Γ of $n^{-1}Z'_n Z_n$ exists and is nonsingular.*

Assumption 6. *(i) $\lim_{n\to\infty} n^{-1}\mathrm{Var}(r_n)$ exists and is nonzero; (ii) $\lim_{n\to\infty} n^{-1}\mathrm{Var}(u'_n E_n u_n + \beta'_0 X'_n G'_n M_n u_n)$ exists and is nonzero.*

Intuitions and related discussions for Assumptions 1–5 are provided in Kelejian and Prucha (2010) and Lee (2001, 2002, 2004). Assumption 1(i) follows naturally when 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 y_n is $y_n = S_n^{-1}X_n\beta_0 + S_n^{-1}u_n$. Under Assumption 3, let $\sigma^2_{i,n} = \mathrm{E}(u^2_{i,n})$, $\Sigma_n = \mathrm{Dg}(\sigma^2_{1,n},\cdots,\sigma^2_{n,n})$, and $\Sigma_n^{(j)} = \mathrm{Dg}(\mu^{(j)}_{1,n},\cdots,\mu^{(j)}_{n,n})$ with $\mu^{(j)}_{i,n} = \mathrm{E}(u^j_{i,n})$, $j = 3,4$. 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 X_n and $G_n X_n \beta_0$ 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]

2.1. The Asymptotic Behavior of the OLS estimator

From Lemma A6 in Appendix A, 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 $h_n \to \infty$, 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 $h_n \to \infty$.

---

3   Multicollinearities can happen, for example, when $X_n = \mathbf{1}_n$ and W_n is row-normalized. Lee (2004) showed that under homoscedasticity, however, the QML estimator can still be consistent in spite of 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.

4   The asymptotic variances are given by $\lim_{n\to\infty} n^{-1}\mathrm{Var}(r_n)/[\mathrm{plim}(d_n/n)]^2 = \lim_{n\to\infty}[\mathrm{Var}(r_n)/n]/[\mathrm{E}(d_n)/n]^2$ and $\lim_{n\to\infty} n^{-1}\mathrm{Var}(u'_n E_n u_n + \beta'_0 X'_n G'_n M_n u_n)/[\mathrm{plim}(d_n/n)]^2 = \lim_{n\to\infty}[\mathrm{Var}(u'_n E_n u_n + \beta'_0 X'_n G'_n M_n u_n)/n]/[\mathrm{E}(d_n)/n]^2$, 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 $\mathrm{plim}_{n\to\infty}(d_n/n) = \mathrm{plim}_{n\to\infty}(y'_n W'_n M_n W_n y_n/n)$ exists and is nonzero. It can be shown (see Appendix A) that $\mathrm{Var}(r_n) = \mathrm{Var}(u'_n M_n G_n u_n) + \beta'_0 X'_n G'_n M_n \Sigma_n M_n G_n X_n \beta_0 + 2\mathrm{Cov}(\beta'_0 X'_n G'_n M_n \Sigma_n M_n G_n X_n \beta_0, u'_n M_n G_n u_n)$, where the covariance term disappears under normality, and $\mathrm{Var}(u'_n E_n u_n + \beta'_0 X'_n G'_n M_n u_n) = \mathrm{Var}(u'_n E_n u_n) + \beta'_0 X'_n G'_n M_n \Sigma_n M_n G_n X_n \beta_0$. When h_n diverges, $\beta'_0 X'_n G'_n M_n \Sigma_n M_n G_n X_n \beta_0$ is the dominating term in $\mathrm{Var}(r_n)$ as well as $\mathrm{Var}(u'_n E_n u_n + \beta'_0 X'_n G'_n M_n u_n)$. Then the usual condition that $\mathrm{plim}_{n\to\infty} n^{-1}Z'_n \Sigma_n Z_n$ 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 $\mathrm{plim}_{n\to\infty} n^{-1}Z'_n \Sigma_n Z_n$, the existence of $\lim_{n\to\infty} n^{-1}\mathrm{Var}(u'_n E_n u_n)$ and $\lim_{n\to\infty} n^{-1}\mathrm{Var}(u'_n M_n G_n u_n)$, where $\mathrm{Var}(u'_n E_n u_n) = O(n/h_n)$ and $\mathrm{Var}(u'_n M_n G_n u_n) = O(n/h_n)$.

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

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

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 $\underset{n\to\infty}{\text{plim}}(h_n/n)r_{1n} = \underset{n\to\infty}{\lim}(h_n/n)\text{tr}(\boldsymbol{\Sigma}_n\boldsymbol{M}_n\boldsymbol{G}_n)$ is typically nonzero.

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

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

indicating that $\hat{\lambda}_n$ converges at rate $\sqrt{n}$, but at this rate it does not converge to λ_0, as in general $\underset{n\to\infty}{\text{plim}} n^{-1/2} r_{1n} = \underset{n\to\infty}{\lim} n^{-1/2}\text{tr}(\boldsymbol{\Sigma}_n\boldsymbol{M}_n\boldsymbol{G}_n)$ is nonzero.

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

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

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 SAR model (1) has the following asymptotic distribution,*

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

where

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

with

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

Remark 1. *The recentering term $\text{tr}(\boldsymbol{\Sigma}_n\boldsymbol{M}_n\boldsymbol{G}_n)/\boldsymbol{y}_n'\boldsymbol{W}_n'\boldsymbol{M}_n\boldsymbol{W}_n\boldsymbol{y}_n$ is in fact $\mathrm{E}(r_n)/d_n$. One could have recentered $\hat{\lambda}_n - \lambda_0$ by $\mathrm{E}(r_n)/\mathrm{E}(d_n)$. (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 (see Appendix B.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 $\underset{n\to\infty}{\lim} \sqrt{n}/h_n = 0$, the recentering term $\text{tr}(\boldsymbol{\Sigma}_n\boldsymbol{M}_n\boldsymbol{G}_n)/\boldsymbol{y}_n'\boldsymbol{W}_n'\boldsymbol{M}_n\boldsymbol{W}_n\boldsymbol{y}_n$ 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 $\text{tr}(\boldsymbol{\Sigma}_n\boldsymbol{M}_n\boldsymbol{G}_n)/\boldsymbol{y}_n'\boldsymbol{W}_n'\boldsymbol{M}_n\boldsymbol{W}_n\boldsymbol{y}_n$; nor does one need to recenter it by $\boldsymbol{y}_n'\boldsymbol{S}_n'\boldsymbol{M}_n\boldsymbol{D}_n\boldsymbol{M}_n\boldsymbol{S}_n\boldsymbol{y}_n/\boldsymbol{y}_n'\boldsymbol{W}_n'\boldsymbol{M}_n\boldsymbol{W}_n\boldsymbol{y}_n$, which is also $o_P(n^{-1/2})$, as in Theorem 2 to be introduced.*

Remark 3. *When h_n diverges, one may single out dominating terms in* $\mathrm{Var}(r_n)$ *and* $\mathrm{E}(d_n)$ *so that a finer expression of* $v = \lim_{n\to\infty}[\mathrm{Var}(r_n)/n]/[\mathrm{E}(d_n)/n]^2$ *can be presented. For example, under divergent h_n, one can replace* $\mathrm{Var}(r_n)$ *with* $\mathrm{Var}(r_{2n}) = \boldsymbol{\beta}_0' \boldsymbol{X}_n' \boldsymbol{G}_n' \boldsymbol{M}_n \boldsymbol{\Sigma}_n \boldsymbol{M}_n \boldsymbol{G}_n \boldsymbol{X}_n \boldsymbol{\beta}_0$ *and replace* $\mathrm{E}(d_n)$ *with its dominating term* $\boldsymbol{\beta}_0' \boldsymbol{X}_n' \boldsymbol{G}_n' \boldsymbol{M}_n \boldsymbol{G}_n \boldsymbol{X}_n \boldsymbol{\beta}_0$. *When h_n is bounded, however, such replacements are not available and one needs to keep all the term in* $\mathrm{Var}(r_n)$ *and* $\mathrm{E}(d_n)$. *Moreover, since* $\mathrm{Var}(r_n)$ *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* $\boldsymbol{\Sigma}_n = \sigma_0^2 \boldsymbol{I}_n$, *namely, under homoscedasticity with* $\sigma_0^2 = \mathrm{E}(u_{i,n}^2)$, *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* $\lim_{n\to\infty} \sqrt{n}/h_n = 0$*) 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 $\mathrm{tr}(\boldsymbol{\Sigma}_n \boldsymbol{M}_n \boldsymbol{G}_n) = \mathrm{E}(\boldsymbol{u}_n' \boldsymbol{M}_n \boldsymbol{G}_n \boldsymbol{u}_n)$ with a term that is of the same order of $\boldsymbol{u}_n' \boldsymbol{M}_n \boldsymbol{G}_n \boldsymbol{u}_n$ and at the same time does not directly involve $\boldsymbol{\Sigma}_n$.

Recall that $\boldsymbol{D}_n = \mathrm{Dg}(\boldsymbol{M}_n \boldsymbol{G}_n)$. Then under a general form of unknown heteroscedasticity, $\mathrm{tr}(\boldsymbol{\Sigma}_n \boldsymbol{M}_n \boldsymbol{G}_n) = \mathrm{tr}(\boldsymbol{\Sigma}_n \boldsymbol{D}_n) = \mathrm{E}(\boldsymbol{u}_n' \boldsymbol{D}_n \boldsymbol{u}_n)$, where $\boldsymbol{u}_n = \boldsymbol{S}_n \boldsymbol{y}_n - \boldsymbol{X}_n \boldsymbol{\beta}_0$. If λ_0 is known, then $\boldsymbol{\beta}_0$ can be consistently estimated by $\tilde{\boldsymbol{\beta}}_n = \tilde{\boldsymbol{\beta}}_n(\lambda_0) = (\boldsymbol{X}_n' \boldsymbol{X}_n)^{-1} \boldsymbol{X}_n' \boldsymbol{S}_n \boldsymbol{y}_n$. Let $\tilde{\boldsymbol{u}}_n = \tilde{\boldsymbol{u}}_n(\lambda_0) = \boldsymbol{S}_n \boldsymbol{y}_n - \boldsymbol{X}_n \tilde{\boldsymbol{\beta}}_n = \boldsymbol{M}_n \boldsymbol{S}_n \boldsymbol{y}_n$. Now one may be able to replace $\mathrm{E}(\boldsymbol{u}_n' \boldsymbol{D}_n \boldsymbol{u}_n)$ with $\tilde{\boldsymbol{u}}_n' \boldsymbol{D}_n \tilde{\boldsymbol{u}}_n = \boldsymbol{y}_n' \boldsymbol{S}_n' \boldsymbol{M}_n \boldsymbol{D}_n \boldsymbol{M}_n \boldsymbol{S}_n \boldsymbol{y}_n$ and use $\boldsymbol{y}_n' \boldsymbol{S}_n' \boldsymbol{M}_n \boldsymbol{D}_n \boldsymbol{M}_n \boldsymbol{S}_n \boldsymbol{y}_n / \boldsymbol{y}_n' \boldsymbol{W}_n' \boldsymbol{M}_n \boldsymbol{W}_n \boldsymbol{y}_n$ as the recentering quantity.

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

$$\sqrt{n}\left(\hat{\lambda}_n - \lambda_0 - \frac{\boldsymbol{y}_n' \boldsymbol{S}_n' \boldsymbol{M}_n \boldsymbol{D}_n \boldsymbol{M}_n \boldsymbol{S}_n \boldsymbol{y}_n}{\boldsymbol{y}_n' \boldsymbol{W}_n' \boldsymbol{M}_n \boldsymbol{W}_n \boldsymbol{y}_n}\right) \xrightarrow{d} \mathrm{N}(0, \eta), \tag{8}$$

where

$$\eta = \lim_{n\to\infty} \frac{n\left\{\mathrm{tr}[\boldsymbol{\Sigma}_n \boldsymbol{E}_n \boldsymbol{\Sigma}_n (\boldsymbol{E}_n + \boldsymbol{E}_n')] + \boldsymbol{\beta}_0' \boldsymbol{X}_n' \boldsymbol{G}_n' \boldsymbol{M}_n \boldsymbol{\Sigma}_n \boldsymbol{M}_n \boldsymbol{G}_n \boldsymbol{X}_n \boldsymbol{\beta}_0\right\}}{[\mathrm{tr}(\boldsymbol{\Sigma}_n \boldsymbol{G}_n' \boldsymbol{M}_n \boldsymbol{G}_n) + \boldsymbol{\beta}_0' \boldsymbol{X}_n' \boldsymbol{G}_n' \boldsymbol{M}_n \boldsymbol{G}_n \boldsymbol{X}_n \boldsymbol{\beta}_0]^2}.$$

Remark 6. *When one replaces* $\mathrm{tr}(\boldsymbol{\Sigma}_n \boldsymbol{M}_n \boldsymbol{G}_n)$ *that involves the unknown $\boldsymbol{\Sigma}_n$ and appears in the recentering term of the OLS estimator, the asymptotic variance η of the newly recentered estimator no longer involves* $\boldsymbol{\Sigma}_n^{(3)}$ *and* $\boldsymbol{\Sigma}_n^{(4)}$. *This stands in contrast to the asymptotic variance v (see Remark 3). So replacing* $\mathrm{tr}(\boldsymbol{\Sigma}_n \boldsymbol{M}_n \boldsymbol{G}_n)$ *with* $\boldsymbol{y}_n' \boldsymbol{S}_n' \boldsymbol{M}_n \boldsymbol{D}_n \boldsymbol{M}_n \boldsymbol{S}_n \boldsymbol{y}_n$ *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 $\boldsymbol{y}_n$ and $\boldsymbol{X}_n$, one can always define the sample binding function. Recall that $\boldsymbol{S}_n(\lambda) = \boldsymbol{I}_n - \lambda \boldsymbol{W}_n$ and $\boldsymbol{D}_n(\lambda) = \mathrm{Dg}(\boldsymbol{M}_n \boldsymbol{G}_n(\lambda)) = \mathrm{Dg}(\boldsymbol{M}_n \boldsymbol{W}_n \boldsymbol{S}_n^{-1}(\lambda))$ are functions of the parameter λ (as well as $\boldsymbol{X}_n$). So the binding function can be defined as

$$b_n(\lambda) = \lambda + \frac{\boldsymbol{y}_n' \boldsymbol{S}_n'(\lambda) \boldsymbol{M}_n \boldsymbol{D}_n(\lambda) \boldsymbol{M}_n \boldsymbol{S}_n(\lambda) \boldsymbol{y}_n}{\boldsymbol{y}_n' \boldsymbol{W}_n' \boldsymbol{M}_n \boldsymbol{W}_n \boldsymbol{y}_n} \tag{9}$$

and the II estimator inverts this binding function:

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

Intuitively, the II estimator defined as such tries to match $\hat{\lambda}_n$ from the observed data to its expectation, 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); however, in simulations one needs to make some distributional assumption to generate the pseudo error term. 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 $\boldsymbol{y}_n$ and $\boldsymbol{X}_n$ 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, $b_n'(\lambda_0) \overset{a.s.}{\to} b_0 \neq 0$, where*

$$b_n'(\lambda_0) = 1 + \frac{\boldsymbol{y}_n' \boldsymbol{S}_n' \boldsymbol{M}_n \mathrm{Dg}(\boldsymbol{M}_n \boldsymbol{G}_n^2) \boldsymbol{M}_n \boldsymbol{S}_n \boldsymbol{y}_n - 2\boldsymbol{y}_n' \boldsymbol{W}_n' \boldsymbol{M}_n \boldsymbol{D}_n \boldsymbol{M}_n \boldsymbol{S}_n \boldsymbol{y}_n}{\boldsymbol{y}_n' \boldsymbol{W}_n' \boldsymbol{M}_n \boldsymbol{W}_n \boldsymbol{y}_n}. \tag{11}$$

One can see from the expression (A.10) in Appendix B.3 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 $O_P(h_n^{-1})$ 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 $\boldsymbol{X}_n$, the weight matrix $\boldsymbol{W}_n$, 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 SAR model (1), under Assumptions 1–7, the II estimator $\hat{\lambda}_n^{II}$ of λ_0, defined as in (10), which is based on the binding function (9), has the following asymptotic distribution:*

$$\sqrt{n}(\hat{\lambda}_n^{II} - \lambda_0) \overset{d}{\to} \mathrm{N}\left(0, \frac{\eta}{b_0^2}\right). \tag{12}$$

---

[5]　The 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, $\boldsymbol{\beta}'_0 \boldsymbol{X}'_n \boldsymbol{G}'_n \boldsymbol{M}_n \boldsymbol{\Sigma}_n \boldsymbol{M}_n \boldsymbol{G}_n \boldsymbol{X}_n \boldsymbol{\beta}_0$ dominates* $\mathrm{Var}(\boldsymbol{u}'_n \boldsymbol{E}_n \boldsymbol{u}_n + \boldsymbol{\beta}'_0 \boldsymbol{X}'_n \boldsymbol{G}'_n \boldsymbol{M}_n \boldsymbol{u}_n)$ *and $\boldsymbol{\beta}'_0 \boldsymbol{X}'_n \boldsymbol{G}'_n \boldsymbol{M}_n \boldsymbol{G}_n \boldsymbol{X}_n \boldsymbol{\beta}_0$ dominates* $\mathrm{E}(d_n)$, *so $\eta = v + o(1)$. In light of Remark 4, this means that under homoscedasticity, the II estimator defined as in (10) is as efficient as the optimal IV estimator and can be as efficient as the ML estimator if the spatial data is normally distributed. (The same conclusion holds for the slightly modified II estimator, designed specifically under homoscedasticity, in Appendix B.6.)*

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 $\lim\limits_{n\to\infty} \sqrt{n}/h_n = 0$, 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 $\boldsymbol{\beta}$ by

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

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

Theorem 4. *For the SAR model (1), under Assumptions 1–7, the OLS estimator of $\boldsymbol{\beta}_0$ defined as in (13) has the asymptotic distribution*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n^{II} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathrm{N}(0, \boldsymbol{V}),$$

and jointly,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n^{II} - \boldsymbol{\theta}_0) = \sqrt{n} \begin{pmatrix} \hat{\lambda}_n^{II} - \lambda_0 \\ \hat{\boldsymbol{\beta}}_n^{II} - \boldsymbol{\beta}_0 \end{pmatrix} \xrightarrow{d} \mathrm{N}\left(\boldsymbol{0}, \begin{pmatrix} \frac{\eta}{b_0^2} & \boldsymbol{\gamma}' \\ \boldsymbol{\gamma} & \boldsymbol{V} \end{pmatrix} \right),$$

where $\boldsymbol{V}$, assumed to exist and be positive definite, is given by (A.14) and $\boldsymbol{\gamma}$ is given by (A.15), respectively, in Appendix B.4.

Remark 9. *One can see (from Appendix B.4) that the expression of $\boldsymbol{V}$ contains the traditional OLS variance term under heteroscedasticity, $\lim\limits_{n\to\infty} n(\boldsymbol{X}'_n \boldsymbol{X}_n)^{-1} \boldsymbol{X}'_n \boldsymbol{\Sigma}_n \boldsymbol{X}_n (\boldsymbol{X}'_n \boldsymbol{X}_n)^{-1}$, as well as terms that signal the additional uncertainty introduced by $\hat{\lambda}_n^{II}$ in the definition of $\hat{\boldsymbol{\beta}}_n^{II}$ 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, $\boldsymbol{V}$, and $\boldsymbol{\gamma}$ in Theorems 3 and 4 by $\eta(\hat{\boldsymbol{\theta}}_n^{II})/[b'_n(\hat{\lambda}_n^{II})]^2$, $\boldsymbol{V}(\hat{\boldsymbol{\theta}}_n^{II}, \hat{\boldsymbol{\Sigma}}_n)$, and $\boldsymbol{\gamma}(\hat{\boldsymbol{\theta}}_n^{II}, \hat{\boldsymbol{\Sigma}}_n)$, respectively, where $\hat{\boldsymbol{\Sigma}}_n = \mathrm{Dg}(\hat{u}_{1,n}^2, \cdots, \hat{u}_{n,n}^2)$ and $\hat{u}_{i,n}$'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 $\boldsymbol{X}_n$, namely, the so-called pure SAR model

$$\boldsymbol{y}_n = \lambda \boldsymbol{W}_n \boldsymbol{y}_n + \boldsymbol{u}_n. \tag{14}$$

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

---

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

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

$$\plim_{n\to\infty}(\hat{\lambda}_n - \lambda_0) = \frac{\plim_{n\to\infty}\frac{h_n}{n}r_{1n}}{\plim_{n\to\infty}\frac{h_n}{n}d_{1n}} = \frac{\lim_{n\to\infty}\frac{h_n}{n}\mathrm{tr}(\mathbf{\Sigma}_n\mathbf{G}_n)}{\lim_{n\to\infty}\frac{h_n}{n}\mathrm{tr}(\mathbf{\Sigma}_n\mathbf{G}_n'\mathbf{G}_n)} \neq 0,$$

regardless of the magnitude of h_n. Proceeding similarly as before, as long as $h_n = o(n)$, one can write $r_{1n} - \mathrm{E}(r_{1n}) = O_P(\sqrt{n/h_n})$, $d_{1n} - \mathrm{E}(d_{1n}) = O_P(\sqrt{n/h_n})$, and $\mathrm{E}(d_{1n}) = O(n/h_n)$. (See Lemma A6 in Appendix A). Then, by a Nagar-type (Nagar (1959)) expansion,

$$\sqrt{\frac{n}{h_n}}\left(\hat{\lambda}_n - \lambda_0 - \frac{\mathrm{E}(r_{1n})}{d_{1n}}\right) = \sqrt{\frac{n}{h_n}}\left(\frac{r_{1n} - \mathrm{E}(r_{1n})}{\mathrm{E}(d_{1n}) + d_{1n} - \mathrm{E}(d_{1n})}\right)$$

$$= \sqrt{\frac{n}{h_n}}\frac{r_{1n} - \mathrm{E}(r_{1n})}{\mathrm{E}(d_{1n})} + o_P(1). \tag{15}$$

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 $\mathbf{D}_n = \mathrm{Dg}(\mathbf{G}_n)$ and $\mathbf{E}_n = \mathbf{G}_n - \mathbf{D}_n$.

Assumption 8. *(i)*

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

exists and is positive; (ii)

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

exists and is positive.

Corollary 1. *Under Assumptions 1–4 and 8, the OLS estimator $\hat{\lambda}_n$ of λ in the pure SAR model (14) has the following asymptotic distribution,*

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

where v is defined as in Assumption 8(i).

Corollary 2. *Under Assumptions 1–4 and 8, the OLS estimator $\hat{\lambda}_n$ of λ in the pure SAR model (14) has the following asymptotic distribution:*

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

where η is defined as in Assumption 8(ii).

Let the sample binding function be

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

Accordingly, Assumption 7 is modified as follows.

Assumption 9. *For all λ in Λ, the binding function (18) is monotonic in λ with probability 1 and $b'_n(\lambda_0) \overset{a.s.}{\to} b_0 \neq 0$, where*

$$b'_n(\lambda_0) = 1 + \frac{y'_n S'_n \mathrm{Dg}(G_n^2) S_n y_n - 2 y'_n W'_n D_n S_n y_n}{y'_n W'_n W_n y_n}. \tag{19}$$

Corollary 3. *For the pure SAR model (14), under Assumptions 1–4, 8, and 9, the II estimator $\hat{\lambda}_n^{II}$ of λ that is based on the binding function (18) has the following asymptotic distribution:*

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

where η is defined as in Assumption 8(ii).

Remark 10. *If $\Sigma_n = \sigma_0^2 I_n$ (namely, under homoscedasticity) and ones uses $\mathrm{E}(r_{1n})/\mathrm{E}(d_{1n}) = \mathrm{tr}(G_n)/\mathrm{tr}(G'_n G_n)$ 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 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 X is present. This is because $y'_n W'_n W_n y_n = O_P(n/h_n)$ for the pure case and $y'_n W'_n M_n W_n y_n = O_P(n)$ when 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, since η defined as in Assumption 8(ii) has the scaling factor n/h_n, one can see that 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 $\sqrt{\mathrm{tr}[\hat{\Sigma}_n \hat{E}_n \hat{\Sigma}_n (\hat{E}_n + \hat{E}'_n)]}/[b'_n(\hat{\lambda}_n^{II}) \mathrm{tr}(\hat{\Sigma}_n \hat{G}'_n \hat{G}_n)]^2$, where $\hat{G}_n = G_n(\hat{\lambda}_n^{II})$, $\hat{E}_n = E_n(\hat{\lambda}_n^{II})$, and $\hat{\Sigma}_n = \mathrm{Dg}(\hat{u}_{1,n}^2, \cdots, \hat{u}_{n,n}^2)$.*

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), $u'_n \Sigma_n^{-1} (G_n(\lambda) - \mathrm{Dg}(G_n(\lambda)) u_n$ and $u'_n \Sigma_n^{-1} (G_n(\lambda) X_n \beta, X_n)$ are used as the optimal moment conditions, see Debarsy et al. (2015). They involve λ (appearing in $G_n(\lambda)$) and β as well as the covariance matrix Σ_n. For λ and β, an initial estimation is constructed from the simple 2SLS with $W_n X_n$ and 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 $u'_n(G_n - \mathrm{Dg}(G_n) u_n$ and $u'_n(G_n X_n \beta, X_n)$ as the moment conditions and the other is to use $u'_n \Sigma_n^{-1}(G_n - \mathrm{Dg}(G_n) u_n$ and $u'_n \Sigma_n^{-1}(G_n X_n \beta, X_n)$ with the true Σ_n (known in simulations) plugged in. The two resulting estimators are denoted by GMM and GMM(Σ_n), respectively, in Tables 1–4. One would expect that in practice the performance of the optimal robust GMM estimator with an estimated Σ_n appearing in the moment conditions would most likely stand between the two.

In each experiment, for each estimator, reported are the bias and root mean squared error (RMSE) from 1000 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 $\hat{\beta}_n^{II}$, 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.

---

[7]   Neither 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 spatial autoregressions (SAR) with variance structure V1 and parameter configuration P1.

R	θ_0	MQML			GMM			GMM(Σ_n)			II		
		Bias	**RMSE**	$P(5\%)$	**Bias**	**RMSE**	$P(5\%)$	**Bias**	**RMSE**	$P(5\%)$	**Bias**	**RMSE**	$P(5\%)$
100	0.2	−0.009	0.067	3.6%	−0.015	0.073	5.8%	0.000	0.01	4.9%	−0.009	0.067	3.8%
	0.8	0.034	0.367	4.5%	0.050	0.377	5.4%	0.001	0.036	5.6%	0.033	0.367	4.7%
	0.2	−0.002	0.099	5.0%	−0.002	0.099	5.1%	0.000	0.008	5.5%	−0.002	0.099	5.0%
	1.5	−0.004	0.088	4.9%	−0.005	0.088	5.1%	0.000	0.007	4.1%	−0.004	0.088	4.9%
	0.6	−0.005	0.035	3.0%	−0.008	0.039	7.2%	0.000	0.005	5.3%	−0.005	0.035	4.1%
	0.8	0.032	0.366	3.8%	0.047	0.378	5.9%	0.001	0.037	5.6%	0.031	0.367	4.7%
	0.2	−0.001	0.102	5.5%	−0.001	0.102	5.5%	0.000	0.008	5.7%	−0.001	0.102	5.5%
	1.5	0.004	0.089	4.9%	0.004	0.089	4.9%	0.000	0.007	4.6%	0.004	0.089	4.9%
	0.9	−0.001	0.009	0.3%	−0.002	0.01	7.7%	0.000	0.001	6.1%	−0.001	0.009	4.3%
	0.8	0.022	0.367	4.0%	0.037	0.378	6.0%	0.003	0.04	6.5%	0.02	0.367	5.6%
	0.2	−0.002	0.102	4.4%	−0.002	0.102	4.4%	0.000	0.008	5.3%	−0.002	0.102	4.4%
	1.5	0.000	0.093	5.9%	0.000	0.094	6.0%	0.000	0.007	4.7%	0.00	0.094	6.1%
200	0.2	0.000	0.047	5.6%	−0.003	0.051	8.0%	0.000	0.007	5.3%	0.000	0.047	6.1%
	0.8	−0.005	0.255	5.1%	0.003	0.261	5.8%	0.001	0.026	6.9%	−0.005	0.255	5.2%
	0.2	0.002	0.071	5.8%	0.001	0.071	5.7%	0.000	0.006	5.4%	0.002	0.071	5.8%
	1.5	0.003	0.061	3.3%	0.002	0.061	3.4%	0.000	0.005	5.4%	0.003	0.061	3.3%
	0.6	−0.003	0.025	2.2%	−0.005	0.028	8.6%	0.000	0.004	5.1%	−0.003	0.025	5.0%
	0.8	0.021	0.256	4.1%	0.030	0.264	6.0%	0.000	0.026	5.2%	0.020	0.256	4.7%
	0.2	−0.001	0.069	4.5%	−0.001	0.069	4.6%	0.000	0.006	5.7%	−0.001	0.069	4.5%
	1.5	0.000	0.062	4.2%	−0.001	0.062	4.3%	0.000	0.005	4.9%	0.000	0.062	4.2%
	0.9	−0.001	0.006	0.7%	−0.001	0.007	8.3%	0.000	0.001	6.0%	−0.001	0.006	4.7%
	0.8	0.022	0.262	4.2%	0.030	0.269	6.7%	0.000	0.026	4.5%	0.021	0.262	5.7%
	0.2	−0.001	0.072	5.5%	−0.001	0.072	5.3%	0.000	0.005	4.8%	−0.001	0.072	5.5%
	1.5	0.000	0.063	3.8%	0.000	0.063	3.9%	0.000	0.005	4.4%	0.000	0.063	3.8%

The weight matrix is a social interaction matrix $W_n = I_R \otimes [(\mathbf{1}_{m_j}\mathbf{1}'_{m_j} - I_{m_j})/(m_j - 1)]$, $m_j \sim$ IID U(3, 20), $j = 1, \cdots, R$. Reported for each estimator are the bias, root mean squared error (RMSE), and the empirical size of the t test for each parameter equal to its true value at 5% from 1000 simulations. The exogenous regressors are $(1, x_{i1}, x_{i2})'$, $x_{i1} \sim$ IID N(3, 1), $x_{i2} \sim$ 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 $1/m_j^2$, $j = 1, \cdots, R$.

Table 2. Estimation of SAR with variance structure V1 and parameter configuration P2.

R	θ_0	MQML			GMM			GMM(Σ_n)			II		
		Bias	**RMSE**	$P(5\%)$	**Bias**	**RMSE**	$P(5\%)$	**Bias**	**RMSE**	$P(5\%)$	**Bias**	**RMSE**	$P(5\%)$
100	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%
200	0.2	−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%

The weight matrix is a social interaction matrix $W_n = I_R \otimes [(\mathbf{1}_{m_j}\mathbf{1}'_{m_j} - I_{m_j})/(m_j-1)]$, $m_j \sim$ IID U(3, 20), $j = 1, \cdots, R$. 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 1000 simulations. The exogenous regressors are $(1, x_{i1}, x_{i2})'$, $x_{i1} \sim$ IID N(3, 1), $x_{i2} \sim$ 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 $1/m_j^2$, $j = 1, \cdots, R$.

Table 3. Estimation of SAR with variance structure V2 and parameter configuration P1.

R	θ_0	MQML			GMM			GMM(Σ_n)			II		
		Bias	**RMSE**	$P(5\%)$	**Bias**	**RMSE**	$P(5\%)$	**Bias**	**RMSE**	$P(5\%)$	**Bias**	**RMSE**	$P(5\%)$
100	0.2	0.000	0.014	4.6%	0.000	0.014	5.2%	−0.001	0.013	4.9%	0.000	0.014	5.0%
	0.8	0.001	0.047	3.6%	0.000	0.047	3.7%	0.003	0.042	3.7%	0.001	0.046	3.7%
	0.2	0.000	0.008	4.1%	0.000	0.008	4.1%	0.000	0.008	5.1%	0.000	0.008	4.2%
	1.5	0.000	0.008	5.3%	0.000	0.008	5.4%	0.000	0.007	4.6%	0.000	0.008	5.3%
	0.6	0.000	0.008	4.3%	0.000	0.008	4.9%	−0.001	0.007	4.3%	0.000	0.008	4.6%
	0.8	0.001	0.049	4.5%	0.002	0.049	5.7%	0.002	0.043	4.3%	0.001	0.049	5.5%
	0.2	0.000	0.008	5.3%	0.000	0.008	5.1%	0.000	0.008	4.9%	0.000	0.008	5.2%
	1.5	0.000	0.008	5.3%	0.000	0.008	5.4%	0.000	0.007	4.7%	0.000	0.008	5.3%
	0.9	0.000	0.002	5.1%	0.000	0.002	7.8%	0.000	0.002	5.9%	0.000	0.002	6.8%
	0.8	0.002	0.050	4.1%	0.002	0.050	5.8%	0.002	0.045	4.9%	0.002	0.049	4.6%
	0.2	0.000	0.008	5.4%	0.000	0.008	5.4%	0.000	0.008	5.6%	0.000	0.008	5.4%
	1.5	0.000	0.008	5.2%	0.000	0.008	5.2%	0.000	0.007	4.7%	0.000	0.008	5.0%
200	0.2	0.000	0.010	3.8%	0.000	0.010	4.7%	0.000	0.009	3.8%	0.000	0.010	4.4%
	0.8	0.001	0.033	4.6%	0.001	0.033	4.1%	0.001	0.030	2.6%	0.001	0.033	4.6%
	0.2	0.000	0.006	4.6%	0.000	0.006	4.6%	0.000	0.005	4.7%	0.000	0.006	4.6%
	1.5	0.000	0.005	4.9%	0.000	0.005	4.6%	0.000	0.005	5.5%	0.000	0.005	4.9%
	0.6	0.000	0.005	3.9%	0.000	0.006	5.4%	0.000	0.005	4.5%	0.000	0.005	4.6%
	0.8	0.000	0.034	3.7%	−0.001	0.034	5.0%	0.000	0.030	3.8%	0.000	0.034	3.9%
	0.2	0.000	0.006	5.9%	0.000	0.006	6.1%	0.000	0.006	5.1%	0.000	0.006	5.9%
	1.5	0.000	0.006	5.5%	0.000	0.006	5.4%	0.000	0.005	4.8%	0.000	0.006	5.6%
	0.9	0.000	0.001	3.9%	0.000	0.001	5.9%	0.000	0.001	5.4%	0.000	0.001	5.0%
	0.8	−0.001	0.035	4.3%	−0.002	0.035	5.4%	0.000	0.031	5.8%	−0.001	0.034	5.0%
	0.2	0.000	0.006	3.8%	0.000	0.006	3.9%	0.000	0.005	4.4%	0.000	0.006	3.8%
	1.5	0.000	0.006	5.1%	0.000	0.006	5.2%	0.000	0.005	4.6%	0.000	0.006	5.5%

The weight matrix is a social interaction matrix $W_n = I_R \otimes [(\mathbf{1}_{m_j}\mathbf{1}'_{m_j} - I_{m_j})/(m_j - 1)]$, $m_j \sim$ IID U(3, 20), $j = 1, \cdots, R$. 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 1000 simulations. The exogenous regressors are $(1, x_{i1}, x_{i2})'$, $x_{i1} \sim$ IID N(3, 1), $x_{i2} \sim$ 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 $1/m_j$, $j = 1, \cdots, R$.

Table 4. Estimation of SAR with variance structure V2 and parameter configuration P2.

R	θ_0	MQML			GMM			GMM(Σ_n)			II		
		Bias	**RMSE**	$P(5\%)$	**Bias**	**RMSE**	$P(5\%)$	**Bias**	**RMSE**	$P(5\%)$	**Bias**	**RMSE**	$P(5\%)$
100	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%
200	0.2	−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%

The weight matrix is a social interaction matrix $W_n = I_R \otimes [(\mathbf{1}_{m_j}\mathbf{1}'_{m_j} - I_{m_j})/(m_j-1)]$, $m_j \sim$ IID U(3, 20), $j = 1, \cdots, R$. 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 1000 simulations. The exogenous regressors are $(1, x_{i1}, x_{i2})'$, $x_{i1} \sim$ IID N(3, 1), $x_{i2} \sim$ 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 $1/m_j$, $j = 1, \cdots, R$.

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 X_n is relatively more biased. Secondly, among the consistent estimators, the optimal robust GMM using the true Σ_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(Σ_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 $u_n'(G_n - \text{Dg}(G_n)u_n$ and $u_n'(G_n X_n \beta, X_n)$ 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 1000 simulated data sets under variance structure V1 and parameter configuration P1 with $R = 100$.[8] Recall that for a given $\theta_0 = (\lambda_0, \beta_0')'$ and the exogenous X_n, the data generating process generates the observable data y_n and the binding function $\lambda + y_n' S_n'(\lambda) M_n D_n(\lambda) M_n S_n(\lambda) y_n / y_n' W_n' M_n W_n y_n$ 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.
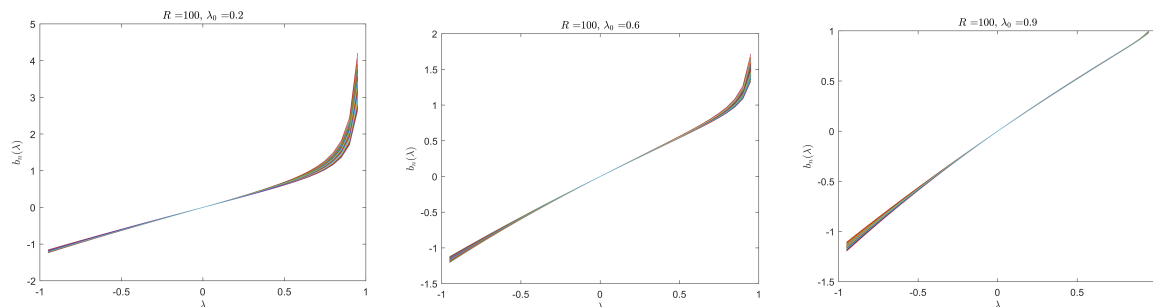
Figure 1. $b_n(\lambda)$ under variance structure V1 and parameter configuration P1, $R = 100$.

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 and 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 were simulated the same as before. From Tables 5 and 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.

---

8   Each sub-figure contains 1000 lines, one for each of the simulated data set.
9   The authors thank a referee for suggesting this comparison. Since one needs to concentrate out the scalar error variance instead of the nuisance matrix Σ_n, the II procedure needs to be modified, see Appendix B.6.

Table 5. Estimation of SAR under homoscedasticity with parameter configuration P1.

R	θ_0	QML			GMM			II		
		Bias	**RMSE**	$P(5\%)$	**Bias**	**RMSE**	$P(5\%)$	**Bias**	**RMSE**	$P(5\%)$
100	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%
200	0.2	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%

The weight matrix is a social interaction matrix $\boldsymbol{W}_n = \boldsymbol{I}_R \otimes [(\mathbf{1}_{m_j}\mathbf{1}'_{m_j} - \boldsymbol{I}_{m_j})/(m_j-1)]$, $m_j \sim$ IID U$(3,20)$, $j = 1, \cdots, R$. 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 1000 simulations. The exogenous regressors are $(1, x_{i1}, x_{i2})'$, $x_{i1} \sim$ IID N$(3,1)$, $x_{i2} \sim$ 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 standard normal distribution.

Table 6. Estimation of SAR under homoscedasticity with parameter configuration P2.

R	θ_0	QML			GMM			II		
		Bias	**RMSE**	**P(5%)**	**Bias**	**RMSE**	**P(5%)**	**Bias**	**RMSE**	**P(5%)**
100	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%
200	0.2	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%

The weight matrix is a social interaction matrix $\mathbf{W}_n = \mathbf{I}_R \otimes [(\mathbf{1}_{m_j}\mathbf{1}'_{m_j} - \mathbf{I}_{m_j})/(m_j - 1)]$, $m_j \sim$ IID U$(3, 20)$, $j = 1, \cdots, R$.
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 1000 simulations. The exogenous regressors are $(1, x_{i1}, x_{i2})'$, $x_{i1} \sim$ IID N$(3, 1)$, $x_{i2} \sim$ 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 standard normal distribution.

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 thoroughly examines 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 recentered 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.

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 $\boldsymbol{u}_n$. The OLS estimator itself is based on an incorrect moment condition, namely, exogeneity of $\boldsymbol{W}_n\boldsymbol{y}_n$. 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 from this approach. Rigorous treatments of these extensions are left for future studies.

10　Very 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 Cheng 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.

Appendix A. Lemmas

This appendix 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).

Lemma A1. *If $\{A_n\}$ and $\{B_n\}$ are uniformly bounded in row and column sums, then so are $\{A_n + B_n\}$ and $\{A_n B_n\}$.*

Lemma A2. *Suppose $\{A_n\}$ has its elements of order $O(h_n^{-1})$. If $\{B_n\}$ is uniformly bounded in column sum, then the elements of $A_n B_n$ are $O(h_n^{-1})$; if $\{B_n\}$ is uniformly bounded in row sum, then the elements of $B_n A_n$ are $O(h_n^{-1})$. In either case, $\mathrm{tr}(A_n B_n) = O(n/h_n)$.*

Lemma A3. *For a product involving (powers of) G_n and G'_n, $\prod_{l=1}^{m}(G_n^{i_1} G_n'^{i2})^{j_l}$, 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 A2, the elements of $G_n = W_n S_n^{-1}$ are $O(h_n^{-1})$ and $\mathrm{tr}(G_n) = O(n/h_n)$. From Lemma A1, $G_n = W_n S_n^{-1}$ is uniformly bounded in row and column sums. By successive using of Lemma A1, $G_n^{i_1-1}$ is uniformly bounded in row and column sums, and through Lemma A2, the elements of $G_n^{i_1} = G_n G_n^{i_1-1}$ are $O(h_n^{-1})$ and $\mathrm{tr}(G_n^{i_1}) = O(n/h_n)$. Similarly, such a claim applies to $G_n'^{i2}$, which is also uniformly bounded in row and column sums. Then $G_n^{i_1} G_n'^{i2}$ is uniformly bounded in row and column sums with its elements being $O(h_n^{-1})$ and $\mathrm{tr}(G_n^{i_1} G_n'^{i2}) = O(n/h_n)$. Proceeding similarly, one can see that the product $\prod_{l=1}^{m}(G_n^{i_1} G_n'^{i2})^{j_l}$ shares these properties too. □

Lemma A4. *For the sequence $\{u_n\}$ with the elements following Assumption 3, let A_n and B_n be nonrandom, then*

$$\mathrm{E}(u'_n A_n u_n) = \mathrm{tr}(\Sigma_n A_n), \tag{A.1}$$

$$\mathrm{E}(u_n u'_n A_n u_n) = \mathrm{dg}(\Sigma_n^{(3)} \odot A_n), \tag{A.2}$$

$$\mathrm{E}(u'_n A_n u_n u'_n B_n u_n) = \mathrm{tr}(\Sigma_n^{(4)} \odot A_n \odot B_n) + \mathrm{tr}(\Sigma_n A_n)\mathrm{tr}(\Sigma_n B_n)$$
$$+ \mathrm{tr}[\Sigma_n A_n \Sigma_n (B_n + B'_n)]. \tag{A.3}$$

Lemma A5. *Under Assumptions 1–5,*

$$u'_n G_n u_n = O_P(n/h_n), \quad u'_n M_n G_n u_n = O_P(n/h_n),$$
$$u'_n G'_n M_n G_n u_n = O_P(n/h_n), \quad \beta'_0 X'_n G'_n M_n u_n = O_P(\sqrt{n}),$$
$$\beta'_0 X'_n G'_n M_n G_n u_n = O_P(\sqrt{n}), \quad \beta'_0 X'_n G'_n M_n G_n X_n \beta_0 = O(n).$$

Proof. From (A.1), $\mathrm{E}(u'_n G_n u_n) = \mathrm{tr}(\Sigma_n G_n)$. Under Assumption 3, Σ_n is uniformly bounded, so $\mathrm{tr}(\Sigma_n G_n) \leq K\mathrm{tr}(G_n) = O(n/h_n)$ from Lemma A3. Lee (2004) shows that, under Assumption 5, $\lim_{n\to\infty} n^{-1}(X_n, G_n X_n \beta_0)'(X_n, G_n X_n \beta_0)$ is nonsingular if and only if both the limits of $n^{-1}X'_n X_n$ and $n^{-1}\beta'_0 X'_n G'_n M_n G_n X_n \beta_0$ are nonsingular, indicating that $X'_n X_n = O(n)$ and $\beta'_0 X'_n G'_n M_n G_n X_n \beta_0 = O(n)$. Also from Lee (2004), M_n is uniformly bounded in row and column sums, and then $\mathrm{E}(u'_n M_n G_n u_n) = \mathrm{tr}(\Sigma_n M_n G_n) \leq K\mathrm{tr}(M_n G_n) = O(n/h_n)$ from Lemmas A2 and A3. Similarly, one can show $\mathrm{E}(u'_n G'_n M_n G_n u_n) = O(n/h_n)$. As for $\beta'_0 X'_n G'_n M_n u_n$, its expectation is zero and its variance is given by $\beta'_0 X'_n G'_n M_n \Sigma_n M_n G_n X_n \beta_0$, which is bounded by $K\beta'_0 X'_n G'_n M_n G_n X_n \beta_0 = O(n)$.

Then it follows that $\beta'_0 X'_n G'_n M_n u_n = O_P(\sqrt{n})$. Similarly, $\beta'_0 X'_n G'_n M_n G_n u_n = O_P(\sqrt{n})$. Note that $G'_n M_n G_n G'_n M_n G_n$ is uniformly bounded in row and column sums through Lemmas A1 and A2. Given that the elements of X_n are uniformly bounded, one has $\beta'_0 X'_n G'_n M_n G_n G'_n M_n G_n X_n \beta_0 = O(n)$. Then it follows that $\beta'_0 X'_n G'_n M_n G_n u_n = O_P(\sqrt{n})$. $\square$

Lemma A6. *Under Assumptions 1–5,* $\mathrm{E}(r_n) = O(n/h_n)$, $\mathrm{Var}(r_n) = O(n)$, $\mathrm{E}(d_n) = O(n)$, *and* $\mathrm{Var}(d_n) = O(n)$. *When there is no* X, *then* $\mathrm{E}(r_{1n}) = O(n/h_n)$, $\mathrm{Var}(r_{1n}) = O(n/h_n)$, $\mathrm{E}(d_{1n}) = O(n/h_n)$, *and* $\mathrm{Var}(d_{1n}) = O(n/h_n)$.

Proof. Given Lemmas A4 and A5,

$$\mathrm{E}(r_n) = \mathrm{tr}(\Sigma_n M_n G_n) = O(n/h_n) \tag{A.4}$$

and

$$\mathrm{E}(d_n) = \mathrm{tr}(\Sigma_n G'_n M_n G_n) + \beta'_0 X'_n G'_n M_n G_n X_n \beta_0 = O(n) \tag{A.5}$$

are obvious. Using Lemma A4,

$$\begin{aligned}
\mathrm{Var}(r_n) &= \mathrm{Var}(u'_n M_n G_n u_n) + \mathrm{Var}(\beta'_0 X'_n G'_n M_n u_n) \\
&\quad + 2\mathrm{Cov}(u'_n M_n G_n u_n, \beta'_0 X'_n G'_n M_n u_n) \\
&= \mathrm{E}[(u'_n M_n G_n u_n)^2] - [\mathrm{E}(u'_n M_n G_n u_n)]^2 + \mathrm{E}(u'_n M_n G_n X_n \beta_0 \beta'_0 X'_n G'_n M_n u_n) \\
&\quad + 2\beta'_0 X'_n G'_n M_n \mathrm{E}(u_n u'_n M_n G_n u_n) \\
&= \mathrm{tr}(\Sigma_n^{(4)} \odot M_n G_n \odot M_n G_n) + \mathrm{tr}[\Sigma_n M_n G_n \Sigma_n (M_n G_n + G'_n M_n)] \\
&\quad + \beta'_0 X'_n G'_n M_n \Sigma_n M_n G_n X_n \beta_0 + 2\beta'_0 X'_n G'_n M_n \mathrm{dg}(\Sigma_n^{(3)} \odot M_n G_n),
\end{aligned} \tag{A.6}$$

where in view of Lemma A3 and A5, $\beta'_0 X'_n G'_n M_n \Sigma_n M_n G_n X_n \beta_0 = O(n)$ is the leading term. Similarly,

$$\begin{aligned}
\mathrm{Var}(d_n) &= \mathrm{E}[(u'_n G'_n M_n G_n u_n)^2] - [\mathrm{E}(u'_n G'_n M_n G_n u_n)]^2 \\
&\quad + 4(u'_n G'_n M_n G_n X_n \beta_0 \beta'_0 X'_n G'_n M_n G_n u_n) \\
&\quad + 4\beta'_0 X'_n G'_n M_n G_n \mathrm{E}(u_n u'_n G'_n M_n G_n u_n) \\
&= \mathrm{tr}(\Sigma_n^{(4)} \odot G'_n M_n G_n \odot G'_n M_n G_n) + 2\mathrm{tr}[\Sigma_n G'_n M_n G_n \Sigma_n G'_n M_n G_n] \\
&\quad + 4\beta'_0 X'_n G'_n M_n G_n \Sigma_n G'_n M_n G_n X_n \beta_0 \\
&\quad + 4\beta'_0 X'_n G'_n M_n G_n \mathrm{dg}(\Sigma_n^{(3)} \odot G'_n M_n G_n),
\end{aligned} \tag{A.7}$$

in which $\beta'_0 X'_n G'_n M_n G_n \Sigma_n G'_n M_n G_n X_n \beta_0 = O(n)$ is the leading term. For the case when there is no X, the results are obvious. $\square$

Lemma A7. *Suppose* $\{A_n\}$ *is a sequence of matrices with uniformly bounded row and column sums. Let* $\{b_n\}$ *be a sequence of constants with uniformly bounded elements and* $\sup_{n\to\infty} n^{-1} \sum_{i=1}^n |b_{i,n}|^{2+\eta_1} < \infty$ *for some* $\eta_1 > 0$. *For the sequence* $\{u_n\}$ *that satisfies Assumption 3, let* $Q_n = b'_n u_n + u'_n A_n u_n$. *Then*

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

Appendix B. Proofs

The proofs of Theorems 1–4 in Section 2 and Corollary 2 in Section 3 are provided in this appendix and those of Corollaries 1 and 3, which follow similarly, are skipped.

Appendix B.1. Proof of Theorem 1

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

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

where, in light of the proof of Lemma A6, $r_{1n} - \mathrm{E}(r_{1n}) = O_P(\sqrt{n/h_n})$, $r_{2n} = O(\sqrt{n})$, and $\mathrm{E}(d_n) = O(n)$. So when h_n diverges,

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

and one can apply Lemma A7 to $r_{2n} = \boldsymbol{\beta}_0' \boldsymbol{X}_n' \boldsymbol{G}_n' \boldsymbol{M}_n \boldsymbol{u}_n$; when h_n is bounded, one can apply Lemma A7 to $r_n = r_{1n} + r_{2n} = \boldsymbol{u}_n' \boldsymbol{M}_n \boldsymbol{G}_n \boldsymbol{u}_n + \boldsymbol{\beta}_0' \boldsymbol{X}_n' \boldsymbol{G}_n' \boldsymbol{M}_n \boldsymbol{u}_n$. From Lemma A6, one sees that $\operatorname{Var}(r_{2n}) = \operatorname{Var}(r_n) + o(n)$ when h_n diverges. Therefore, regardless of h_n,

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

□

Appendix B.2. Proof of Theorem 2

Proof. Note that

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

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

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

It follows from Lemma A7 that

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

where

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

by using Lemma A4 and the fact that $\mathrm{dg}(\boldsymbol{E}_n) = \boldsymbol{0}$. □

Appendix B.3. Proof of Theorem 3

Proof. 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 $s_n \to \infty$ and $s_n/\sqrt{n} \to 0$,

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

The first derivative of the binding function is

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

By substituting $\boldsymbol{y}_n = \boldsymbol{S}_n^{-1} \boldsymbol{X}_n \boldsymbol{\beta}_0 + \boldsymbol{S}_n^{-1} \boldsymbol{u}_n$ and using Lemmas A2–A4, 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

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

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. □

Appendix B.4. Proof of Theorem 4

Proof. Upon substitution,

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_n^{II} &= (\boldsymbol{X}'_n \boldsymbol{X}_n)^{-1} \boldsymbol{X}'_n \hat{\boldsymbol{S}}_n \boldsymbol{S}_n^{-1} \boldsymbol{X}_n \boldsymbol{\beta}_0 + (\boldsymbol{X}'_n \boldsymbol{X}_n)^{-1} \boldsymbol{X}'_n \hat{\boldsymbol{S}}_n \boldsymbol{S}_n^{-1} \boldsymbol{u}_n \\
&= (\boldsymbol{X}'_n \boldsymbol{X}_n)^{-1} \boldsymbol{X}'_n \left[\boldsymbol{S}_n - (\hat{\lambda}_n^{II} - \lambda_0) \boldsymbol{W}_n \right] \boldsymbol{S}_n^{-1} \boldsymbol{X}_n \boldsymbol{\beta}_0 \\
&\quad + (\boldsymbol{X}'_n \boldsymbol{X}_n)^{-1} \boldsymbol{X}'_n \left[\boldsymbol{S}_n - (\hat{\lambda}_n^{II} - \lambda_0) \boldsymbol{W}_n \right] \boldsymbol{S}_n^{-1} \boldsymbol{u}_n \\
&= \boldsymbol{\beta}_0 + (\boldsymbol{X}'_n \boldsymbol{X}_n)^{-1} \boldsymbol{X}'_n \boldsymbol{u}_n - (\hat{\lambda}_n^{II} - \lambda_0) \left[(\boldsymbol{X}'_n \boldsymbol{X}_n)^{-1} \boldsymbol{X}'_n \boldsymbol{G}_n \boldsymbol{X}_n \boldsymbol{\beta}_0 + (\boldsymbol{X}'_n \boldsymbol{X}_n)^{-1} \boldsymbol{X}'_n \boldsymbol{G}_n \boldsymbol{u}_n \right],
\end{aligned}
$$

where $\hat{\lambda}_n^{II} - \lambda_0 = O_p(n^{-1/2})$ (from Theorem 3). Further, $(X_n'X_n)^{-1}X_n'u_n = O_p(n^{-1/2})$, $(X_n'X_n)^{-1}X_n'G_nX_n\beta_0 = O(1)$, and $(X_n'X_n)^{-1}X_n'G_nu_n = O_p(n^{-1/2})$ (from Lemma A5). Note that if one expands $\hat{\lambda}_n^{II} - \lambda_0 = b_n^{-1}(\hat{\lambda}_n) - b_n^{-1}(b_n(\lambda_0))$,

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

From the proof of Theorem 2,

$$\hat{\lambda}_n - b_n(\lambda_0) = \frac{r_n - u_n'D_nu_n}{\mathrm{E}(d_n)} + o_P(n^{-1/2}) = \frac{u_n'E_nu_n + \beta_0'X_n'G_n'M_nu_n}{\mathrm{E}(d_n)} + o_P(n^{-1/2}). \tag{A.12}$$

Given the above results, one has

$$\begin{aligned}
&\sqrt{n}(\hat{\beta}_n^{II} - \beta_0)\\
&= \sqrt{n}(X_n'X_n)^{-1}X_n'u_n - \sqrt{n}(\hat{\lambda}_n^{II} - \lambda_0)(X_n'X_n)^{-1}X_n'G_nX_n\beta_0 + o_P(1)\\
&= \sqrt{n}(X_n'X_n)^{-1}X_n'u_n - \left[\frac{\sqrt{n}\,(\hat{\lambda}_n - b_n(\lambda_0))}{b_n'(\lambda_0)}\right](X_n'X_n)^{-1}X_n'G_nX_n\beta_0 + o_P(1)\\
&= \sqrt{n}(X_n'X_n)^{-1}X_n'u_n - \frac{\sqrt{n}(X_n'X_n)^{-1}X_n'G_nX_n\beta_0}{b_n'(\lambda_0)}\frac{u_n'E_nu_n + \beta_0'X_n'G_n'M_nu_n}{\mathrm{E}(d_n)} + o_P(1). \tag{A.13}
\end{aligned}$$

One can check that Lemma A7 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,

$$\begin{aligned}
V = \lim_{n\to\infty} \Big\{ &n(X_n'X_n)^{-1}X_n'\Sigma_nX_n(X_n'X_n)^{-1}\\
&+ \frac{\eta}{b_0^2}(X_n'X_n)^{-1}X_n'G_nX_n\beta_0\beta_0'X_n'G_n'X_n(X_n'X_n)^{-1}\\
&- \frac{n(X_n'X_n)^{-1}X_n'G_nX_n\beta_0\beta_0'X_n'G_n'M_n\Sigma_nX_n(X_n'X_n)^{-1}}{b_0[\mathrm{tr}(\Sigma_nG_n'M_nG_n) + \beta_0'X_n'G_n'M_nG_nX_n\beta_0]}\\
&- \frac{n(X_n'X_n)^{-1}X_n'\Sigma_nM_nG_nX_n\beta_0\beta_0'X_n'G_n'X_n(X_n'X_n)^{-1}}{b_0[\mathrm{tr}(\Sigma_nG_n'M_nG_n) + \beta_0'X_n'G_n'M_nG_nX_n\beta_0]} \Big\}. \tag{A.14}
\end{aligned}$$

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

$$\gamma = \lim_{n\to\infty}\left[\frac{n}{b_0\mathrm{E}(d_n)}(X_n'X_n)^{-1}X_n'\Sigma_nM_nG_nX_n\beta_0 - \frac{\eta}{b_0^2}(X_n'X_n)^{-1}X_n'G_nX_n\beta_0\right]. \tag{A.15}$$

□

Appendix B.5. Proof of Corollary 2

Proof. By substitution and using the Nagar (1959) expansion,

$$\sqrt{\frac{n}{h_n}}\left(\hat{\lambda}_n - \lambda_0 - \frac{y_n' S_n' D_n S_n y_n}{y_n' W_n' W_n y_n}\right) = \sqrt{\frac{n}{h_n}}\left(\frac{r_{1n}}{d_{1n}} - \frac{u_n' D_n u_n}{d_{1n}}\right)$$

$$= \sqrt{\frac{n}{h_n}}\left(\frac{u_n' E_n u_n}{d_{1n}}\right)$$

$$= \sqrt{\frac{n}{h_n}}\frac{u_n' E_n u_n}{\mathrm{E}(d_{1n})} + o_P(1),$$

where $\mathrm{E}(u_n' E_n u_n) = 0$, $\mathrm{Var}(u_n' E_n u_n) = \mathrm{tr}[\Sigma_n E_n \Sigma_n (E_n + E_n')] = O(n/h_n)$, and and the asymptotic distribution follows when one applies Lemma A7 to the quadratic form $u_n' E_n u_n$. □

Appendix B.6. The Case of Homoscedastic Error Term

If $\Sigma_n = \sigma_0^2 I_n$ (and further $\Sigma_n^{(j)} = \mu_j I_n$, $\mu_j = \mathrm{E}(u_{i,n}^j)$, $j = 3, 4$), then the recentering term in (7) becomes $\sigma_0^2 \mathrm{tr}(M_n G_n)/y_n' W_n' M_n W_n y_n$. To make the II procedure feasible, one may replace $\sigma_0^2 \mathrm{tr}(M_n G_n)/y_n' W_n' M_n W_n y_n$ by $n^{-1} y_n' S_n' M_n S_n y_n \mathrm{tr}(M_n G_n)/y_n' W_n' M_n W_n y_n$. Similar to the proof of Theorem 2, one can put

$$\sqrt{n}\left(\hat{\lambda}_n - \lambda_0 - \frac{n^{-1} y_n' S_n' M_n S_n y_n \mathrm{tr}(M_n G_n)}{y_n' W_n' M_n W_n y_n}\right) = \sqrt{n}\left(\frac{r_n^* - \mathrm{E}(r_n^*)}{\mathrm{E}(d_n)}\right) + o_P(1) \xrightarrow{d} \mathrm{N}(0, \omega),$$

where $\omega = \lim_{n\to\infty} n\mathrm{Var}(r_n^*)/\{[\mathrm{E}(d_n)]^2\}$, $r_n^* = u_n' M_n G_n^* u_n + \beta_0' X_n' G_n' M_n u_n$, $G_n^* = G_n - n^{-1}\mathrm{tr}(M_n G_n)$. In particular,

$$\mathrm{Var}(r_n^*) = \mu_4 \mathrm{tr}(M_n G_n^* \odot M_n G_n^*) + \sigma_0^2 \mathrm{tr}[M_n G_n^*(M_n G_n^* + G_n^{*\prime} M_n)]$$
$$+ \sigma_0^2 \beta_0' X_n' G_n' M_n M_n G_n X_n \beta_0 + 2\mu_3 \beta_0' X_n' G_n' M_n \mathrm{dg}(M_n G_n^*).$$

Define the binding function as $b_n(\lambda) = \lambda + n^{-1} y_n' S_n'(\lambda) M_n S_n(\lambda) y_n \mathrm{tr}(M_n G_n(\lambda))/y_n' W_n' M_n W_n y_n$ and $b_n'(\lambda_0) = 1 + (n y_n' W_n' M_n W_n y_n)^{-1}[y_n' S_n' M_n S_n y_n \mathrm{tr}(M_n G_n^2) - 2 y_n' W_n' M_n S_n y_n \mathrm{tr}(M_n G_n)]$. Assume $b_n'(\lambda_0) \xrightarrow{a.s.} b_0 \neq 0$. The asymptotic distribution of $(\hat{\lambda}_n^{II}, \hat{\beta}_n^{II\prime})'$ resulting from this binding function follows similarly from the proofs of Theorems 3 and 4 , given by

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

where

$$V = \lim_{n\to\infty}\left\{ n\sigma_0^2(X_n'X_n)^{-1} + \frac{\omega}{b_0^2}(X_n'X_n)^{-1}X_n'G_nX_n\beta_0\beta_0'X_n'G_n'X_n(X_n'X_n)^{-1} \right.$$

$$- \frac{n(X_n'X_n)^{-1}X_n'G_nX_n\beta_0[\sigma_0^2\beta_0'X_n'G_n'M_n + \mu_3\mathrm{dg}(M_nG_n^*)']X_n(X_n'X_n)^{-1}}{b_0[\sigma_0^2\mathrm{tr}(G_n'M_nG_n) + \beta_0'X_n'G_n'M_nG_nX_n\beta_0]}$$

$$\left. - \frac{n(X_n'X_n)^{-1}X_n'[\sigma_0^2M_nG_nX_n\beta_0 + \mu_3\mathrm{dg}(M_nG_n^*)]\beta_0'X_n'G_n'X_n(X_n'X_n)^{-1}}{b_0[\sigma_0^2\mathrm{tr}(G_n'M_nG_n) + \beta_0'X_n'G_n'M_nG_nX_n\beta_0]} \right\}$$

and

$$\gamma = \lim_{n\to\infty}\left\{\frac{n}{b_0\mathrm{E}(d_n)}(X_n'X_n)^{-1}X_n'[\sigma_0^2M_nG_nX_n\beta_0 + \mu_3\mathrm{dg}(M_nG_n^*)] - \frac{\omega}{b_0^2}(X_n'X_n)^{-1}X_n'G_nX_n\beta_0\right\}.$$

References

Badinger, Harald, and Peter Egger. 2011. Estimation of higher-order spatial autoregressive cross-section models with heteroscedastic disturbances. *Papers in Regional Science* 90: 213–35. [CrossRef]

Case, Anne C. 1991. Spatial patterns in household demand. *Econometrica* 59: 953–65. [CrossRef]

Cheng, Xu, Zhipeng Liao, and Ruoyao Shi. 2019. On uniform asymptotic risk of averaging GMM estimators. *Quantitative Economics* 3: 931–97. [CrossRef]

Cliff, Andrew David, and J. Keith Ord. 1981. *Spatial Processes: Models and Applications*. London: Pion Ltd.

Debarsy, Nicolas, Fei Jin, and Lung-Fei Lee. 2015. Large sample properties of the matrix exponential spatial specification with an applicationto FDI. *Journal of Econometrics* 188: 1–21. [CrossRef]

Gospodinov, Nikolay, Ivana Komunjer, and Serena Ng. 2017. Simulated minimum distance estimation of dynamic models with errors-in-variables. *Journal of Econometrics* 200: 181–93. [CrossRef]

Gouriéroux, Christian, Alain Monfort, and Eric Renault. 1993. Indirect inference. *Journal of Applied Econometrics* 8: S85–S118. [CrossRef]

Jin, Fei, and Lung-Fei Lee. 2019. GEL estimation and tests of spatial autoregressive models. *Journal of Econometrics* 208: 585–612. [CrossRef]

Kelejian, Harry H., and Ingmar R. Prucha. 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40: 509–33. [CrossRef]

Kelejian, Harry H., and Ingmar R. Prucha. 2001. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104: 219–57. [CrossRef]

Kelejian, Harry H., and Ingmar R. Prucha. 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157: 53–67. [CrossRef]

Kyriacou, Maria, Peter C. B. Phillips, and Francesca Rossi. 2017. Indirect inference in spatial autoregression. *The Econometrics Journal* 20: 168–89. [CrossRef]

Lam, Clifford, and Pedro C. L. Souza. 2019. Estimation and selection of spatial weight matrix in a spatial lag model. *Journal of Business and Economic Statistics* 3: 693–710. [CrossRef]

Lee, Lung-Fei. 2001. *Asymptotic Distribution of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models I: Spatial Autoregressive Processes*. Working Paper. Columbus: Department of Econommics, Ohio State University.

Lee, Lung-Fei. 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models *Econometric Theory* 18: 252–77. [CrossRef]

Lee, Lung-Fei. 2004. Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72: 1899–925. [CrossRef]

Lee, Lung-Fei. 2007. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137: 489–514. [CrossRef]

Lin, Xu, and Lung-Fei Lee. 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics* 157: 34–52. [CrossRef]

Liu, Shew Fan, and Zhenlin Yang. 2015. Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and nonnormality. *Regional Science and Urban Economics* 52: 50–70. [CrossRef]

Nagar, Anirudh L. 1959. The bias and moment matrix of the general *k*-class estimators of the parameters in simultaneous equations. *Econometrica* 27: 575–95. [CrossRef]

Phillips, Peter C. B. 2012. Folklore theorems, implicit maps, and indirect inference. *Econometrica* 80: 425–54.

Robinson, Peter M. 2008. Correlation testing in time series, spatial and cross-sectional data. *Journal of Econometrics* 147: 5–16. [CrossRef]

Smith, Anthony A., Jr. 1993. Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics* 8: S63–S84. [CrossRef]

Zhang, Xinyu, and Jihai Yu. 2018. Spatial weights matrix election and model averaging for spatial autoregressive models. *Journal of Econometrics* 203: 1–18. [CrossRef]