Indirect Inference Estimation of Higher-order Spatial Autoregressive Models*

Yong Bao

Department of Economics
Purdue University
E-mail: ybao@purdue.edu

Abstract

This paper proposes estimating parameters in higher-order spatial autoregressive models, where the error term also follows a spatial autoregression and its innovations are heteroskedastic, by matching the simple ordinary least squares estimator with its analytical approximate expectation, following the principle of indirect inference. The resulting estimator is shown to be consistent, asymptotically normal, simulation-free, and robust to unknown heteroskedasticity. Monte Carlo simulations demonstrate its good finite-sample properties in comparison with existing estimators. An empirical study of Airbnb rental prices in the city of Asheville illustrates that the structure of spatial correlation and effects of various factors at the early stage of the COVID-19 pandemic are quite different from those during the second summer. Notably, during the pandemic, safety is valued more and on-line reviews are valued much less.

Key Words: spatial autoregressive models; indirect inference; ordinary least squares; heteroskedasticity

JEL classification: C21, C31

*The author is grateful to Essie Maasoumi (the editor), an associate editor, and two anonymous referees for their constructive feedback. The author also thanks Xiaoyan Zhou for his careful reading of an early version of this paper.

1. Introduction

Correlation in space has attracted lots of attention in various disciplines of the social sciences. It arises naturally due to aggregation, competition, copycatting, externality, capacity constraint, among others. The spatial autoregressive (SAR) model and its various extensions have been used in many areas, including economics and finance, real estate, criminology, political science, and sociology. In terms of estimation and inference, there are mainly two approaches in the literature. One is based on the principle of (quasi) maximum likelihood ((Q)ML), see, for example, Anselin (1988) and Lee (2004). The other is based on the principle of instrumental variables (IVs) or moment conditions, see, *inter alia*, Kelejian and Prucha (1998, 2010), Lee and Liu (2010), Lin and Lee (2010), and Jin and Lee (2019).

The (Q)ML approach typically relies on some stringent assumptions on the data generating process (DGP). Under error heteroskedasticity, the (Q)ML estimator is inconsistent. Instead, the generalized spatial 2-stage least squares (GS2SLS) estimator of Kelejian and Prucha (1998, 2010), also see Badinger and Egger (2011), and the generalized method of moments (GMM) estimator of Jin and Lee (2019) can be used, which are robust to unknown heteroskedasticity.

Very recently, a third approach emerged in the literature, started by Kyriacou et al. (2017) that focused on the pure first-order spatial autoregressive (SAR(1)) model. The idea is to derive the analytical approximate expectation of the simple and yet inconsistent ordinary least squares (OLS) estimator of the SAR parameter. Note that the analytical expression involves the true parameter. Then one can solve for the SAR parameter by matching the OLS estimator with its analytical expectation. Kyriacou et al. (2017) named their approach the indirect inference (II), even though they used analytical approximation instead of simulations as in the original works of Gouriéroux et al. (1993) and Smith Jr. (1993). This new approach is straightforward to implement, simulation-free, and does not rely on the choice of IV or moment conditions associated with the error term. Kyriacou et al. (2021) extended their earlier work on SAR(1) to include exogenous regressors and allow for error heteroskedasticity. The analytical approximation in Kyriacou et al. (2017, 2021) is based on the approximate moment result of Lieberman (1994) on a ratio of quadratic forms by taking expectations of both the numerator and denominator of the ratio that defines the OLS estimator of the SAR parameter. Bao et al. (2020) also considered the SAR(1) model with exogenous regressors and possible heteroskedastic errors, but they approximated the ratio by taking expectation of the numerator only.[1] The latter approximation makes it much easier to derive the asymptotic distribution of the recentered OLS estimator and the resulting II estimator.

Note that Kyriacou et al. (2017, 2021) and Bao et al. (2020) focused on SAR(1), where the error process does not have spatial correlation. In this case, one needs to solve for a scalar binding function, whether random or deterministic, to estimate the SAR parameter. Kyriacou et al. (2021) suggested possible extensions to higher-order SAR or a SAR with spatially correlated errors by adding more II conditions via suitable binding functions, but fell short of mentioning how to construct these additional binding functions. Bao and Liu (2021) took this endeavor to extend the II approach to a spatial autoregressive model with autoregressive error (SARAR), where the orders of spatial autoregressions in the outcome and error equations are both 1, namely, SARAR(1,1). They recognized that the OLS estimator of the SAR parameter in the outcome equation has (approximate) expectation that depends on, in addition to itself, the SAR parameter in the error process and thus a single sample binding function is not enough for one to solve for two unknown SAR parameters. Bao and Liu (2021) tackled this difficulty by designing a second sample binding function that is based on the OLS estimator of the SAR parameter in the error process using a consistent residual vector. So in the end, given the sample data, they used a system of two equations to solve for two unknown SAR parameters.

In practice, one may be faced with many choices of spatial weight matrices that characterize cross-sectional correlations possibly arising from geographic, social, economic, and demographic factors. For example, in studying housing prices in 377 metropolitan statistical areas (MSAs) in the US, Yang (2021) tried with distance measures based on geographic distance, migration flow, and bilateral house price correlation in a SAR(1) framework. In principle, she could have used a higher-order model to include three spatial weight matrices based on the three distance measures. Dall'erba et al. (2009) studied productivity growth in service industry in European regions. Recognizing that economic interactions decrease very substantially when a national border is passed, they included two spatial weight matrices, one based on nearest within-nation neighbors and the other based on nearest neighbors in the bordering nations. On many occasions, even if a single measure of spatial correlation is used, it may still be an open question what threshold value of the measure should be used to define neighbors and thus higher-order models,

[1] The function in Bao et al. (2020) that is used to solve for the SAR parameter is thus a *sample* binding function, which is random. Its probability limit is known as the binding function, which is used in Kyriacou et al. (2017, 2021).

based on different threshold values, need to be experimented with. A SAR(1) or SARAR(1,1) specification may fail to capture the full spectrum of cross-sectional correlation. For instance, instead of using a coarse 10-mile distance ring, Gupta and Robinson (2015) reduced the radius of distance rings to 1 mile and included in total ten 1-mile distance rings in studying the level of venture capital funding in the US biotechnology industry. After trying with a higher-order model with ten spatial lags, Gupta and Robinson (2015) suggested that only the first two spatial lags matter and thus spatial dependence is restricted to a smaller radius. Similarly, in the example of Airbnb prices in Section 5, a SAR(1) specification using a spatial weight matrix constructed from the 20-nearest neighbors would be strongly rejected based on a Moran-type statistic (to be introduced in Section 3) for the 2020 data. Spatial correlation spreads as far as 100-nearest neighbors. A lump-sum spatial weight matrix using 100-nearest neighbors might give rise to a simple SAR(1), but it fails to reveal the different degrees of correlation arising from neighbors of different distances. Using finer neighborhood rings, Section 5 shows that closer neighbors have higher impact than neighbors that are farther away. Higher-order models are more appropriate and appealing to accommodate this kind of correlation pattern.

This paper aims to fill this gap by studying the II method of estimation for the more general SARAR(p, q), where the finite integers p and q denote the orders of spatial autoregressions in the outcome and error equations, respectively. Since there are multiple parameters, one may be tempted to write the OLS estimator (of the parameter vector in the outcome equation) using the familiar matrix inverse formula. However, a generalization of Lieberman (1994) to the multivariate case is not available and thus one may not be able to employ the approach of Kyriacou et al. (2017, 2021) to come up with a system of deterministic binding functions. It seems more appealing to use instead the approach based on random sample binding functions as in Bao et al. (2020) and Bao and Liu (2021) that takes the inverse part in the OLS formula as given and evaluates only the expectation of the "numerator" part. Still, as in the SARAR(1,1) case, the OLS estimator of the p SAR parameters in the outcome equation depends on, in addition to themselves, the q SAR parameters in the error process. Similarly, the OLS estimator of the q SAR parameters in the error process depends on both sets of SAR parameters. Therefore, one needs to find a practicable way of expressing these relationships as a system of $p + q$ equations given the observable sample data. In the SARAR(1,1) case, each sample binding function is related to an approximate expectation of a scalar estimator. For the general SARAR(p, q), seeking a scalar expression and thus a corresponding approximate expectation for each element from the OLS matrix formula, even if one uses only the "numerator" part, is not an easy extension, since in general these elements are correlated with and nonlinear functions of each other and taking approximate expectation one by one is extremely difficult, if not impossible. If one takes a different route to approximate the whole matrix, it can become very complicated as the approximation will eventually involve matrix inverses. So even though conceptually extending the II approach from SARAR(1,1) to SARAR(p, q) may be straightforward, its practical implementation is not.

The strategy in this paper is to take a conditional approach by working out analytical approximate expectation of each element of the OLS estimator of the SAR parameters in the outcome equation conditional on all the other SAR parameters in the model, whereas for each element of the OLS estimator, based on a properly defined residual vector, of the parameters in the error process, its analytical approximate expectation is conditional on all the SAR parameters. Thus, one arrives at a system of $p+q$ sample binding functions that can be used to solve for the two sets of SAR parameters. Once the SAR parameters are consistently estimated, estimation of the parameter vector associated with the exogenous regressors follows from the usual OLS procedure.

As in the set-up of Bao and Liu (2021), the error innovations in this paper are possibly heteroskedastic. The motivation of using II approach again is to avoid the choice of IV and/or moment conditions. It may also possess some degree of numerical advantage relative to the GMM approach of Jin and Lee (2019) since it searches for parameters over a $(p+q)$-dimensional parameter space, whereas the GMM approach searches over a $(p+q+k_x)$-dimensional parameter space, where k_x is the number of exogenous regressors.[2] The GS2SLS approach of Kelejian and Prucha (2010) estimates the outcome equation parameters first by 2SLS, totally ignoring the correlation in the error process, and then using the resulting residuals to estimate the SAR parameters in the error process by GMM. In contrast, the II approach in this paper estimates first both sets of SAR parameters jointly and then estimates the remaining parameters associated with exogenous regressors in the outcome equation. Lee (2007) and Yang (2015) emphasized that the spatial coefficients are the main source of bias in model estimation and the main cause of difficulty in bias correction in SAR models. One would expect that ignoring the spatial correlation in the error process when estimating parameters in the outcome equation by the GS2SLS approach could lead to some undesirable results. If one brushes aside the complexity of the optimal weight

[2]Examples of high-dimensional parameter space include, among others, Helbich et al. (2014) and Rico-Juan and Taltavull de La Paz (2021), where 23 and 47 (non-constant) covariates are used respectively in their hedonic pricing models.

matrices in the GMM objective functions in Jin and Lee (2019) and (the second step of) Kelejian and Prucha (2010), the numerical cost, in terms of parameter dimension, of the II procedure lies between those of GMM and GS2SLS. This paper also offers insights into how the sample binding functions are related to the best moment conditions of the GMM estimator as discussed in Liu et al. (2010).

This paper is organized as follows. The next section introduces notation and assumptions on SARAR(p, q) models. Section 3 provides the main results regarding the II estimator. In particular, it shows that the II estimator is consistent, asymptotically normal, and robust to heteroskedasticity in error innovations. Section 4 contains Monte Carlo results by comparing the II estimator with the GMM and GS2SLS estimators. Section 5 provides an empirical study of Airbnb rental prices using two data sets from the city of Asheville, North Carolina. It is found that the correlation structure and marginal effects of various factors at the early stage of the COVID-19 pandemic are very different from those during the second summer. The last section concludes. Technical details and proofs are collected in the Appendix. Additional results are contained in a supplementary appendix.

2. Model Specification

Consider the following SARAR(p, q) model

$$\boldsymbol{y}_n = \boldsymbol{X}_n\boldsymbol{\beta} + \sum_{i=1}^{p} \lambda_i \boldsymbol{W}_{in}\boldsymbol{y}_n + \boldsymbol{u}_n, \quad \boldsymbol{u}_n = \sum_{j=1}^{q} \rho_j \boldsymbol{M}_{jn}\boldsymbol{u}_n + \boldsymbol{v}_n, \tag{1}$$

where n is the sample size, $\boldsymbol{y}_n$ is an $n \times 1$ vector of observations on the outcome variable, $\boldsymbol{X}_n$ is an $n \times k_x$ matrix of observations on k_x exogenous deterministic regressors with coefficient vector $\boldsymbol{\beta}$, $\boldsymbol{W}_{in}$ and $\boldsymbol{M}_{jn}$ are $n \times n$ nonstochastic spatial weight matrices, λ_i and ρ_j are scalar spatial dependence parameters, and $\boldsymbol{v}_n$ is an $n \times 1$ vector of innovation terms.

2.1. Notation

To ease notation burden, the subscript n is dropped (in $\boldsymbol{y}_n$, $\boldsymbol{W}_{in}$, $\boldsymbol{M}_{jn}$, $\boldsymbol{v}_n$, and other terms to follow) with the understanding that, for example, $\boldsymbol{y} = \boldsymbol{y}_n$ in fact denotes a triangular array. lim, $\rightarrow$, and sup are with respect to n going to infinity. $\overset{a.s.}{\rightarrow}$ and $\overset{d}{\rightarrow}$ denote convergences almost surely and in distribution, respectively. Matrix/vector dimension subscripts are also dropped when the dimensions can be read from the context. A parameter with subscript 0 is used to signify the parameter's true value. When a matrix/vector is presented without its argument(s), it means that it is evaluated at the true parameter value(s). (Sometimes the argument(s) may be added explicitly to emphasize the dependency.) For a vector $\boldsymbol{a}$ and a square matrix $\boldsymbol{A}$, $\mathrm{Dg}(\boldsymbol{a})$ denotes a square diagonal matrix with the vector $\boldsymbol{a}$ spanning the main diagonal, $\mathrm{dg}(\boldsymbol{A})$ is a column vector that collects in order the diagonal elements of $\boldsymbol{A}$, $\mathrm{Dg}(\boldsymbol{A}) = \mathrm{Dg}(\mathrm{dg}(\boldsymbol{A}))$, and $\boldsymbol{A} \succ 0$ denotes $\boldsymbol{A}$ being positive definite. tr and $\odot$ denote matrix trace and Hadamard (element by element) product operators, respectively. For any matrix $\boldsymbol{A}$, $||\boldsymbol{A}||_\infty$ and $||\boldsymbol{A}||_1$ denote maximum absolute row sum norm and maximum absolute column sum norm, respectively, and $\boldsymbol{A}^\star = \boldsymbol{A} + \boldsymbol{A}'$. When there is no confusion, a typical element of $\boldsymbol{A}$ (or $\boldsymbol{a}$) is denoted by a_{ij} (or a_i). Throughout, $\boldsymbol{\lambda} = (\lambda_1, \cdots, \lambda_p)'$, $\boldsymbol{\lambda}_{(-i)} = (\lambda_1, \cdots, \lambda_{i-1}, \lambda_{i+1}, \cdots, \lambda_p)'$, $\boldsymbol{\rho} = (\rho_1, \cdots, \rho_q)'$, $\boldsymbol{\rho}_{(-j)} = (\rho_1, \cdots, \rho_{j-1}, \rho_{j+1}, \cdots, \rho_q)'$, $\boldsymbol{\gamma} = (\boldsymbol{\lambda}', \boldsymbol{\rho}')'$, and

$$\begin{aligned}
&\boldsymbol{S}(\boldsymbol{\lambda}) = \boldsymbol{I} - \sum_{i=1}^{p} \lambda_i \boldsymbol{W}_i, \quad \boldsymbol{S}_{(-i)}(\boldsymbol{\lambda}) = \boldsymbol{S}(\boldsymbol{\lambda}_{(-i)}) = \boldsymbol{S}(\boldsymbol{\lambda}) + \lambda_i \boldsymbol{W}_i, \quad \boldsymbol{G}_i(\boldsymbol{\lambda}) = \boldsymbol{W}_i \boldsymbol{S}^{-1}(\boldsymbol{\lambda}),\\
&\boldsymbol{R}(\boldsymbol{\rho}) = \boldsymbol{I} - \sum_{j=1}^{q} \rho_j \boldsymbol{M}_j, \quad \boldsymbol{R}_{(-j)}(\boldsymbol{\rho}) = \boldsymbol{R}(\boldsymbol{\rho}_{(-j)}) = \boldsymbol{R}(\boldsymbol{\rho}) + \rho_j \boldsymbol{M}_j, \quad \boldsymbol{F}_j(\boldsymbol{\rho}) = \boldsymbol{M}_j \boldsymbol{R}^{-1}(\boldsymbol{\rho}),\\
&\boldsymbol{H}(\boldsymbol{\rho}) = \boldsymbol{I} - \boldsymbol{R}(\boldsymbol{\rho})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}'(\boldsymbol{\rho})\boldsymbol{R}(\boldsymbol{\rho})\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'(\boldsymbol{\rho}),\\
&\boldsymbol{D}_i(\gamma) = \mathrm{Dg}[\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{R}(\boldsymbol{\rho})\boldsymbol{G}_i(\boldsymbol{\lambda})\boldsymbol{R}^{-1}(\boldsymbol{\rho})], \quad \boldsymbol{E}_i(\gamma) = \boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{R}(\boldsymbol{\rho})\boldsymbol{G}_i(\boldsymbol{\lambda})\boldsymbol{R}^{-1}(\boldsymbol{\rho}) - \boldsymbol{D}_i(\gamma),\\
&\boldsymbol{K}_j(\boldsymbol{\rho}) = \mathrm{Dg}[\boldsymbol{F}_j(\boldsymbol{\rho})], \quad \boldsymbol{L}_j(\boldsymbol{\rho}) = \boldsymbol{F}_j(\boldsymbol{\rho}) - \boldsymbol{K}_j(\boldsymbol{\rho}),
\end{aligned}$$

where $\boldsymbol{I}$ is the identity matrix. The parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\gamma}', \boldsymbol{\beta}')'$ and the total number of parameters to be estimated is $k = p + q + k_x$.

2.2. Assumptions

The following assumptions are made throughout.

Assumption 1 *$\forall i = 1, \cdots, p$ and $\forall j = 1, \cdots, q$, (i) $||\boldsymbol{W}_i||_1 < \infty$, $||\boldsymbol{W}_i||_\infty < \infty$, $||\boldsymbol{M}_j||_1 < \infty$, and $||\boldsymbol{M}_j||_\infty < \infty$; (ii) the diagonal elements of $\boldsymbol{W}_i$ and $\boldsymbol{M}_j$ are all zero.*

Assumption 2 *(i) $\boldsymbol{S}^{-1}$ and $\boldsymbol{R}^{-1}$ exist. (ii) $||\boldsymbol{S}^{-1}||_1 < \infty$, $||\boldsymbol{S}^{-1}||_\infty < \infty$, $||\boldsymbol{R}^{-1}||_1 < \infty$, and $||\boldsymbol{R}^{-1}||_\infty < \infty$.*

Assumption 3 *The innovation terms are independent of each other with $\mathrm{E}(\boldsymbol{v}) = \boldsymbol{0}$, $\mathrm{Var}(\boldsymbol{v}) = \boldsymbol{\Sigma} = \mathrm{Dg}(\sigma_1^2, \cdots, \sigma_n^2)$, and $\mathrm{E}(|v_i|^{4+\eta}) < \infty$, $i = 1, \cdots, n$, for some positive constant η.*

Assumption 4 *(i)* $\boldsymbol{\gamma}_0$ *is contained in a compact parameter space* $\boldsymbol{\Gamma}$*. (ii) For any* $\boldsymbol{\gamma} = (\boldsymbol{\lambda}', \boldsymbol{\rho}')' \in \boldsymbol{\Gamma}$ *in a neighborhood of* $\boldsymbol{\gamma}_0$*,* $||\boldsymbol{S}^{-1}(\boldsymbol{\lambda})||_1 < \infty$*,* $||\boldsymbol{S}^{-1}(\boldsymbol{\lambda})||_\infty < \infty$*,* $||\boldsymbol{R}^{-1}(\boldsymbol{\rho})||_1 < \infty$*, and* $||\boldsymbol{R}^{-1}(\boldsymbol{\rho})||_\infty < \infty$*.*

Assumption 5 *(i) All the elements of* $\boldsymbol{X}$ *are uniformly bounded. (ii) The following limits exist and are non-singular/non-zero:* $\lim n^{-1}\boldsymbol{X}'\boldsymbol{X}$*,* $\lim n^{-1}\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{X}\boldsymbol{\beta}_0$ *(when* $\boldsymbol{\beta}_0 \neq \boldsymbol{0}$*),* $i = 1, \cdots, p$*, and* $\lim n^{-1}\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{F}_j'\boldsymbol{F}_j\boldsymbol{H}\boldsymbol{R}\boldsymbol{X}\boldsymbol{\beta}_0$ *(when* $\boldsymbol{\beta}_0 \neq \boldsymbol{0}$*),* $j = 1, \cdots, q$*.*

Assumption 1.(ii) can be regarded as a normalization rule and it excludes the so-called "self influence." Assumption 1.(i) and Assumption 2 are originated by Kelejian and Prucha (1998) that aim to limit the degree of spatial dependence. A sufficient condition for $\boldsymbol{S}^{-1}$ to be bounded in $|| \cdot ||_\infty$ in Assumption 2 is $\sum_{i=1}^p |\lambda_{0i}| < 1/\max_i(||\boldsymbol{W}_i||_\infty)$, so for row-normalized $\boldsymbol{W}$ matrices, this amounts to $\sum_{i=1}^p |\lambda_{0i}| < 1$, see Lee and Liu (2010) and Badinger and Egger (2011).[3] Assumptions 1 and 2 together imply that $\boldsymbol{S}^{-1}(\boldsymbol{\lambda})$ and $\boldsymbol{R}^{-1}(\boldsymbol{\rho})$ have bounded norms in a neighborhood of $\boldsymbol{\gamma}_0$, see Lee (2004). It is listed explicitly as Assumption 4.(ii) for convenience. Assumption 3 is the same as in Kelejian and Prucha (2010) and Jin and Lee (2019), allowing the independent error innovations to be heteroskedastic. When $q = 0$, the first two limits in Assumption 5.(ii) are similar to those in Lee (2002) that are related to an identification condition for estimation in the least squares and IV frameworks and rule out possible multicollinearities among $\boldsymbol{X}$ and $\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{X}\boldsymbol{\beta}_0$ for large n. When $q \neq 0$, an additional condition (the third limit in Assumption 5.(ii)) is needed so that possible multicollinearities are ruled out in residual equations that involve the parameter vector $\boldsymbol{\rho}$. Under Assumptions 1 and 2, the equilibrium solution of the outcome variable is $\boldsymbol{y} = \boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{v}$. (Accordingly, $\boldsymbol{W}_i\boldsymbol{y} = \boldsymbol{G}_i\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{v}$.)

Note that nothing has been said about the degrees of denseness of the spatial weight matrices, except that implied by Assumption 1. Consider a scenario where under a particular spatial weight specification, the number of neighboring units for each member increases, though at a slower rate, as the sample size increases. It means, for the specific spatial weight matrix, when normalized, its elements are $O(h_n^{-1})$ such that $h_n \to \infty$ and $h_n/n \to 0$ as $n \to \infty$. The appendix provides a general discussion of this case. In particular, such a modification does not affect the estimation procedure presented in this paper, nor will it affect the resulting inference procedure, as one can still construct the t-ratio, for example, without any information on the degrees of denseness of the weight matrices. It may affect the convergence rate of the II estimator of the SAR parameters in the error process. In particular, it may become more difficult to estimate

[3] However, Elhorst et al. (2012) pointed out that this constraint may be too restrictive.

precisely some elements of $\boldsymbol{\rho}_0$ in finite samples when their corresponding weight matrices become denser.[4]

3. Main Results

Given the notation and assumptions in the previous section, this section provides the main results. First it shows how to approximate the bias of each of the estimated SAR parameters under heteroskedastic error innovations. When it is properly recentered, the OLS estimator is in fact consistent and asymptotically normal. Based on this, a set of $p+q$ binding functions is designed and the II estimator, together with its asymptotic distribution, is developed. Section 3.2 outlines a model specification testing procedure to detect possible left-over spatial correlation. Section 3.3 discusses how the II method is related to the best GMM of Liu et al. (2010).

3.1. The II Estimator

The II estimator, as discussed in the introduction, is based on matching the OLS estimator of each of the SAR parameters with its analytical approximate expectation. Suppose $\boldsymbol{\rho}$ and $\boldsymbol{\lambda}_{(-i)}$ are known (equal to their true values), (1) becomes $\boldsymbol{R}\boldsymbol{S}_{(-i)}\boldsymbol{y} = \boldsymbol{R}\boldsymbol{X}\boldsymbol{\beta} + \lambda_i \boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y} + \boldsymbol{v}$, which makes its possible to apply the OLS formula to the i-th SAR parameter in the outcome equation. Let $r_i = \boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}$ and $d_i = \boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}$. Then

$$\hat{\lambda}_i = \frac{\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}_{(-i)}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}} = \lambda_{0i} + \frac{r_i}{d_i} \tag{2}$$

gives the OLS estimator of λ_{0i}. Note that $\hat{\lambda}_i = \hat{\lambda}_i(\boldsymbol{\lambda}_{0,(-i)}, \boldsymbol{\rho}_0)$, depending on the known values of $\boldsymbol{\lambda}_{(-i)}$ and $\boldsymbol{\rho}$. If $p=1$ as considered in Bao and Liu (2021), then one needs to set $\boldsymbol{S}_{(-i)} = \boldsymbol{I}$ in (2), and if further $q=0$ as considered in Bao et al. (2020), then one needs to set $\boldsymbol{R} = \boldsymbol{I}$ in (2). In general, for SARAR(p,q), given the observable sample data $\boldsymbol{y}$ and $\boldsymbol{X}$ (which appears in the matrix $\boldsymbol{H}$), the OLS estimator of each SAR parameter in the outcome equation is a function of all the other SAR parameters in the outcome equation and all the SAR parameters in the error process. As can be expected, $\hat{\lambda}_i$ is generally correlated with $\hat{\lambda}_{i'}$ for any $i \neq i'$, and the associated

[4] When the GMM and GS2SLS estimators use moment conditions that are based on these dense weight matrices, the magnitudes of various quadratic terms that appear in the moment conditions and the associated derivatives will similarly depend on the degrees of denseness of the spatial weight matrices. One would expect the convergence rates of the resulting estimators may also be affected. Kelejian and Prucha (2010) and Jin and Lee (2019) did not consider explicitly the case of dense weight matrices. A rigorous treatment of this issue in the frameworks of GMM and GS2SLS is beyond the scope of this paper.

random binding functions to be introduced later are also correlated with each other.

Substituting $\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y} = \boldsymbol{R}\boldsymbol{G}_i\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{v}$ into r_i and d_i, one has

$$
\begin{aligned}
r_i &= \boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v} + \boldsymbol{v}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{v},\\
d_i &= \boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{v}'\boldsymbol{R}^{-1\prime}\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{v} + 2\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{v},
\end{aligned}
$$

where $\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v} = O_p(n^{1/2})$, $\boldsymbol{v}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{v} = O_p(n)$, $\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{v} = O_p(n^{1/2})$, $\boldsymbol{v}'\boldsymbol{R}^{-1\prime}\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{v} = O_p(n)$, and $\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{X}\boldsymbol{\beta}_0 = O(n)$. (These orders can be verified in view of Lemma A.2 in the appendix.) As such, this estimator of λ_{0i} is biased and inconsistent. One possible way to achieve consistency is to recenter $\hat{\lambda}_i - \lambda_{0i}$ by $\mathrm{E}(r_i)/d_i = \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1})/\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}$.[5] Unfortunately, this recentering term involves the nuisance covariance matrix $\boldsymbol{\Sigma}$. If $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$ are known (equal to their true values), a consistent estimator of $\boldsymbol{\beta}_0$ is $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\lambda}_0, \boldsymbol{\rho}_0) = (\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}$ and the resulting residual vector is $\tilde{\boldsymbol{v}} = \tilde{\boldsymbol{v}}(\boldsymbol{\lambda}_0, \boldsymbol{\rho}_0) = \boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}$. It is obvious that $\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}) = \mathrm{E}(\boldsymbol{v}'\boldsymbol{D}_i\boldsymbol{v})$. (Recall that $\boldsymbol{D}_i = \mathrm{Dg}(\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1})$.) If $\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1})$ is replaced with $\tilde{\boldsymbol{v}}'\boldsymbol{D}_i\tilde{\boldsymbol{v}} = \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{D}_i\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}$, then $\hat{\lambda}_i - \lambda_{0i}$ is linked to $\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{D}_i\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}/\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}$, indicating that, given the sample data $\boldsymbol{y}$ and $\boldsymbol{X}$, the OLS estimator $\hat{\lambda}_i$ is approximately a function of $\boldsymbol{\lambda}_0$ (appearing in $\boldsymbol{S}$ and $\boldsymbol{D}_i$) and $\boldsymbol{\rho}_0$ (appearing in $\boldsymbol{R}$, $\boldsymbol{H}$, and $\boldsymbol{D}_i$).

Lemma 1 *Let* $\hat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \cdots, \hat{\lambda}_p)'$, $\boldsymbol{b}_{\boldsymbol{\lambda}} = (b_{\lambda_1}, \cdots, b_{\lambda_p})'$, $\boldsymbol{s}_{\boldsymbol{\lambda}} = (s_{\lambda_1}, \cdots, s_{\lambda_p})'$ *, where*

$$
\begin{aligned}
b_{\lambda_i} &= \frac{\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{D}_i\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}},\\
s_{\lambda_i} &= \frac{\boldsymbol{v}'\boldsymbol{E}_i\boldsymbol{v} + \boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}}{\mathrm{E}(d_i)},\\
\mathrm{E}(d_i) &= \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{R}^{-1\prime}\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}) + \boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{X}\boldsymbol{\beta}_0.
\end{aligned}
$$

Under Assumptions 1 to 5, if $\boldsymbol{\Xi}_{\boldsymbol{\lambda}} \succ 0$, *then*

$$
\sqrt{n}\boldsymbol{\Xi}_{\boldsymbol{\lambda}}^{-1/2}(\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0 - \boldsymbol{b}_{\boldsymbol{\lambda}}) = \sqrt{n}\boldsymbol{\Xi}_{\boldsymbol{\lambda}}^{-1/2}\boldsymbol{s}_{\boldsymbol{\lambda}} + o_p(1) \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_p), \tag{3}
$$

[5] One may be tempted to use the non-random $\mathrm{E}(r_i)/\mathrm{E}(d_i)$ as the recentering term by following the traditional notion of bias correction, see Kyriacou et al. (2017, 2021). While both $\hat{\lambda}_i - \mathrm{E}(r_i)/d_i$ and $\hat{\lambda}_i - \mathrm{E}(r_i)/\mathrm{E}(d_i)$ are consistent, one can show that the latter has a more complicated asymptotic variance in the sense that it involves the variances of $\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}$ and $\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}$ as well as their covariance. Moreover, using the latter complicates further the variance of the resulting II estimator to be introduced. See discussions in Bao et al. (2020) and Bao and Liu (2021) for the case of $p = 1$.

where $\boldsymbol{\Xi}_{\boldsymbol{\lambda}}$ *has*

$$\xi_{\lambda,ij} = \mathrm{Cov}(s_{\lambda_i}, s_{\lambda_j}) = \mathrm{E}(s_{\lambda_i} s_{\lambda_j}) = \frac{n[\mathrm{tr}(\boldsymbol{\Sigma E}_i \boldsymbol{\Sigma E}_j^{\star}) + \boldsymbol{\beta}_0' \boldsymbol{X}' \boldsymbol{G}_i' \boldsymbol{R}' \boldsymbol{H \Sigma H R G}_j \boldsymbol{X} \boldsymbol{\beta}_0]}{\mathrm{E}(d_i)\mathrm{E}(d_j)} \tag{4}$$

in its (i,j)*-th position,* $i,j = 1, \cdots, p$*.*

The recentering vector $\boldsymbol{b}_{\boldsymbol{\lambda}}$ in Lemma 1 depends on both $\boldsymbol{\lambda}_0$ and $\boldsymbol{\rho}_0$. If one knows the true value of $\boldsymbol{\rho}$, or if there is no spatial correlation in the error process, then by properly recentering the simple OLS estimator, one can correct its inconsistency. Of course, the traditional plug-in approach of bias correction will not work, since one cannot directly estimate the recentering term consistently. The II approach tackles this by treating $\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0 - \boldsymbol{b}_{\boldsymbol{\lambda}}$ as a system of functions and then solving for the unknown parameters. When $\boldsymbol{\rho}$ is unknown, one needs to come up with a second system of equations for $\boldsymbol{\rho}$. If the error vector $\boldsymbol{u}$ is observable, then one should be able to arrive at a similar result provided by Lemma 1 regarding the OLS estimator of $\boldsymbol{\rho}_0$ based on the known $\boldsymbol{u}$. With the unknown error vector $\boldsymbol{u}$, one needs to use a proper residual vector so that it can be used to define the OLS estimator of $\boldsymbol{\rho}_0$.

Given (the infeasible) $\tilde{\boldsymbol{v}}$, define $\tilde{\boldsymbol{u}} = \boldsymbol{R}^{-1}\tilde{\boldsymbol{v}} = \boldsymbol{R}^{-1}\boldsymbol{HRSy}$ and suppose that one estimates ρ_{0j} based on this residual vector, namely,

$$\hat{\rho}_j = \hat{\rho}_j(\boldsymbol{\lambda}_0, \boldsymbol{\rho}_0) = \frac{\tilde{\boldsymbol{u}}' \boldsymbol{R}_{(-j)}' \boldsymbol{M}_j \tilde{\boldsymbol{u}}}{\tilde{\boldsymbol{u}}' \boldsymbol{M}_j' \boldsymbol{M}_j \tilde{\boldsymbol{u}}} = \frac{\boldsymbol{y}' \boldsymbol{S}' \boldsymbol{R}' \boldsymbol{H} \boldsymbol{R}^{-1\prime} \boldsymbol{R}_{(-j)}' \boldsymbol{F}_j \boldsymbol{HRSy}}{\boldsymbol{y}' \boldsymbol{S}' \boldsymbol{R}' \boldsymbol{H} \boldsymbol{F}_j' \boldsymbol{F}_j \boldsymbol{HRSy}}. \tag{5}$$

In contrast to the case of $\hat{\lambda}_i - \lambda_{0i}$, a proper recentering term for $\hat{\rho}_j - \rho_{0j}$ is not obvious. Note that

$$\tilde{\boldsymbol{u}} - \boldsymbol{u} = \boldsymbol{R}^{-1}\boldsymbol{HRSy} - \boldsymbol{R}^{-1}\boldsymbol{v} = \boldsymbol{R}^{-1}(\boldsymbol{H} - \boldsymbol{I})\boldsymbol{v},$$

where all the elements of $\boldsymbol{H} - \boldsymbol{I}$ are $O(n^{-1})$. One can see that, in light of Lemma A.2 in the appendix,

$$(\tilde{\boldsymbol{u}} - \boldsymbol{u})' \boldsymbol{R}_{(-j)}' \boldsymbol{M}_j (\tilde{\boldsymbol{u}} - \boldsymbol{u}) = \boldsymbol{v}' (\boldsymbol{H} - \boldsymbol{I})' \boldsymbol{R}^{-1\prime} \boldsymbol{R}_{(-j)}' \boldsymbol{M}_j \boldsymbol{R}^{-1} (\boldsymbol{H} - \boldsymbol{I}) \boldsymbol{v} = O_p(1)$$

and

$$\boldsymbol{u}' \boldsymbol{R}_{(-j)}' \boldsymbol{M}_j (\tilde{\boldsymbol{u}} - \boldsymbol{u}) = \boldsymbol{v}' \boldsymbol{R}^{-1\prime} \boldsymbol{R}_{(-j)}' \boldsymbol{M}_j \boldsymbol{R}^{-1} (\boldsymbol{H} - \boldsymbol{I}) \boldsymbol{v} = O_p(1).$$

So

$$\begin{aligned}\tilde{\boldsymbol{u}}'\boldsymbol{R}'_{(-j)}\boldsymbol{M}_j\tilde{\boldsymbol{u}} &= (\boldsymbol{u}+\tilde{\boldsymbol{u}}-\boldsymbol{u})'\boldsymbol{R}'_{(-j)}\boldsymbol{M}_j(\boldsymbol{u}+\tilde{\boldsymbol{u}}-\boldsymbol{u})\\ &= \boldsymbol{u}'\boldsymbol{R}'_{(-j)}\boldsymbol{M}_j\boldsymbol{u}+(\tilde{\boldsymbol{u}}-\boldsymbol{u})'\boldsymbol{R}'_{(-j)}\boldsymbol{M}_j(\tilde{\boldsymbol{u}}-\boldsymbol{u})+2\boldsymbol{u}'\boldsymbol{R}'_{(-j)}\boldsymbol{M}_j(\tilde{\boldsymbol{u}}-\boldsymbol{u})\\ &= \boldsymbol{u}'\boldsymbol{R}'_{(-j)}\boldsymbol{M}_j\boldsymbol{u}+O_p(1)\end{aligned}$$

and similarly $\tilde{\boldsymbol{u}}'\boldsymbol{M}'_j\boldsymbol{M}_j\tilde{\boldsymbol{u}} = \boldsymbol{u}'\boldsymbol{M}'_j\boldsymbol{M}_j\boldsymbol{u}+O_p(1)$. Then

$$\begin{aligned}\hat{\rho}_j &= \frac{\boldsymbol{u}'\boldsymbol{R}'_{(-j)}\boldsymbol{M}_j\boldsymbol{u}}{\boldsymbol{u}'\boldsymbol{M}'_j\boldsymbol{M}_j\boldsymbol{u}}+O_p(n^{-1})\\ &= \rho_{0j}+\frac{\boldsymbol{u}'\boldsymbol{M}'_j\boldsymbol{v}}{\boldsymbol{u}'\boldsymbol{M}'_j\boldsymbol{M}_j\boldsymbol{u}}+O_p(n^{-1})\\ &= \rho_{0j}+\frac{\boldsymbol{v}'\boldsymbol{F}_j\boldsymbol{v}}{\boldsymbol{v}'\boldsymbol{F}'_j\boldsymbol{F}_j\boldsymbol{v}}+O_p(n^{-1}).\end{aligned}$$

Now $\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}_j)/\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{F}'_j\boldsymbol{F}_j\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}$ is a natural choice to recenter $\hat{\rho}_j-\rho_{0j}$. As before, one needs to take care of the nuisance covariance matrix $\boldsymbol{\Sigma}$ in the recentering term. The strategy is to use $\tilde{\boldsymbol{v}}'_n\mathrm{Dg}(\boldsymbol{F}_j)\tilde{\boldsymbol{v}}_n = \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{K}_j\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}$ instead of $\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}_j)=\mathrm{E}(\boldsymbol{v}'\mathrm{Dg}(\boldsymbol{F}_j)\boldsymbol{v})$ in the recentering term. This links $\hat{\rho}_j-\rho_{0j}$ to $\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{K}_j\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}/\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{F}'_j\boldsymbol{F}_j\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}$ (recall that $\boldsymbol{K}_j = \mathrm{Dg}(\boldsymbol{F}_j)$), indicating that, given the sample data $\boldsymbol{y}$ and $\boldsymbol{X}$, the OLS estimator $\hat{\rho}_j$ is approximately a function of $\boldsymbol{\lambda}_0$ (appearing in $\boldsymbol{S}$) and $\boldsymbol{\rho}_0$ (appearing in $\boldsymbol{R}$, $\boldsymbol{H}$, and $\boldsymbol{K}_i$).

Lemma 2 *Let* $\hat{\boldsymbol{\rho}}=(\hat{\rho}_1,\cdots,\hat{\rho}_q)'$, $\boldsymbol{b}_{\boldsymbol{\rho}}=(b_{\rho_1},\cdots,b_{\rho_q})'$, *and* $\boldsymbol{s}_{\boldsymbol{\rho}}=(s_{\rho_1},\cdots,s_{\rho_q})'$, *where*

$$b_{\rho_j}=\frac{\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{K}_j\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{F}'_j\boldsymbol{F}_j\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}},\quad s_{\rho_j}=\frac{\boldsymbol{v}'\boldsymbol{L}_j\boldsymbol{v}}{\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}'_j\boldsymbol{F}_j)}.$$

Under Assumptions 1–5, if $\boldsymbol{\Xi}_{\boldsymbol{\rho}}\succ 0$, *then*

$$\sqrt{n}\boldsymbol{\Xi}_{\boldsymbol{\rho}}{}^{-1/2}(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0-\boldsymbol{b}_{\boldsymbol{\rho}})=\sqrt{n}\boldsymbol{\Xi}_{\boldsymbol{\rho}}{}^{-1/2}\boldsymbol{s}_{\boldsymbol{\rho}}+o_p(1)\xrightarrow{d}\mathrm{N}(\boldsymbol{0},\boldsymbol{I}_q),\tag{6}$$

where $\boldsymbol{\Xi}_{\boldsymbol{\rho}}$ *has*

$$\xi_{\rho,ij}=\mathrm{Cov}(s_{\rho_i},s_{\rho_j})=\mathrm{E}(s_{\rho_i}s_{\rho_j})=\frac{n\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{L}_i\boldsymbol{\Sigma}\boldsymbol{L}^{\star}_j)}{\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}'_i\boldsymbol{F}_i)\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}'_j\boldsymbol{F}_j)}\tag{7}$$

in its (i,j)*-th position,* $i,j=1,\cdots,q$.

With $\hat{\boldsymbol{\lambda}}-\boldsymbol{\lambda}_0-\boldsymbol{b}_{\boldsymbol{\lambda}}=\boldsymbol{s}_{\boldsymbol{\lambda}}+o_p(n^{-1/2})$ and $\hat{\boldsymbol{\rho}}_n-\boldsymbol{\rho}_0-\boldsymbol{b}_{\boldsymbol{\rho}}=\boldsymbol{s}_{\boldsymbol{\rho}}+o_p(n^{-1/2})$ from Lemmas 1 and 2, respectively, where both $\boldsymbol{s}_{\boldsymbol{\lambda}}$ and $\boldsymbol{s}_{\boldsymbol{\rho}}$ are linear and quadratic forms in $\boldsymbol{v}$, their joint asymptotic distribution follows straightforwardly.

Lemma 3 *Let* $\hat{\boldsymbol{\gamma}} = (\hat{\boldsymbol{\lambda}}', \hat{\boldsymbol{\rho}}')'$, $\boldsymbol{b} = (\boldsymbol{b}'_{\boldsymbol{\lambda}}, \boldsymbol{b}'_{\boldsymbol{\rho}})'$, *and* $\boldsymbol{\Xi} = ((\boldsymbol{\Xi}'_{\boldsymbol{\lambda}}, \boldsymbol{\Xi}_{\boldsymbol{\lambda\rho}})', (\boldsymbol{\Xi}'_{\boldsymbol{\lambda\rho}}, \boldsymbol{\Xi}'_{\boldsymbol{\rho}})')$, *where* $\boldsymbol{\Xi}_{\boldsymbol{\lambda\rho}}$ *has*

$$\xi_{\lambda\rho,ij} = \mathrm{Cov}(s_{\lambda_i}, s_{\rho_j}) = \mathrm{E}(s_{\lambda_i} s_{\rho_j}) = \frac{n\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{E}_i\boldsymbol{\Sigma}\boldsymbol{L}_j^{\star})}{\mathrm{E}(d_i)\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}'_j\boldsymbol{F}_j)} \tag{8}$$

in its (i,j)*-th position,* $i = 1, \cdots, p$, $j = 1, \cdots, q$. *Under Assumptions 1 to 5, if* $\boldsymbol{\Xi} \succ 0$, *then*

$$\sqrt{n}\boldsymbol{\Xi}^{-1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0 - \boldsymbol{b}) \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_{p+q}). \tag{9}$$

The asymptotic distribution result (9) can be used to design an estimator of $\boldsymbol{\gamma}_0$ in the spirit of indirect inference by matching $\hat{\boldsymbol{\gamma}}$ with its (approximate) expectation $\boldsymbol{\gamma}_0 + \boldsymbol{b}$. Keep in mind that $\hat{\boldsymbol{\gamma}}$ and $\boldsymbol{b}$ depend on $\boldsymbol{\gamma}_0$. So the II estimator of $\boldsymbol{\gamma}_0$ is the root of the vector-valued function $\boldsymbol{\psi}(\boldsymbol{\gamma}) = \hat{\boldsymbol{\gamma}}(\boldsymbol{\gamma}) - \boldsymbol{\gamma} - \boldsymbol{b}(\boldsymbol{\gamma}) = (\psi_1(\boldsymbol{\gamma}), \cdots, \psi_p(\boldsymbol{\gamma}), \psi_{p+1}(\boldsymbol{\gamma}), \cdots, \psi_{p+q}(\boldsymbol{\gamma}))'$, where the first p elements are

$$\begin{aligned}\psi_i(\boldsymbol{\gamma}) &= \psi_{\lambda_i}(\boldsymbol{\gamma})\\ &= \frac{\boldsymbol{y}'\boldsymbol{W}'_i\boldsymbol{R}'(\boldsymbol{\rho})\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{R}(\boldsymbol{\rho})\boldsymbol{S}_{(-i)}(\boldsymbol{\lambda})\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{W}'_i\boldsymbol{R}'(\boldsymbol{\rho})\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{R}(\boldsymbol{\rho})\boldsymbol{W}_i\boldsymbol{y}} - \lambda_i - \frac{\boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{\lambda})\boldsymbol{R}'(\boldsymbol{\rho})\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{D}_i(\gamma)\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{R}(\boldsymbol{\rho})\boldsymbol{S}(\boldsymbol{\lambda})\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{W}'_i\boldsymbol{R}'(\boldsymbol{\rho})\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{R}(\boldsymbol{\rho})\boldsymbol{W}_i\boldsymbol{y}},\end{aligned}$$

$i = 1, \cdots, p$, and the second q elements are

$$\begin{aligned}\psi_{p+j}(\boldsymbol{\gamma}) &= \psi_{\rho_j}(\boldsymbol{\gamma})\\ &= \frac{\boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{\lambda})\boldsymbol{R}'(\boldsymbol{\rho})\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{R}^{-1\prime}(\boldsymbol{\rho})\boldsymbol{R}'_{(-j)}(\boldsymbol{\rho})\boldsymbol{F}_j(\boldsymbol{\rho})\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{R}(\boldsymbol{\rho})\boldsymbol{S}(\boldsymbol{\lambda})\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{\lambda})\boldsymbol{R}'(\boldsymbol{\rho})\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{F}'_j(\boldsymbol{\rho})\boldsymbol{F}_j(\boldsymbol{\rho})\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{R}(\boldsymbol{\rho})\boldsymbol{S}(\boldsymbol{\lambda})\boldsymbol{y}} - \rho_j\\ &\quad - \frac{\boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{\lambda})\boldsymbol{R}'(\boldsymbol{\rho})\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{K}_j(\boldsymbol{\rho})\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{R}(\boldsymbol{\rho})\boldsymbol{S}(\boldsymbol{\lambda})\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{\lambda})\boldsymbol{R}'(\boldsymbol{\rho})\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{F}'_j(\boldsymbol{\rho})\boldsymbol{F}_j(\boldsymbol{\rho})\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{R}(\boldsymbol{\rho})\boldsymbol{S}(\boldsymbol{\lambda})\boldsymbol{y}},\end{aligned}$$

$j = 1, \cdots, q$. Succinctly, the II procedure is equivalent to solving

$$\boldsymbol{\psi}(\boldsymbol{\gamma}) = \boldsymbol{0}. \tag{10}$$

Assumption 6 *For* $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, *(i)* $\Pr(\lim \boldsymbol{\psi}(\boldsymbol{\gamma}) = \boldsymbol{0}) = 1$ *at* $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$ *and* $\Pr(\lim \boldsymbol{\psi}(\boldsymbol{\gamma}) \neq \boldsymbol{0}) = 1$ *for any* $\boldsymbol{\gamma} \neq \boldsymbol{\gamma}_0$, *and (ii) the Jacobian* $\boldsymbol{\Psi}(\boldsymbol{\gamma}) = \partial\boldsymbol{\psi}(\boldsymbol{\gamma})/\partial\boldsymbol{\gamma}'$ *is nonsingular almost surely.*

Essentially, Assumption 6 ensures the existence and uniqueness of the root of the sample binding function $\boldsymbol{\psi}(\boldsymbol{\gamma})$, at least in large samples. It would be desirable to set up some primitive conditions that lead to the existence and uniqueness of the root for any given sample. These conditions will depend on the structure of the data matrix, the characteristics of the weight matrices, and the parameter space. For a given sample, however, one may check the Jacobian's determinant, condition number, or inverse error bound, over a grid of values of $\boldsymbol{\gamma}$ to verify numerically validity of this assumption.[6] Assumption 6 is also needed for deriving the asymptotic distribution of the resulting II estimator.

Theorem 1 *Let $\boldsymbol{V}_{\boldsymbol{\gamma}} = \boldsymbol{\Psi}^{-1}\boldsymbol{\Xi}\boldsymbol{\Psi}^{-1\prime} = ((\boldsymbol{V}'_{\boldsymbol{\lambda}}, \boldsymbol{V}_{\boldsymbol{\lambda\rho}})', (\boldsymbol{V}'_{\boldsymbol{\lambda\rho}}, \boldsymbol{V}'_{\boldsymbol{\rho}})')$, partitioned in accordance with $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$. Under Assumptions 1 to 6, if $\boldsymbol{V}_{\boldsymbol{\gamma}} \succ 0$, the asymptotic distribution of the II estimator of $\boldsymbol{\gamma}_0$ that solves (10) is*

$$\sqrt{n}\boldsymbol{V}_{\boldsymbol{\gamma}}^{-1/2}(\hat{\boldsymbol{\gamma}}_{II} - \boldsymbol{\gamma}_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{I}_{p+q}). \tag{11}$$

Once $\boldsymbol{\gamma}_0$ is estimated by $\hat{\boldsymbol{\gamma}}$, one can estimate $\boldsymbol{\beta}_0$ by the usual OLS procedure, namely,

$$\hat{\boldsymbol{\beta}}_{II} = \hat{\boldsymbol{\beta}}_{II}(\hat{\boldsymbol{\gamma}}_{II}) = (\boldsymbol{X}'\boldsymbol{R}'(\hat{\boldsymbol{\rho}}_{II})\boldsymbol{R}(\hat{\boldsymbol{\rho}}_{II})\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'(\hat{\boldsymbol{\rho}}_{II})\boldsymbol{R}(\hat{\boldsymbol{\rho}}_{II})\boldsymbol{S}(\hat{\boldsymbol{\lambda}}_{II})\boldsymbol{y}. \tag{12}$$

Given that $\hat{\boldsymbol{\lambda}}_{II}$ and $\hat{\boldsymbol{\rho}}_{II}$ are consistent, $\hat{\boldsymbol{\beta}}_{II}$ defined as such is necessarily consistent. Let $\boldsymbol{\Psi}^{-1} = ((\boldsymbol{\Psi}_{\boldsymbol{\lambda}}^{(-1)\prime}, \boldsymbol{\Psi}_{\boldsymbol{\lambda\rho}}^{(-1)})', (\boldsymbol{\Psi}_{\boldsymbol{\lambda\rho}}^{(-1)\prime}, \boldsymbol{\Psi}_{\boldsymbol{\rho}}^{(-1)\prime})')$, partitioned in accordance with $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$. The following theorem gives the joint distribution of $\hat{\boldsymbol{\gamma}}_{II}$ and $\hat{\boldsymbol{\beta}}_{II}$.

Theorem 2 *Let $\boldsymbol{V} = ((\boldsymbol{V}'_{\boldsymbol{\gamma}}, \boldsymbol{V}_{\boldsymbol{\gamma\beta}})', (\boldsymbol{V}'_{\boldsymbol{\gamma\beta}}, \boldsymbol{V}'_{\boldsymbol{\beta}})')$, where*

$$\boldsymbol{V}_{\boldsymbol{\beta}} = n(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{\Sigma}\boldsymbol{R}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1} + \boldsymbol{J}_1\boldsymbol{V}_{\boldsymbol{\lambda}}\boldsymbol{J}'_1 - n\boldsymbol{J}_1\boldsymbol{\Psi}_{\boldsymbol{\lambda}}^{(-1)}\boldsymbol{J}'_2 - n\boldsymbol{J}_2\boldsymbol{\Psi}_{\boldsymbol{\lambda}}^{(-1)\prime}\boldsymbol{J}'_1$$

and $\boldsymbol{V}_{\boldsymbol{\gamma\beta}} = (\boldsymbol{V}'_{\boldsymbol{\lambda\beta}}, \boldsymbol{V}'_{\boldsymbol{\rho\beta}})'$, in which

$$\begin{aligned}
\boldsymbol{V}_{\boldsymbol{\lambda\beta}} &= n\boldsymbol{\Psi}_{\boldsymbol{\lambda}}^{(-1)}\boldsymbol{J}'_2 - \boldsymbol{V}_{\boldsymbol{\lambda}}\boldsymbol{J}'_1, \quad \boldsymbol{V}_{\boldsymbol{\rho\beta}} = n(\boldsymbol{\Psi}_{\boldsymbol{\rho\lambda}}^{(-1)}\boldsymbol{J}'_2 - \boldsymbol{V}_{\boldsymbol{\rho\lambda}}\boldsymbol{J}'_1), \\
\boldsymbol{J}_1 &= [(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{G}_1\boldsymbol{X}\boldsymbol{\beta}_0, \cdots, (\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{G}_p\boldsymbol{X}\boldsymbol{\beta}_0], \\
\boldsymbol{J}_2 &= \left[\frac{(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{\Sigma}\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_1\boldsymbol{X}\boldsymbol{\beta}_0}{\mathrm{E}(d_1)}, \cdots, \frac{(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{\Sigma}\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_p\boldsymbol{X}\boldsymbol{\beta}_0}{\mathrm{E}(d_p)}\right].
\end{aligned}$$

[6] Note that matrix determinant may not be a good indicator for assessing how close a matrix is to being singular. The condition number is a more reliable measure. However, there is no agreed rule on how large the condition number should be to regard a matrix as an ill-conditioned one. A more useful measure in practice is the matrix inverse relative error bound $\varepsilon = \varepsilon(\boldsymbol{\Psi}^{-1}(\boldsymbol{\gamma}))$, see Wilkinson (1961). It measures the upper bound of the error in the computed inverse relative to the exact inverse.

Under Assumptions 1 to 6, if $\boldsymbol{V} \succ 0$, the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{II} = (\hat{\boldsymbol{\gamma}}'_{II}, \hat{\boldsymbol{\beta}}'_{II})'$ is

$$\sqrt{n}\boldsymbol{V}^{-1/2}(\hat{\boldsymbol{\theta}}_{II} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{I}_k). \tag{13}$$

The first part of $\boldsymbol{V}_{\boldsymbol{\beta}}$ is the usual variance formula of the OLS estimator of $\boldsymbol{\beta}_0$ under heteroskedasticity in the case of known values of $\boldsymbol{\lambda}$ and $\boldsymbol{\rho}$. The remaining terms are due to the estimation uncertainty associated with $\hat{\boldsymbol{\lambda}}_{II}$.[7] Note that $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\theta}_0, \boldsymbol{\Sigma})$. In practice, one can estimate $\boldsymbol{V}$ by $\hat{\boldsymbol{V}} = \boldsymbol{V}(\boldsymbol{y}, \boldsymbol{X}, \hat{\boldsymbol{\theta}}_{II}, \mathrm{Dg}(\hat{\boldsymbol{v}}^2))$, where $\hat{\boldsymbol{v}} = \boldsymbol{H}(\hat{\boldsymbol{\rho}}_{II})\boldsymbol{R}(\hat{\boldsymbol{\rho}}_{II})\boldsymbol{S}(\hat{\boldsymbol{\lambda}}_{II})\boldsymbol{y} \equiv \hat{\boldsymbol{H}}\hat{\boldsymbol{R}}\hat{\boldsymbol{S}}\boldsymbol{y}$.

3.2. A Moran-Type Test

After estimation, one may follow Kelejian and Prucha (2001) and Robinson (2008), also see Kyriacou et al. (2021), to check whether there is still left-over spatial correlation in the error process. Under the null of correct model specification, $\boldsymbol{v}$ from (1) should have a diagonal variance structure, namely, $\mathrm{E}(\boldsymbol{v}\boldsymbol{v}') = \boldsymbol{\Sigma}$. For this purpose, a Moran-type statistic based on $\hat{\boldsymbol{v}}'\boldsymbol{A}\hat{\boldsymbol{v}}$ can be used, where $\boldsymbol{A}$ is a matrix that signals possible spatial correlation and also satisfies Assumption 1.[8] By substituting $\hat{\boldsymbol{v}} = (\boldsymbol{H} + \hat{\boldsymbol{H}} - \boldsymbol{H})(\boldsymbol{R} + \hat{\boldsymbol{R}} - \boldsymbol{R})(\boldsymbol{S} + \hat{\boldsymbol{S}} - \boldsymbol{S})\boldsymbol{y}$ into $\hat{\boldsymbol{v}}'\boldsymbol{A}\hat{\boldsymbol{v}}$ and expanding, one has

$$\hat{\boldsymbol{v}}'\boldsymbol{A}\hat{\boldsymbol{v}} = \tilde{\boldsymbol{v}}'\boldsymbol{A}\tilde{\boldsymbol{v}} + 2\tilde{\boldsymbol{v}}'\boldsymbol{A}\boldsymbol{H}\boldsymbol{R}(\hat{\boldsymbol{S}} - \boldsymbol{S})\boldsymbol{y} + 2\tilde{\boldsymbol{v}}'\boldsymbol{A}\boldsymbol{H}(\hat{\boldsymbol{R}} - \boldsymbol{R})\boldsymbol{S}\boldsymbol{y} + 2\tilde{\boldsymbol{v}}'\boldsymbol{A}(\hat{\boldsymbol{H}} - \boldsymbol{H})\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} + o_p(n^{-1/2}),$$

where recall that $\tilde{\boldsymbol{v}} = \tilde{\boldsymbol{v}}(\boldsymbol{\lambda}_0, \boldsymbol{\rho}_0) = \boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}$. By another round of substitution of $\tilde{\boldsymbol{v}} = \boldsymbol{v} + (\boldsymbol{H} - \boldsymbol{I})\boldsymbol{v}$, $\hat{\boldsymbol{S}} - \boldsymbol{S} = -\sum_{i=1}^{p}(\hat{\lambda}_{II,i} - \lambda_{0i})\boldsymbol{W}_i$, $\hat{\boldsymbol{R}} - \boldsymbol{R} = -\sum_{j=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})\boldsymbol{M}_j$, $\hat{\boldsymbol{H}} - \boldsymbol{H} = \boldsymbol{R}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}' - \hat{\boldsymbol{R}}\boldsymbol{X}(\boldsymbol{X}'\hat{\boldsymbol{R}}'\hat{\boldsymbol{R}}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{R}}'$, and (see the proof of Theorem 2 in the appendix),

$$\begin{aligned}(\boldsymbol{X}'\hat{\boldsymbol{R}}'\hat{\boldsymbol{R}}\boldsymbol{X})^{-1} &= (\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1} \\ &\quad + \sum_{j=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{R}\boldsymbol{M}_j)^{\star}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1} + o_p(n^{-3/2}),\end{aligned}$$

[7] From the expansion in the proof of Theorem 2 in the Appendix, one sees that the estimation uncertainty of $\hat{\boldsymbol{\rho}}_{II}$ does not contribute to the variance of $\hat{\boldsymbol{\beta}}_{II}$.

[8] The following discussion is based on all the spatial matrices ($\boldsymbol{W}'s$, $\boldsymbol{M}$'s, and $\boldsymbol{A}$) being sparse. Otherwise, while the first equality in equation (14) to be introduced still holds, $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ in (14) may need to have different scaling factors in their definitions in the remaining equalities (so that their limits can be properly defined) and the three terms in the last equality may have different magnitudes.

one can write

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\hat{\boldsymbol{v}}'\boldsymbol{A}\hat{\boldsymbol{v}} \\
&= \frac{1}{\sqrt{n}}\boldsymbol{v}'\boldsymbol{A}\boldsymbol{v} - \frac{2}{\sqrt{n}}\sum_{i=1}^{p}(\hat{\lambda}_{II,i} - \lambda_{0i})\boldsymbol{v}'\boldsymbol{AHRG}_i(\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{R}^{-1}\boldsymbol{v}) \\
&- \frac{2}{\sqrt{n}}\sum_{j=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})\boldsymbol{v}'\boldsymbol{AH}(\boldsymbol{M}_j\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{F}_j\boldsymbol{v}) + o_p(1) \\
&+ \frac{2}{\sqrt{n}}\sum_{j=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})\boldsymbol{v}'\boldsymbol{A}\left[\boldsymbol{M}_j\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\right]^{\star}(\boldsymbol{R}\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{v}) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\boldsymbol{v}'\boldsymbol{A}\boldsymbol{v} - 2\boldsymbol{t}_1'\sqrt{n}(\hat{\boldsymbol{\lambda}}_{II} - \boldsymbol{\lambda}_0) - 2\boldsymbol{t}_2'\sqrt{n}(\hat{\boldsymbol{\rho}}_{II} - \boldsymbol{\rho}_0) + o_p(1) \\
&= \frac{1}{\sqrt{n}}\boldsymbol{v}'\boldsymbol{A}\boldsymbol{v} - 2\boldsymbol{t}_1'\sqrt{n}(\boldsymbol{\Psi}_{\boldsymbol{\lambda}}^{(-1)}\boldsymbol{s}_{\boldsymbol{\lambda}} + \boldsymbol{\Psi}_{\boldsymbol{\lambda\rho}}^{(-1)}\boldsymbol{s}_{\boldsymbol{\rho}}) - 2\boldsymbol{t}_2'\sqrt{n}(\boldsymbol{\Psi}_{\boldsymbol{\rho}}^{(-1)}\boldsymbol{s}_{\boldsymbol{\rho}} + \boldsymbol{\Psi}_{\boldsymbol{\rho\lambda}}^{(-1)}\boldsymbol{s}_{\boldsymbol{\lambda}}) + o_p(1) \\
&= \begin{pmatrix} 1 & -2(\boldsymbol{t}_1'\boldsymbol{\Psi}_{\boldsymbol{\lambda}}^{(-1)} + \boldsymbol{t}_2'\boldsymbol{\Psi}_{\boldsymbol{\rho\lambda}}^{(-1)}) & -2(\boldsymbol{t}_1'\boldsymbol{\Psi}_{\boldsymbol{\lambda\rho}}^{(-1)} + \boldsymbol{t}_2'\boldsymbol{\Psi}_{\boldsymbol{\rho}}^{(-1)}) \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{n}}\boldsymbol{v}'\boldsymbol{A}\boldsymbol{v} \\ \sqrt{n}\boldsymbol{s}_{\boldsymbol{\lambda}} \\ \sqrt{n}\boldsymbol{s}_{\boldsymbol{\rho}} \end{pmatrix} + o_p(1), \qquad (14)
\end{aligned}
$$

where $\boldsymbol{t}_1'$ horizontally stacks $n^{-1}\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{AHRG}_i\boldsymbol{R}^{-1})$, $i = 1, \cdots, p$, and $\boldsymbol{t}_2'$ horizontally stacks $n^{-1}\mathrm{tr}\left(\boldsymbol{\Sigma}\boldsymbol{A}\left(\boldsymbol{HF}_j - \left[\boldsymbol{M}_j\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\right]^{\star}\right)\right)$, $j = 1, \cdots, q$. It follows that

$$
\frac{1}{\sqrt{n}}\hat{\boldsymbol{v}}'\boldsymbol{A}\hat{\boldsymbol{v}} \xrightarrow{d} \mathrm{N}(0, z^2), \qquad (15)
$$

where

$$
\begin{aligned}
z^2 = \lim \frac{1}{n}\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^{\star}) + 4\lim\begin{pmatrix} \boldsymbol{t}_1' & \boldsymbol{t}_2' \end{pmatrix}\boldsymbol{V}_{\boldsymbol{\gamma}}\begin{pmatrix} \boldsymbol{t}_1' & \boldsymbol{t}_2' \end{pmatrix}' \\
- 4\lim\begin{pmatrix} \boldsymbol{t}_1'\boldsymbol{\Psi}_{\boldsymbol{\lambda}}^{(-1)} + \boldsymbol{t}_2'\boldsymbol{\Psi}_{\boldsymbol{\rho\lambda}}^{(-1)} & \boldsymbol{t}_1'\boldsymbol{\Psi}_{\boldsymbol{\lambda\rho}}^{(-1)} + \boldsymbol{t}_2'\boldsymbol{\Psi}_{\boldsymbol{\rho}}^{(-1)} \end{pmatrix}\boldsymbol{c},
\end{aligned}
$$

and $\boldsymbol{c} = (c_1, \cdots, c_{p+q})'$, $c_i = \lim \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{E}_i\boldsymbol{\Sigma}\boldsymbol{A}^{\star})/\mathrm{E}(d_i)$, $c_{p+j} = \lim \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{L}_j\boldsymbol{\Sigma}\boldsymbol{A}^{\star})/\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}_j'\boldsymbol{F}_j)$, $i = 1, \cdots, p$, $j = 1, \cdots, q$. By using the sample analogues and the II estimator, the variance z^2 can be estimated. Denote such an estimator by $\hat{z}^2$. Then a Moran-type statistic is

$$
\mathcal{I} = \frac{(\hat{\boldsymbol{v}}'\boldsymbol{A}\hat{\boldsymbol{v}})^2}{n\hat{z}^2}, \qquad (16)
$$

which converges to a chi-squared distribution with one degree of freedom under the null of correct model specification.

3.3. In Relation to GMM

From Section 3.1, one can write

$$\hat{\boldsymbol{\gamma}}(\boldsymbol{\gamma}_0) - \boldsymbol{\gamma}_0 - \boldsymbol{b}(\boldsymbol{\gamma}_0) = \begin{pmatrix} \frac{\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_1'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}+\boldsymbol{v}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_1\boldsymbol{R}^{-1}\boldsymbol{v}-\boldsymbol{v}'\boldsymbol{H}\boldsymbol{D}_1\boldsymbol{H}\boldsymbol{v}}{\boldsymbol{y}'\boldsymbol{W}_1'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_1\boldsymbol{y}} \\ \vdots \\ \frac{\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_p'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}+\boldsymbol{v}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_p\boldsymbol{R}^{-1}\boldsymbol{v}-\boldsymbol{v}'\boldsymbol{H}\boldsymbol{D}_p\boldsymbol{H}\boldsymbol{v}}{\boldsymbol{y}'\boldsymbol{W}_p'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_p\boldsymbol{y}} \\ \frac{\boldsymbol{v}'\boldsymbol{F}_1\boldsymbol{v}-\boldsymbol{v}'\boldsymbol{H}\boldsymbol{K}_1\boldsymbol{H}\boldsymbol{v}}{\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{F}_1'\boldsymbol{F}_1\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}} \\ \vdots \\ \frac{\boldsymbol{v}'\boldsymbol{F}_q\boldsymbol{v}-\boldsymbol{v}'\boldsymbol{H}\boldsymbol{K}_q\boldsymbol{H}\boldsymbol{v}}{\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{F}_q'\boldsymbol{F}_q\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}} \end{pmatrix} + O_p(n^{-1})$$

$$= \begin{pmatrix} \frac{\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_1'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}+\boldsymbol{v}'\boldsymbol{E}_1\boldsymbol{v}}{\boldsymbol{y}'\boldsymbol{W}_1'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_1\boldsymbol{y}} \\ \vdots \\ \frac{\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_p'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}+\boldsymbol{v}'\boldsymbol{E}_p\boldsymbol{v}}{\boldsymbol{y}'\boldsymbol{W}_p'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_p\boldsymbol{y}} \\ \frac{\boldsymbol{v}'\boldsymbol{L}_1\boldsymbol{v}}{\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{F}_1'\boldsymbol{F}_1\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}} \\ \vdots \\ \frac{\boldsymbol{v}'\boldsymbol{L}_q\boldsymbol{v}}{\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{F}_q'\boldsymbol{F}_q\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}} \end{pmatrix} + O_p(n^{-1}). \tag{17}$$

Given $\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}$ and $\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{F}_j'\boldsymbol{F}_j\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}$ are nonzero in large samples (see Assumption 5) and are both of order $O_p(n)$, (17) can be written as

$$\frac{1}{n}\begin{pmatrix} \boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_1'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}+\boldsymbol{v}'\boldsymbol{E}_1\boldsymbol{v} \\ \vdots \\ \boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_p'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}+\boldsymbol{v}'\boldsymbol{E}_p\boldsymbol{v} \\ \boldsymbol{v}'\boldsymbol{L}_1\boldsymbol{v} \\ \vdots \\ \boldsymbol{v}'\boldsymbol{L}_q\boldsymbol{v} \end{pmatrix} \times \frac{n}{O_p(n)} + O_p(n^{-1})$$

$$= \frac{1}{n}\begin{pmatrix} \boldsymbol{v}'\boldsymbol{E}_1\boldsymbol{v} \\ \vdots \\ \boldsymbol{v}'\boldsymbol{E}_p\boldsymbol{v} \\ \boldsymbol{v}'\boldsymbol{L}_1\boldsymbol{v} \\ \vdots \\ \boldsymbol{v}'\boldsymbol{L}_q\boldsymbol{v} \end{pmatrix} \times \frac{n}{O_p(n)} + O_p(n^{-1/2}) + O_p(n^{-1}) = \boldsymbol{0} + O_p(n^{-1/2}) + O_p(n^{-1}), \tag{18}$$

where the $O_p(n^{-1/2})$ term includes the omitted $n^{-1}\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}$ when one uses only the quadratic terms collected in $\boldsymbol{m}(\boldsymbol{\gamma}_0) = n^{-1}(\boldsymbol{v}'\boldsymbol{E}_1\boldsymbol{v}, \cdots, \boldsymbol{v}'\boldsymbol{E}_p\boldsymbol{v}, \boldsymbol{v}'\boldsymbol{L}_1\boldsymbol{v}, \cdots, \boldsymbol{v}'\boldsymbol{L}_q\boldsymbol{v})'$, up to some scaling constant, to solve for the II estimator. This set of $p+q$ conditions is conditional on the parameter vector $\boldsymbol{\beta}_0$, since $\boldsymbol{v} = \boldsymbol{R}\boldsymbol{S}\boldsymbol{y} - \boldsymbol{R}\boldsymbol{X}\boldsymbol{\beta}_0$. The II procedure replaces $\boldsymbol{v}$ with $\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}$ in (18), where $\boldsymbol{H}$ concentrates out $\boldsymbol{\beta_0}$. However, the matrix $\boldsymbol{\Xi}$, whose probability limit is related to the asymptotic variance of the II estimator of $\boldsymbol{\gamma}_0$, involves $\boldsymbol{H}$ and thus does take into account the effects of this concentration. The effects of the omitted $n^{-1}\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}$ terms in the moment conditions for λ_{0i} enter directly the top-left block of $\boldsymbol{\Xi}$ that pertains to $\boldsymbol{\lambda}_0$, see (4), as well as the top-right and lower-left blocks, see (8). This set of $p+q$ conditions alone is not enough for one to construct a GMM-type estimator of $\boldsymbol{\theta}_0$. For one to be able to estimate jointly the whole parameter vector $\boldsymbol{\theta}_0$, they can be augmented with moment conditions related to $\boldsymbol{\beta}_0$ so that

$$\boldsymbol{g}(\boldsymbol{\theta}_0) = \frac{1}{n}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{v}, \boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_1'\boldsymbol{R}'\boldsymbol{v}, \cdots, \boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_p'\boldsymbol{R}'\boldsymbol{v}, \boldsymbol{v}'\boldsymbol{E}_1\boldsymbol{v}, \cdots, \boldsymbol{v}'\boldsymbol{E}_p\boldsymbol{v}, \boldsymbol{v}'\boldsymbol{L}_1\boldsymbol{v}, \cdots, \boldsymbol{v}'\boldsymbol{L}_q\boldsymbol{v})', \tag{19}$$

where $\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{v}$ follows from the exogeneity of $\boldsymbol{X}$, $\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{v}$ follows from the part $\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}$ (with $\boldsymbol{H}$ replaced by $\boldsymbol{I}$ since $\boldsymbol{\beta}_0$ is not concentrated out and is estimated jointly) that is not used directly in the II estimation procedure, and $\boldsymbol{E}_i$ in $\boldsymbol{v}'\boldsymbol{E}_i\boldsymbol{v}$ is $\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1} - \mathrm{Dg}(\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1})$ (where again $\boldsymbol{H}$ is replaced by $\boldsymbol{I}$ in the original definition of $\boldsymbol{E}_i$).

When v_i is normal and independent and identically distributed (i.i.d.), Liu et al. (2010) showed that the above set of moment conditions is the best in the sense that the resulting GMM estimator is as efficient as the ML estimator. In the presence of heteroskedasticity, the best moment conditions involve the unknown matrix $\boldsymbol{\Sigma}$ and using the best moment conditions with estimated $\boldsymbol{\Sigma}$ may make them no longer the best.[9] In the first-order SAR framework, Lin and Lee (2010) recommended using the estimated moment conditions $n^{-1}(\boldsymbol{X}'\boldsymbol{v}, \boldsymbol{X}'\hat{\boldsymbol{G}}\boldsymbol{v}, \boldsymbol{v}'(\hat{\boldsymbol{G}} - \mathrm{Dg}(\hat{\boldsymbol{G}}))\boldsymbol{v})'$. Similarly, Liu et al. (2010) recommended using some initially consistently estimated $\hat{\boldsymbol{G}}_i$, $\hat{\boldsymbol{E}}_i$, $\hat{\boldsymbol{R}}_j$, and $\hat{\boldsymbol{L}}_j$ and they showed that this does not affect the asymptotic efficiency of the best GMM estimator under homoskedasticity.

Setting aside the effects of $\boldsymbol{\Sigma}$, one can see the differences between the II and best GMM estimators. First, the II procedure estimates $\boldsymbol{\gamma}_0$ and then $\boldsymbol{\beta}_0$ in the second step and in estimating $\boldsymbol{\gamma}_0$ the moment conditions $\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}$ are not taken into account directly. In contrast, the best GMM considers a broader set of moment conditions. When there is no $\boldsymbol{X}$, the two approaches

[9] Suppose one defines $\hat{\lambda}_i = \boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{H}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{R}\boldsymbol{S}_{(-i)}\boldsymbol{y}/\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{H}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}$, where $\boldsymbol{H} = \boldsymbol{I} - \boldsymbol{\Sigma}^{-1/2}\boldsymbol{R}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{\Sigma}^{-1}\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{\Sigma}^{-1/2}$ and $\hat{\rho}_j = \boldsymbol{u}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{R}_{(-j)}'\boldsymbol{M}_j\boldsymbol{\Sigma}^{-1/2}\boldsymbol{u}/\boldsymbol{u}'\boldsymbol{\Sigma}^{-1/2}\boldsymbol{M}_j'\boldsymbol{M}_j\boldsymbol{\Sigma}^{-1/2}\boldsymbol{u}$ as the "best" OLS estimators, then the same procedure as in Section 3.1 will lead to a set of moment conditions that involves $\boldsymbol{\Sigma}$.

are essentially based on the same set of moment conditions under normality.[10] Second, the II procedure estimates $\boldsymbol{\gamma}_0$ without any initial consistently estimated $\hat{\boldsymbol{G}}_i$, $\hat{\boldsymbol{E}}_i$, $\hat{\boldsymbol{R}}_j$, and $\hat{\boldsymbol{L}}_j$, but the best GMM estimator usually uses these estimates in constructing the moment conditions, as recommended by Lin and Lee (2010) and Liu et al. (2010).

If the II estimator had used simply the set of moment conditions $\boldsymbol{m}(\boldsymbol{\gamma}_0)$ that is based on only the numerator parts of $\hat{\boldsymbol{\lambda}} - \boldsymbol{\lambda}_0 - \boldsymbol{b}_{\boldsymbol{\lambda}}$ and $\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0 - \boldsymbol{b}_{\boldsymbol{\rho}}$, when there is no $\boldsymbol{X}$, one can see that, by comparing (11) and (25) in the appendix and modifying the Jacobian function accordingly, the II estimator could be as efficient as the best GMM estimator under the assumptions in Liu et al. (2010).

4. Monte Carlo Results

To assess the performance of the II estimator proposed in this paper, this section provides some Monte Carlo evidence. For comparison, the best robust GMM estimator (based on (19)) of Jin and Lee (2019) with $\boldsymbol{G}_i$, $\boldsymbol{E}_i$, $\boldsymbol{R}_j$, and $\boldsymbol{L}_j$ estimated and the GS2SLS estimator of Kelejian and Prucha (2010) are included.[11]

In Table 1, four data generating processes are considered, denoted by DGPs 1–4, respectively, corresponding to SARAR(2,2) with various degrees of spatial correlation in the outcome variable and error term. Specifically, DGP 1 has $\boldsymbol{\lambda}_0 = (0.15, 0.1)'$ and $\boldsymbol{\rho}_0 = (0.5, 0.4)'$; DGP 2 has $\boldsymbol{\lambda}_0 = (0.5, 0.4)'$ and $\boldsymbol{\rho}_0 = (0.15, 0.1)'$; DGP 3 has $\boldsymbol{\lambda}_0 = (0.15, 0.1)'$ and $\boldsymbol{\rho}_0 = (0.05, 0.02)'$; DGP 4 has $\boldsymbol{\lambda}_0 = (0.5, 0.4)'$ and $\boldsymbol{\rho}_0 = (0.3, 0.2)'$. Similar to the experimental design in Kyriacou et al. (2021), the exogenous regressors include a constant term and two independent uniform random variables on the interval $[0, 1]$. The parameter vector $\boldsymbol{\beta}_0$ is fixed at $(0.2, 0.1, -0.3)'$. Throughout, error innovations are simulated as $v_i = \sigma_i \epsilon_i$, where σ_i is drawn from a χ^2 distribution with 5 degrees of freedom and ϵ_i is i.i.d. following a student-t distribution with 5 degrees of freedom. $\boldsymbol{W}_1$

[10] When error innovations are i.i.d. and non-normal, the best moment conditions involve the skewness and kurtosis of error innovations. When both non-normality and heteroskedasticity are present, it is expected that the best moment conditions are to involve the possible heterogeneous third and fourth moments. But then in practice, how to arrive at a feasible version of the best moment conditions is a challenge.

[11] Additional experiments are conducted when $\boldsymbol{G}_i$, $\boldsymbol{E}_i$, $\boldsymbol{R}_j$, and $\boldsymbol{L}_j$ are treated as functions of the parameter vector and/or when the variance matrix $\boldsymbol{\Sigma}$, either estimated or of the true value, is used to scale these matrices. The resulting best GMM estimates perform no better. A robust GMM with moment conditions $n^{-1}[\boldsymbol{v}'\boldsymbol{Q}, \boldsymbol{v}'\boldsymbol{W}_1\boldsymbol{v}, \cdots, \boldsymbol{v}'\boldsymbol{W}_p\boldsymbol{v}, \boldsymbol{v}'(\boldsymbol{W}_1^2 - \text{Dg}(\boldsymbol{W}_1^2))\boldsymbol{v}, \cdots, \boldsymbol{v}'(\boldsymbol{W}_p^2 - \text{Dg}(\boldsymbol{W}_p^2))]$, where $\boldsymbol{Q} = [\boldsymbol{X}, \boldsymbol{W}_1\boldsymbol{X}_*, ..., \boldsymbol{W}_p\boldsymbol{X}_*, \boldsymbol{W}_1^2\boldsymbol{X}_*, ..., \boldsymbol{W}_p^2\boldsymbol{X}_*]$ with $\boldsymbol{X}_*$ denoting the part of $\boldsymbol{X}$ excluding the constant term, is also tried and its bias and RMSE performances are comparable with the best robust GMM, but its size performance is much worse. To save space, these results are not reported but are available upon request. For the GS2SLS estimator of Kelejian and Prucha (2010), the matrix of instrumental variables is $\boldsymbol{Q}$ in the first step and $n^{-1}[\boldsymbol{v}'\boldsymbol{P}_1\boldsymbol{v}, \boldsymbol{v}'\boldsymbol{P}_2\boldsymbol{v}]$, where $\boldsymbol{P}_1 = [\boldsymbol{M}_1, \cdots, \boldsymbol{M}_q]$ and $\boldsymbol{P}_2$ consists of $\boldsymbol{M}_j'\boldsymbol{M}_i - \text{Dg}(\boldsymbol{M}_j'\boldsymbol{M}_i)$ for $j \leq i \leq q$, is used as the set of moment conditions in the second step.

in the outcome equation is a distance-based matrix with its (i,j)-th entry defined as $\exp(-|l_i - l_j|)\mathbb{1}(|l_i - l_j| < \log(n))$, where the locations $l_1, \cdots, l_n$ are n random numbers from a uniform distribution on the interval $[0, n]$. The resulting matrix is then normalized by its $||\cdot||_\infty$ norm. $\boldsymbol{W}_2$ is constructed as a random symmetric matrix of zeros and ones with the number of ones restricted to be 10% of the total entries. It is then row normalized. The two matrices in the error process are generated as first-order and second-order contiguity matrices, respectively. Namely, $\boldsymbol{M}_1$ has its (i,j)-th entry equal to 1 if $|i-j| = 1$ and zero otherwise; $\boldsymbol{M}_2$ has its (i,j)-th entry equal to 1 if $|i-j| = 2$ and zero otherwise. Both are then row-normalized. Under these configurations, Assumption 2 and Assumption 4.(ii) are satisfied for the four DGPs.

Insert Table 1 here.

The bias, root mean squared error (RMSE), and empirical rejection probability of the two-sided 5% t-test of the specific parameter equal to its true value, denoted by $P(5\%)$ in Table 1, are based on 1000 simulations. To save space, only results related to the spatial parameters are reported. The following observations can be made. First, the GMM and II methods estimate the SAR parameters with little bias. No one seems to universally dominate the other. The GS2SLS estimator substantially overestimates λ_{01} across all DGPs. Second, in terms of RMSE, the II procedure delivers the smallest values in all cases, whereas the GS2SLS method gives the highest values in many cases. However, under DGPs 2 and 3, when the spatial correlation in the error process is mild, the GMM method reports the highest RMSE in estimating $\boldsymbol{\rho}_0$. Third, in terms of hypothesis testing, the II method has the best size performance, with its empirical size close to the nominal size, especially when $n = 200$. GMM also performs as well as II when $n = 200$, though there are significant size distortions when n is relatively small. The GS2SLS, however, can have serious size distortions under DGPs 1 and 4, where the degree of spatial correlation in the error term is relatively strong. Recall that in the first step of GS2SLS, the moment conditions do not necessarily take into account the spatial correlation in the error process. This kind of disconnection between its first and second steps may help explain its poor performances in these situations. Under DGPs 2 and 3 when the degree of spatial correlation in the error term is relatively weak, the GS2SLS-based t-test has reasonable size performance when $n = 100, 200$ for testing λ_{02}, ρ_{01}, and ρ_{02}, but it has trouble with testing λ_{01}. The supplementary appendix contains more simulation results, where the scenarios of strong exogenous regressors, many covariates, and a dense $\boldsymbol{W}_2$ matrix are considered. The superb performance of the II estimator is again documented across different situations.

As discussed in the previous section, the II procedure relies on a key assumption for identification, namely, Assumption 6. For a given sample, one may numerically verify that the sample binding function is invertible by checking the sample Jacobian. Figure 1 plots under each DGP and n, the cumulative distribution function (CDF) across 1000 simulations of the maximum (over the parameter space of $\boldsymbol{\gamma}$) inverse error bound of $\boldsymbol{\Psi}(\boldsymbol{\gamma})$, denoted by $\max \varepsilon(\boldsymbol{\Psi}^{-1}(\boldsymbol{\gamma}))$.[12] It evidently reveals that there is no problem of numerically inverting the sample binding function under each n-DGP combination, up to 6 to 9 decimal digits of accuracy. This helps explain that the II method does not encounter numerical failure in the simulations that lead to Table 1.

Insert Figure 1 here.

Recall that the GMM method searches over a $(p+q+k_x)$-dimensional parameter space, whereas the II approach searches over a $(p+q)$-dimensional space. So in terms of numerical cost, the GMM estimator may be preferred in situations of large n and small k_x, but the II approach may be preferred in cases of moderate n but large k_x. Figure 2 plots the average (out of 1000 simulations) ratio of computation time from best robust GMM over that from II across the four DGPs with $k_x = 2, 3, \cdots, 31$ and $n = 50, 100, 200$. Figure 3 plots the average time ratio when $n = 200, 400, 600, \cdots, 2000$ while $k_x = 3, 5, 7$.[13] All the non-constant covariates are simulated as independent uniform random variables on the interval $[0, 1]$ and $\boldsymbol{\beta}_0 = (0.2, 0.1, -0.3, 0.1, -0.3, \cdots)'$. (So for example, $\boldsymbol{\beta}_0 = (0.2, 0.1)'$ when $k_x = 2$, $(0.2, 0.1, -0.3)'$ when $k_x = 3$, $(0.2, 0.1, -0.3, 0.1)'$ when $k_x = 4$, and so on.) They indicate clearly the strength, in terms of computation time, of the II approach relative to the GMM method when the number of covariates in the outcome equation increases. When the sample size is relatively large (and k_x is small), the GMM approach may start to have some advantage relative to the II method. With $k_x = 7$, however, even under $n = 2000$, the GMM approach still takes longer than II. It is interesting to observe from Figure 3 that when n is not too large (≤ 400), the time ratio of GMM relative to II actually goes up as n increases. This suggests the extra nonlinearity in parameters appearing in the matrices in the quadratic forms that form the sample binding functions from II may not be of a practical concern in such cases. The GMM method is still

[12] The search over the parameter space is conducted by a 4-round adaptive grid search with $|\lambda_1| + |\lambda_2| < 1$ and $|\rho_1| + |\rho_2| < 1$ imposed. For each element (λ_1, λ_2, ρ_1, and ρ_2), the initial grid is $[-0.99, -0.98, -0.9, -0.6, -0.3, 0, 0.3, 0.6, 0.9, 0.98, 0.99]$, consisting of 11 grid points with search interval of 1.98. In the first round of finer grid search, the search interval is halved. After 4 rounds, the search interval is $1.98/2^4 = 0.12375$ with 11 grid points of grid length 0.012375.

[13] All the simulation results in this section and the Supplementary Appendix reported in various tables and Figure 1 are based on parallel computation (`parfor` over 1000 simulations) in Matlab 2021b. To precisely capture computation time per simulated sample for each estimator, sequential computation (`for` over 1000 simulations) is used to produce Figures 2 and 3.

nonlinear in nature, in spite of using estimates of these matrices in forming the sample quadratic moment conditions. The nonlinearity, compounded with extra computational burden (due to the extra search over the k_x beta parameters) from GMM, may increase in n at a faster rate compared with II when n is not too large.

Insert Figure 2 here.

Insert Figure 3 here.

5. An Empirical Study

This section provides an empirical study by fitting SARAR models to Airbnb log prices in Asheville, North Carolina in the United States. Bao and Liu (2021) used a data set collected on March 21, 2020, consisting of 2247 observations, to estimate a SARAR(1,1) and concluded that the SAR parameter is very significant in the outcome equation but the SAR parameter in the error process is not.

The study of Bao and Liu (2021) is not issue-free, however. First, the variable indicating free parking has an estimated coefficient with a negative sign. When one is to conduct a one-tailed test of a value of zero against the alternative of a negative value for this coefficient, the evidence is to reject the null at 5%. The calculated marginal effects also indicate a rejection of the null. This is quite counter-intuitive. Second, historical data at Acuweather.com reveal that on March 20, the high temperature was 79F (and 73F on March 19). A typical guest was more likely to pay a price premium for a room/house with air conditioning (AC) when searching around in advance. This AC variable was not included in their study. Third, out of the 2247 listings, only 30 did not provide Wi-Fi. Given that a constant term was included in the set of exogenous regressors, this would create a high degree of multicollinearity.

This study tries to address these issues. Note that free parking on premises and free street parking were categorized as a single free parking variable in Bao and Liu (2021). Usually, street parking indicates non-existence of dedicated parking space and there is a risk of car theft or vandalism at night, whereas parking on premises, either in garage or inside some gated community, is generally considered safer. Out of the 2247 listings, only 73 did not provide free parking. It is not clear from the data source whether this indicates parking was not free or there was simply no parking available. Thus, a dummy variable "PremisePk" is created in this study for free parking on premises to capture possible price premium of reserved parking space relative to either street parking or no parking. This study includes an AC dummy variable and excludes the 30 observations with no Wi-Fi, leading to in total 2217 observations.

Moreover, instead of a single spatial weight matrix in terms of the number (20, 50, or 100) of nearest neighbors (for instance, a J-nearest neighbor weight matrix is constructed such that for a unit, the J nearest units, in terms of geographic distances, are considered as this unit's neighbors), five possible distance rings are used to better trace and quantify the effects of neighbors of different distances. More specifically, a $(J_1, J_2]$-nearest weight matrix, $J_1 < J_2$, denoted by $\boldsymbol{W}_{(J_1,J_2]}$ is defined such that for a unit, if some other units are considered as its J_2-nearest neighbors but not its J_1-nearest neighbors, then the corresponding entries in this matrix are non-zero. This kind of weight matrix classification based on distance / neighborhood rings is used in Kolympiris et al. (2011) and Gupta and Robinson (2015). By definition, the resulting weight matrices do not overlap, which is recommended by Elhorst et al. (2012) in consideration of stationary parameter region in the estimation of higher-order models. In total 36 SARAR(p, q) models are tried, $p = 0, \cdots, 5$, $q = 0, \cdots, 5$, where $\boldsymbol{W}_1 = \boldsymbol{W}_{(J_0,J_1]}$, $\cdots$, $\boldsymbol{W}_p = \boldsymbol{W}_{(J_{p-1},J_p]}$, $J_0 = 0$, $J_1 = 20$, $J_2 = 50$, $J_3 = 100$, $J_4 = 200$, and $J_5 = 300$, are used in the outcome equation and the same matrices are used as $\boldsymbol{M}$'s in the error process. (When $p = q = 0$, it is a simple linear regression model and the OLS estimator is used.) Therefore, for example, $\boldsymbol{W}_{(20,50]}\boldsymbol{y}$ in the outcome equation corresponds to the spatially lagged dependent variable collecting neighbors that are from 21-nearest to 50-nearest distances. One could have included more distance rings. Nevertheless, it is quite reasonable that when a consumer or host is searching around, he/she is unlikely to compare more than 300 listings that are nearby.

The date of March 21, 2020 was about two weeks after COVID-19 was declared as a national emergency. The State of North Carolina identified the first case of COVID-19 on March 3 and the governor issued an executive order declaring a state of emergency on March 10 and two days later, the county of Buncombe, where Asheville is the county seat, declared a local state of emergency. The data on March 21, 2020 were collected at the early stage of the pandemic and travelers and hosts might not have taken into full consideration the pandemic. To investigate the possible effects of this pandemic on the Airbnb economy, a new data set was collected on July 10, 2021, when the nation evolved into the second year of the pandemic and the highly contagious Delta variant was surging. With 8 listings of shared room and 16 listings of no Wi-Fi removed, this new data set comprises 2102 observations.

Insert Table 2 here.

Table 2 reports the sample statistics of both data sets.[14] Notably, the listing price on average has been going up, largely in line with the somewhat high inflation rate during the pandemic. Host count jumps substantially, which may be explained by the eviction moratorium set up by the Centers for Disease Control and Prevention (CDC).[15] So does the proportion of listings of entire home/house, and consequently variables like the numbers of people that can be accommodated, bedrooms, and bathrooms also go up. The proportions of listings with AC (which typically does not circulate fresh air from outside) and breakfast both go down, together with that of instant bookable rentals. This may suggest that in the second year of the pandemic people are more aware of the severity of COVID-19 and become more cautious. On the other hand, more rental properties provide TV and the number of minimum nights goes up. This suggests that once guests are admitted, hosts may prefer that they stay more inside and longer, hopefully to reduce the chance of exposure to COVID. Finally, while the average review score stays about the same, the number of reviews goes up substantially with much more variation, coinciding with the observed phenomena of various kinds of misinformation and disinformation during the pandemic.

There are possibly 36 SARAR(p, q) specifications for each data set. For a model with a general variance structure, Shi and Tsai (2002) suggested using the so-called residual information criterion (RIC) for the purpose of model selection. In the framework of SARAR, since $\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} = \boldsymbol{R}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{v}$, where $\mathrm{Var}(\boldsymbol{v}) = \boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Sigma}_0$ by defining $\sigma^2 = n^{-1}\sum_{i=1}^{n}\sigma_i^2$, the RIC of a SARAR$(p, q)$ estimated by II is

$$\mathrm{RIC} = (n - k_x)\log(\tilde{\sigma}^2) + \log|\hat{\boldsymbol{\Sigma}}_0| + k_x\log(n) - k_x + \frac{4}{n - k_x - 2},$$

where $\tilde{\sigma}^2 = n^{-1}\sum_{i=1}^{n}\hat{\boldsymbol{u}}'\hat{\boldsymbol{R}}_{i\circ}'\hat{\boldsymbol{R}}_{i\circ}\hat{\boldsymbol{u}}$, $\hat{\boldsymbol{u}} = \boldsymbol{S}(\hat{\boldsymbol{\lambda}}_{II})\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}_{II}$, $\hat{\boldsymbol{\Sigma}}_0 = \mathrm{Dg}(\hat{\boldsymbol{u}}'\hat{\boldsymbol{R}}_{1\circ}'\hat{\boldsymbol{R}}_{1\circ}\hat{\boldsymbol{u}}, \cdots, \hat{\boldsymbol{u}}'\hat{\boldsymbol{R}}_{n\circ}'\hat{\boldsymbol{R}}_{n\circ}\hat{\boldsymbol{u}})/\tilde{\sigma}^2$, and $\hat{\boldsymbol{R}} = \boldsymbol{R}(\hat{\boldsymbol{\rho}}_{II})$ with its i-th row denoted by $\hat{\boldsymbol{R}}_{i\circ}$.[16] Based on RIC, SARAR$(3, 0)$ is initially

[14] They are retrieved from a third-party website http://insideairbnb.com/ that provides data collected from publicly available information at https://www.airbnb.com. Note that review scores in the data set of July 10, 2021 were recorded on a 5-pioint scale and they are converted to scores on a 100-point scale, to be consistent with those from the data set of March 21, 2020.

[15] The eviction moratorium does not explicitly apply to Airbnb rentals. Court rulings in several states like California and New York treat Airbnb hosts as landlords. Rosendahl (2017) argued that Airbnb hosts should be treated like innkeepers but not landlords. Airbnb announced on June 15, 2021 that from July 1 it would ban landlords from listing on its site if they evict a nonpaying tenant. On June 24, 2021, the CDC extended the eviction moratorium that was scheduled to expire on June 30, 2021 to July 31, 2021.

[16] Another measure for model comparison may be constructed from the predictor vector in Kelejian and Piras (2011), which can be calculated as $\hat{\boldsymbol{y}} = \boldsymbol{S}^{-1}(\hat{\boldsymbol{\lambda}}_{II})\boldsymbol{X}\hat{\boldsymbol{\beta}}_{II}$. Then the root mean squared prediction error (RMSPE) is $\sqrt{(\boldsymbol{y} - \hat{\boldsymbol{y}})'(\boldsymbol{y} - \hat{\boldsymbol{y}})/n}$. Note that this measure does not directly take into consideration possible spatial dependence in the error process and the number of exogenous regressors. A rigorous treatment of model selection in the spatial framework is beyond the scope of this paper.

chosen for the data of March 21, 2020 and SARAR$(2,2)$ for July 10, 2021 when the II method is used to estimate SARAR(p,q) models, out of the 36 (p,q) combinations.[17]

Table 3 presents the estimation results for the 2020 data, together with the marginal effects of the covariates. (The absolute values of t-ratios are inside brackets.) The estimated effects of error innovations on the dependent variable are also reported. (See the appendix for more details on calculation of the average direct impact (ADI), average indirect impact (AII), and average total impact (ATI) and the associated standard errors.)

Insert Table 3 here.

The SARAR(3,0), or SAR(3), fitted to the data of March 21, 2020 suggests that there exists spatial correlation between the listing price of a unit and those of its 20-nearest neighbors as well as those of its $(50,100]$-nearest neighbors, but the correlation between it and those from the $(20,50]$ ring is very weak and negative.[18] This kind of correlation pattern cannot be revealed by a single distance ring specification as in Bao and Liu (2021). The Moran-type statistic (16) reports a p-value of 0.84166 (0.74625) when $\boldsymbol{W}_{(100,200]}$ ($\boldsymbol{W}_{(200,300]}$) is used as the test matrix, suggesting that this SAR(3) specification appears to be enough to capture spatial correlation embedded in the data. Given that the estimate of $\boldsymbol{W}_{(20,50]}\boldsymbol{y}$ is very insignificant, a SAR(2) model is also used to fit the data by using $\boldsymbol{W}_{(0,50]}$ and $\boldsymbol{W}_{(50,100]}$ as two distance-ring based weight matrices. Under this specification, the p-values from the Moran-type statistic (16) are 0.83424 and 0.94412, corresponding to $\boldsymbol{A} = \boldsymbol{W}_{(100,200]}$ and $\boldsymbol{A} = \boldsymbol{W}_{(200,300]}$, respectively. So this also gives support of adequacy of this seemingly different specification. The estimated SAR parameters are somewhat different, but the estimated marginal effects of the various covariates, which are of ultimate interests, are largely consistent across the two specifications. Important factors include the property type, number of bathrooms, and availability of TV, which were also found in Bao and Liu (2021) to have very significant effects. Free parking on premises has a marginally positive effect, standing in contrast to the somewhat negative effect of a lump-sum free parking variable in Bao and Liu (2021). Moreover, the AC dummy variable has a very strong

[17] The best robust GMM estimation results are quite sensitive to how the relevant matrices in the quadratic forms are estimated for the Airbnb data. The robust GMM and GS2SLS (see Footnote 11) select SAR(3) for the 2020 data. For the 2021 data, GS2LS selects SARAR(1,2) and GMM selects SARAR(5,5), which reports some counter-intuitive results. The Supplementary Appendix provides estimation results from SAR(3) for the 2020 data and SARAR(1,2) for the 2021 data for comparison purpose. One can see that GMM reports somewhat strange results for the 2021 data, including positive impact of distance to city center and huge error indirect and total impacts with small t-ratios.

[18] For the interpretation of a negative spatial coefficient, see Griffith and Arbia (2010), Elhorst et al. (2012), Kao and Bera (2016), and Griffith (2019). Basically, it demonstrates dissimilarity, typically a kind of competition or backwash effect or rebellious behavior.

effect and now it ranks the second, above the bathroom variable, suggesting the importance of taking the weather information into consideration when deciding which factors may affect rental prices.

When time moves to the second year of the pandemic, the data provide a somewhat different picture of the correlation structure and the effects of various factors. The initially chosen SARAR(2,2) gives no evidence of left-over correlation based on the Moran-type statistic (16). The estimated coefficient of $\boldsymbol{W}_{(20,50]}\boldsymbol{y}$ is very insignificant in the outcome equation and the estimated coefficient of $\boldsymbol{M}_{(20,50]}\boldsymbol{u}$ in the error process is only marginally significant. So in Tables 4 and 5 three other specifications are also reported: a SARAR(1,1) using $\boldsymbol{W}_{(0,20]}\boldsymbol{y}$ and $\boldsymbol{M}_{(0,20]}\boldsymbol{u}$, a second SARAR(1,1) using $\boldsymbol{W}_{(0,50]}\boldsymbol{y}$ and $\boldsymbol{M}_{(0,50]}\boldsymbol{u}$, and a SARAR(1,2) using $\boldsymbol{W}_{(0,20]}\boldsymbol{y}$, $\boldsymbol{M}_{(0,20]}\boldsymbol{u}$, and $\boldsymbol{M}_{(20,50]}\boldsymbol{u}$. Again, in spite of the somewhat different spatial parameter estimates, the estimated marginal effects are similar across the four specifications. They indicated that the correlation structure is not completely captured by the observable outcome variable and there exists significant correlation in the error process. While the property type, number of bathrooms, and availability of TV are still important factors, two other important factors also emerge, namely, free parking on premises and the number of bedrooms. These two variables may signal some sort of security and safety during the pandemic and thus travelers are willing to pay more. The AC variable, while still has a positive and statistically important effect, does not witness an attached price premium as high as that on March 21, 2020. This is not surprising, given that the second data set was collected during the summer season and that people may not value it that much due to concerns over internal air circulation during the pandemic. Interestingly, the number of reviews has an estimated coefficient that is only marginally negative, so does the review score, standing in stark contrast to findings based on the data of March 21, 2020. This may suggest that during the pandemic (and also after the 2020 presidential election), many people might have lost their trust in on-line reviews.

Moreover, there exists a striking difference across the estimated SARAR models fitted to the two data sets, one at the early stage and one during the summer surge of the second year, in terms of the direct impacts and indirect or spillover effects of the covariates on the price variable. For the earlier data, the direct and spillover effects are of similar magnitudes and about the same levels of significance. For the later data, the ADI is typically larger than the AII for a given covariate. For example, the earlier data suggest that a unit having AC would on average push up its listing price by about 21% and the price of neighboring units by about 25%, resulting in a total impact of more than 45%. But from the later data, on average an AC-equipped unit would hike its price by about 7%, which is statistically significant, and that of its neighbors by about only 2%, leading to a total impact around 9%.

Lastly, the marginal effects of error innovations reveal that for both data sets, listing prices can be affected substantially, both directly and indirectly, by the unobservable embodied in the error process. For the early-stage data, both the explanatory variables and error innovations produce sizable spillover effects, but they are dominantly from the unobservable during the second year, evidenced by the very high value of the estimated $\mathrm{AII}(v)$. This implies that when the economy is deep into the pandemic, the interaction in Airbnb listing prices is much more complicated, far beyond that captured in the observed prices. The market may have behaved more unpredictably at the later stage, given the chaos, frustration, uncertainty, and all kinds of information of different natures regarding the pandemic.

Insert Table 4 here.

Insert Table 5 here.

Figure S.1 in the supplementary appendix reports under each model specification for the two data sets, the cumulative distributions of the inverse error bound of $\mathbf{\Psi}(\boldsymbol{\gamma})$ around its maximum value in the relevant parameter space.[19] They indicate that Assumption 6, up to at least 9 decimal digits of accuracy, are evidently satisfied in this empirical study.

6. Concluding Remarks

This paper has proposed a new estimator that is based on the principle of indirect inference by matching the simple OLS estimator of the SAR parameters in both the outcome and error equations in $\mathrm{SARAR}(p, q)$ models with its analytical approximate expectation. This is possible when one writes down a set of sample binding functions that arise from, for each parameter in the outcome equation, an OLS regression conditional on all the other SAR parameters. For each parameter in the error equation, one can use an OLS regression conditional on all the SAR parameters. The resulting II estimator is shown to be consistent, asymptotically normal, simulation-free, and robust to unknown heteroskedasticity in error innovations. With the full set of SAR parameters estimated in $\mathrm{SARAR}(p, q)$, a Moran-type specification testing procedure,

[19] An adaptive grid search as in Footnote 12 is used to search for the maximum value.

based on the residual vector, for possible left-over spatial correlation is also proposed. Furthermore, this paper discusses how the sample binding functions are related to the best moment conditions of the GMM estimator. The II method treats the relevant matrices in the system of sample binding functions as functions of model parameters, but the GMM method usually estimates these matrices first and then formulates the best moment conditions. The relevant matrices in the sample binding functions are nonlinear in model parameters. Thus, one would expect that the GMM estimator, when using the best moment conditions with the matrices in them estimated first, may be preferred in situations when the sample size is huge but the number of exogenous regressors is small, whereas the II estimator may be preferred in cases of moderate sample sizes and a lot of exogenous regressors. Simulation results demonstrate the II estimator's good finite-sample properties, relative to the GMM and GS2SLS estimators. They also highlight its relative strength/weakness relative to GMM when one takes into account the number of covariates and sample size under study.

Note that Liu and Yang (2015) proposed a modified QML (MQML) estimation method for SAR(1) that is robust against unknown heteroskedasticity by correcting the score function of the QML under heteroskedasticity. Monte Carlo evidence in Bao et al. (2020) showed that the MQML of Liu and Yang (2015) tends to deliver substantially under-sized t-test regarding the SAR parameter when its value is relatively high in SAR(1). Liu and Yang (2015) also outlined the MQML approach for SAR(1) when the error is also SAR(1), namely, SARAR(1,1). However, the associated inference procedure is not problem-free. The author has conducted Monte Carlo experiments and found that the estimated variance matrix (of the MQML estimator) is not guaranteed to be positive definite. Checking Section 5 of Liu and Yang (2015), one can see that this problematic variance estimator can happen largely due to two reasons: (i) the numerical outer-product of gradient (OPG) to approximate the variance of estimated SAR parameters, and (ii) the estimated third moment of the error term based on the sample residuals for the estimated parameter vector pertaining to the exogenous regressors. The latter could be really unreliable. Using the sample residuals (raised to power 3) could not even guarantee the estimated variance to be consistent, unlike the White-type estimator that was used for SAR(1) in Liu and Yang (2015). This was also pointed out by Liu and Yang (2015) in their paper. They proposed OPG because the analytical variance (of SAR parameters part) involves (again) the third and fourth error moments. The MQML has yet to be extended to higher-order models and one would expect the same issues will arise when conducting inference. The II estimator proposed in this paper avoids the two complications completely.

The new estimation procedure is used to study Airbnb rental prices in the city of Asheville on two different dates, one at the early stage of the COVID pandemic and one in the second summer. It finds empirically different spatial patterns across the two data sets and reveals that during the pandemic people value more safety and much less on-line reviews and the spillover effects from the unobservable are more prominent. Note that no attempt has been made in the empirical study to identify possible factors that appear in the error innovations and potential channels that make their spillover effects more substantial than those from the observable covariates. One can think of factors like people's attitude toward the pandemic, hygiene habits, and trust in different sources of information, which may have changed substantially during the second year of the pandemic, as potential terms that can create spillover effects but are hard to observe and/or measure. It would be worthwhile in a future empirical study to consider how to incorporate these factors.

A second possible direction of future research is to combine the different consistent estimators for SARAR(p, q) models. Under heteroskedasticity and non-normality, the best optimal GMM moment conditions are unknown. The popular choice of moment conditions in applied works appears to be those based on linear and quadratic forms in $\boldsymbol{v}$, where the matrices inside the linear quadratic forms are typically functions of the non-stochastic weight matrices, but not functions of model parameters. The II estimator in this paper involves quadratic forms in $\boldsymbol{v}$ and the matrices inside them are functions of model parameters. For GS2SLS, the first-step IV moment conditions and the quadratic moment conditions in the second step can be collected together as a vector of linear and quadratic forms in $\boldsymbol{v}$. Thus, a linear combination, out of any two, or all, of the three estimators may be constructed and the asymptotic distribution of the combined estimator can be straightforwardly derived. For example,

$$\check{\boldsymbol{\theta}} = (1-\omega)\hat{\boldsymbol{\theta}}_{II} + \omega\hat{\boldsymbol{\theta}}_1, \tag{20}$$

where $\hat{\boldsymbol{\theta}}_1$ is the GMM or GS2SLS estimator. Suppose $\sqrt{n}\boldsymbol{V}_1^{-1/2}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{I})$. Obviously, $\sqrt{n}[(1-\omega)^2\boldsymbol{V} + \omega^2\boldsymbol{V}_1 + \omega(1-\omega)(\boldsymbol{C}_1 + \boldsymbol{C}_1')]^{-1/2}(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{I})$, where $\boldsymbol{C}_1 = \mathrm{Cov}(\hat{\boldsymbol{\theta}}_{\mathrm{II}}, \hat{\boldsymbol{\theta}}_1) = \mathrm{Cov}(\hat{\boldsymbol{\theta}}_{II} - \boldsymbol{\theta}_0, \hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)$. Given that $\hat{\boldsymbol{\theta}}_{II} - \boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0$ can be written in terms of linear and quadratic terms in $\boldsymbol{v}$ (see (23) in the appendix for $\hat{\boldsymbol{\theta}}_{II} - \boldsymbol{\theta}_0$), it is straightforward to derive $\boldsymbol{C}_1$. Suppose one is interested in finding a value of ω such that it minimizes the global asymptotic

MSE (AMSE), namely,

$$\check{\omega} = \arg\min_{\omega} n\mathrm{E}[(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'(\check{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)] = \frac{v - c_1}{v + m_1 - 2c_1}, \tag{21}$$

where $m_1 = \mathrm{E}[(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)'(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)]$, $c_1 = \mathrm{E}[(\hat{\boldsymbol{\theta}}_{II} - \boldsymbol{\theta}_0)'(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0)]$, and $v = \mathrm{E}[(\hat{\boldsymbol{\theta}}_{II} - \boldsymbol{\theta}_0)'(\hat{\boldsymbol{\theta}}_{II} - \boldsymbol{\theta}_0)]$. In practice, this optimal weight may be estimated by $\mathrm{tr}(\hat{\boldsymbol{V}} - \hat{\boldsymbol{C}}_1)/\mathrm{tr}(\hat{\boldsymbol{V}} + \hat{\boldsymbol{V}}_1 - 2\hat{\boldsymbol{C}}_1)$. However, it is well known in the forecasting literature that once the combination weight is estimated, it could in fact make the combined forecasting worse in finite samples, given the uncertainty introduced by the estimated weight. How to take this into account is still an open question and it can be much complicated in the spatial framework.

The spatial weight matrices are assumed to be given in this paper. In stead of higher-order models, Zhang and Yu (2018) and Lam and Souza (2020) proposed combining spatial weight matrices in recognition of different choices of the weight matrices. Zhang and Yu (2018) suggested averaging across a set of candidate models based on different weight matrices. They used the Mallows C_p criterion to estimate the averaging weight vector, where the C_p criterion is constructed from the estimated candidate models and each candidate model is estimated by either ML (under homoskedasticity) or GMM (under heteroskedasticity). Lam and Souza (2020) suggested the LASSO strategy in selection of the weight matrices, where the LASSO objective function is based on some distance measure constructed using IVs. It would be interesting to explore using the II estimator in the C_p criterion and the sample binding functions in the LASSO objective function.

Table 1: Finite-Sample Performances of GMM, GS2SLS, and II in Estimating SARAR(2,2)

			GMM			GS2SLS			II		
n	DGP	$\boldsymbol{\theta}_0$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$
50	1	$\lambda_{01} = 0.15$	−0.009	0.287	24.6%	0.822	1.202	16.4%	−0.028	0.196	10.8%
		$\lambda_{02} = 0.1$	−0.014	0.178	17.2%	0.083	0.577	3.1%	−0.005	0.118	5.6%
		$\rho_{01} = 0.5$	0.032	0.154	17.3%	−0.074	0.162	17.2%	0.023	0.123	12.3%
		$\rho_{02} = 0.4$	−0.069	0.187	21.0%	−0.078	0.178	17.7%	−0.050	0.136	13.7%
	2	$\lambda_{01} = 0.5$	−0.112	0.391	16.9%	0.674	0.930	14.4%	−0.098	0.292	8.5%
		$\lambda_{02} = 0.4$	−0.079	0.303	19.4%	0.078	0.658	3.6%	−0.045	0.201	7.1%
		$\rho_{01} = 0.15$	−0.015	0.199	16.7%	−0.038	0.169	9.7%	−0.018	0.154	9.5%
		$\rho_{02} = 0.1$	−0.065	0.258	16.9%	−0.044	0.167	13.7%	−0.041	0.156	12.3%
	3	$\lambda_{01} = 0.15$	−0.027	0.489	19.7%	0.981	1.309	11.7%	−0.060	0.397	11.8%
		$\lambda_{02} = 0.1$	−0.081	0.363	19.0%	−0.127	0.758	2.9%	−0.019	0.233	7.9%
		$\rho_{01} = 0.05$	−0.020	0.207	17.7%	−0.015	0.167	9.1%	−0.022	0.163	9.6%
		$\rho_{02} = 0.02$	−0.067	0.270	21.1%	−0.041	0.171	15.6%	−0.047	0.161	14.1%
	4	$\lambda_{01} = 0.5$	−0.092	0.378	19.7%	0.645	0.905	12.4%	−0.082	0.272	9.4%
		$\lambda_{02} = 0.4$	−0.083	0.303	17.2%	−0.044	0.598	3.2%	−0.033	0.183	5.1%
		$\rho_{01} = 0.3$	−0.010	0.175	12.4%	−0.051	0.164	10.8%	−0.015	0.140	10.3%
		$\rho_{02} = 0.2$	−0.097	0.274	16.3%	−0.063	0.173	16.4%	−0.040	0.151	12.4%
100	1	$\lambda_{01} = 0.15$	0.024	0.200	15.1%	0.907	1.163	16.7%	−0.000	0.132	10.0%
		$\lambda_{02} = 0.1$	−0.006	0.149	7.3%	0.057	0.909	1.7%	0.001	0.103	5.8%
		$\rho_{01} = 0.5$	0.018	0.103	11.3%	−0.075	0.142	24.8%	0.004	0.079	9.4%
		$\rho_{02} = 0.4$	−0.039	0.121	11.8%	−0.054	0.136	14.8%	−0.017	0.085	8.4%
	2	$\lambda_{01} = 0.5$	−0.048	0.297	11.8%	0.732	0.907	15.7%	−0.062	0.207	6.2%
		$\lambda_{02} = 0.4$	−0.093	0.299	11.2%	0.091	0.911	3.9%	−0.033	0.194	4.1%
		$\rho_{01} = 0.15$	−0.007	0.132	9.9%	−0.033	0.120	7.7%	−0.002	0.101	8.9%
		$\rho_{02} = 0.1$	−0.056	0.192	11.9%	−0.029	0.123	10.1%	−0.013	0.110	10.9%
	3	$\lambda_{01} = 0.15$	−0.021	0.396	11.3%	1.057	1.289	14.9%	−0.050	0.284	7.3%
		$\lambda_{02} = 0.1$	−0.036	0.349	12.2%	−0.149	1.003	1.6%	0.005	0.228	7.5%
		$\rho_{01} = 0.05$	−0.017	0.145	10.5%	−0.018	0.125	6.6%	−0.012	0.107	8.1%
		$\rho_{02} = 0.02$	−0.049	0.195	12.8%	−0.012	0.119	9.9%	−0.014	0.102	9.4%
	4	$\lambda_{01} = 0.5$	−0.056	0.284	9.6%	0.661	0.822	17.8%	−0.062	0.188	5.4%
		$\lambda_{02} = 0.4$	−0.082	0.289	10.5%	0.198	0.918	3.5%	−0.025	0.177	3.9%
		$\rho_{01} = 0.3$	0.004	0.120	8.9%	−0.045	0.113	8.1%	0.001	0.091	8.3%
		$\rho_{02} = 0.2$	−0.057	0.184	11.3%	−0.038	0.126	10.9%	−0.015	0.097	9.1%
200	1	$\lambda_{01} = 0.15$	0.007	0.123	7.2%	1.021	1.219	16.8%	−0.005	0.092	6.8%
		$\lambda_{02} = 0.1$	−0.005	0.145	6.3%	0.148	1.426	3.5%	−0.003	0.099	4.8%
		$\rho_{01} = 0.5$	0.011	0.068	8.8%	−0.070	0.120	25.0%	0.005	0.055	7.2%
		$\rho_{02} = 0.4$	−0.021	0.077	9.3%	−0.056	0.112	18.6%	−0.011	0.059	6.5%
	2	$\lambda_{01} = 0.5$	−0.027	0.255	7.5%	0.857	1.015	18.8%	−0.051	0.169	5.8%
		$\lambda_{02} = 0.4$	−0.120	0.294	8.5%	0.156	1.333	2.4%	−0.033	0.193	3.4%
		$\rho_{01} = 0.15$	−0.002	0.097	9.7%	−0.023	0.087	6.5%	−0.002	0.075	9.6%
		$\rho_{02} = 0.1$	−0.025	0.121	9.0%	−0.021	0.087	8.8%	−0.009	0.074	7.6%
	3	$\lambda_{01} = 0.15$	−0.001	0.309	9.9%	1.141	1.382	16.9%	−0.019	0.213	7.8%
		$\lambda_{02} = 0.1$	−0.062	0.327	7.9%	−0.075	1.403	2.7%	−0.017	0.230	5.4%
		$\rho_{01} = 0.05$	−0.004	0.091	8.0%	−0.017	0.091	5.0%	−0.004	0.072	6.9%
		$\rho_{02} = 0.02$	−0.023	0.116	8.9%	−0.010	0.088	8.7%	−0.007	0.073	7.8%
	4	$\lambda_{01} = 0.5$	−0.023	0.241	9.4%	0.815	0.995	19.0%	−0.044	0.154	6.0%
		$\lambda_{02} = 0.4$	−0.093	0.266	7.1%	0.138	1.366	3.8%	−0.031	0.179	3.3%
		$\rho_{01} = 0.3$	−0.003	0.075	6.9%	−0.044	0.093	12.7%	−0.000	0.065	8.5%
		$\rho_{02} = 0.2$	−0.032	0.109	7.1%	−0.029	0.091	9.1%	−0.012	0.071	8.8%

Note: All results are based on 1000 simulations and $P(5\%)$ denotes the empirical size of the two-sided 5% t-test of the specific parameter equal to its true value.

Table 2: Summary Statistics of Airbnb Listings in Asheville, NC

	March 21, 2020		July 10, 2021		
Variable	Mean	SD	Mean	SD	Definition
log(Price)	4.748	0.683	4.995	0.655	Natural logarithm of listing price
Superhost	0.721	0.449	0.757	0.429	Host is experienced (a super host)
Hostcount	10.075	54.425	18.926	82.969	# of rentals listed by host
EnHome	0.768	0.422	0.829	0.376	Entire home/apartment
Accomm	4.091	2.638	4.424	2.765	# of people can be accommodated
Bdrms	1.591	1.272	1.820	1.083	# of bedrooms
Barms	1.360	0.733	1.473	0.767	# of bathrooms
DisCenter	3.338	2.551	3.537	2.747	Distance (in miles) to city center
PremisePk	0.606	0.489	0.550	0.498	Offer free parking on premises
AC	0.946	0.225	0.904	0.294	Offer air conditioning
TV	0.830	0.375	0.872	0.335	Offer TV
Bkfst	0.140	0.347	0.099	0.299	Offer breakfast
InsBook	0.632	0.482	0.575	0.495	Offer instant booking
MinNights	4.046	18.426	4.794	13.200	Minimum # of nights
Reviews	3.103	2.444	5.598	11.118	# of reviews per month
ReScore	97.706	3.304	97.568	5.205	Overall review scores

Table 3: SAR(3) and SAR(2) Fitted to March 21, 2020 Airbnb Data in Asheville, NC

Variable	Est.	ADI	AII	ATI	Est.	ADI	AII	ATI
$\boldsymbol{W}_{(0,20]}\boldsymbol{y}$	0.447							
	[7.939]							
$\boldsymbol{W}_{(20,50]}\boldsymbol{y}$	−0.033							
	[0.726]							
$\boldsymbol{W}_{(0,50]}\boldsymbol{y}$					0.446			
					[9.270]			
$\boldsymbol{W}_{(50,100]}\boldsymbol{y}$	0.131				0.090			
	[2.905]				[1.883]			
Constant	0.144				0.041			
	[0.366]				[0.101]			
Superhost	0.008	0.008	0.009	0.017	0.010	0.010	0.011	0.021
	[0.304]	[0.304]	[0.303]	[0.303]	[0.373]	[0.373]	[0.372]	[0.373]
Hostcount	0.001	0.001	0.002	0.003	0.001	0.002	0.002	0.003
	[5.809]	[5.824]	[4.227]	[5.129]	[5.551]	[5.551]	[4.118]	[4.972]
EnHome	0.332	0.336	0.395	0.731	0.336	0.338	0.385	0.723
	[14.748]	[14.840]	[6.048]	[9.499]	[14.870]	[14.889]	[5.954]	[9.532]
Accomm	0.079	0.080	0.094	0.173	0.083	0.083	0.095	0.178
	[7.250]	[7.250]	[4.936]	[6.306]	[7.728]	[7.729]	[4.937]	[6.516]
Bdrms	0.011	0.011	0.013	0.024	0.004	0.004	0.005	0.009
	[0.364]	[0.364]	[0.363]	[0.363]	[0.147]	[0.147]	[0.147]	[0.147]
Barms	0.182	0.185	0.217	0.401	0.187	0.188	0.215	0.403
	[8.056]	[8.069]	[5.185]	[6.838]	[8.103]	[8.106]	[5.118]	[6.848]
DisCenter	−0.036	−0.037	−0.043	−0.080	−0.040	−0.040	−0.046	−0.086
	[9.226]	[9.232]	[5.969]	[8.224]	[9.944]	[9.952]	[5.871]	[8.437]
PremisePk	0.009	0.009	0.010	0.019	0.018	0.019	0.021	0.040
	[0.482]	[0.482]	[0.483]	[0.483]	[1.025]	[1.025]	[1.022]	[1.027]
AC	0.211	0.214	0.251	0.464	0.222	0.223	0.254	0.477
	[4.875]	[4.882]	[3.997]	[4.596]	[4.921]	[4.923]	[3.985]	[4.627]
TV	0.140	0.141	0.166	0.307	0.153	0.153	0.175	0.329
	[6.362]	[6.386]	[4.833]	[5.910]	[6.955]	[6.962]	[4.978]	[6.323]
Bkfst	0.072	0.073	0.086	0.159	0.083	0.083	0.095	0.178
	[2.864]	[2.867]	[2.633]	[2.794]	[3.244]	[3.245]	[2.890]	[3.130]
InsBook	0.077	0.078	0.092	0.170	0.079	0.080	0.091	0.170
	[4.226]	[4.231]	[3.637]	[4.058]	[4.281]	[4.282]	[3.665]	[4.116]
MinNights	−0.002	−0.002	−0.002	−0.003	−0.002	−0.002	−0.002	−0.004
	[1.214]	[1.214]	[1.166]	[1.192]	[1.278]	[1.278]	[1.222]	[1.253]
Reviews	−0.041	−0.042	−0.049	−0.091	−0.043	−0.043	−0.049	−0.092
	[9.349]	[9.380]	[5.376]	[7.424]	[9.417]	[9.419]	[5.330]	[7.475]
ReScore	0.011	0.011	0.013	0.023	0.012	0.012	0.013	0.025
	[2.769]	[2.768]	[2.598]	[2.729]	[3.008]	[3.008]	[2.767]	[2.948]
ADI(v)	1.012				1.006			
	[328.160]				[834.364]			
AII(v)	1.188				1.147			
	[6.515]				[6.335]			
ATI(v)	2.200				2.153			
	[11.989]				[11.829]			

Note: The absolute values of t-ratios are inside brackets. ADI, AII, and ATI denote respectively the average direct, indirect, and total impacts on the outcome from the observable covariates. ADI(v), AII(v), and ATI(v) denote the corresponding measures from the error innovation v.

Table 4: SARAR(2,2) and SARAR(1,1) Fitted to July 10, 2021 Airbnb Data in Asheville, NC

Variable	Est.	ADI	AII	ATI	Est.	ADI	AII	ATI
$\boldsymbol{W}_{(0,20]}\boldsymbol{y}$	0.282				0.247			
	[4.257]				[2.803]			
$\boldsymbol{W}_{(20,50]}\boldsymbol{y}$	−0.044							
	[0.305]							
$\boldsymbol{W}_{(0,50]}\boldsymbol{y}$								
$\boldsymbol{M}_{(0,20]}\boldsymbol{u}$	0.319				0.480			
	[2.779]				[4.466]			
$\boldsymbol{M}_{(20,50]}\boldsymbol{u}$	0.313							
	[1.510]							
$\boldsymbol{M}_{(0,50]}\boldsymbol{u}$								
Constant	3.249				3.247			
	[4.034]				[5.620]			
Superhost	0.064	0.064	0.020	0.083	0.061	0.061	0.020	0.081
	[2.606]	[2.606]	[1.153]	[2.392]	[2.479]	[2.479]	[1.648]	[2.409]
Hostcount	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	[1.656]	[1.657]	[1.064]	[1.641]	[1.665]	[1.665]	[1.379]	[1.666]
EnHome	0.295	0.296	0.091	0.387	0.300	0.301	0.098	0.399
	[10.356]	[10.373]	[1.243]	[4.735]	[10.466]	[10.492]	[2.128]	[7.015]
Accomm	0.054	0.054	0.017	0.070	0.052	0.052	0.017	0.069
	[7.428]	[7.422]	[1.231]	[4.295]	[7.114]	[7.108]	[2.013]	[5.365]
Bdrms	0.114	0.114	0.035	0.150	0.114	0.114	0.037	0.151
	[5.783]	[5.786]	[1.213]	[3.841]	[5.764]	[5.767]	[2.016]	[4.860]
Barms	0.208	0.209	0.064	0.273	0.213	0.214	0.069	0.284
	[9.942]	[9.946]	[1.237]	[4.628]	[10.114]	[10.118]	[2.090]	[6.618]
DisCenter	−0.040	−0.041	−0.013	−0.053	−0.046	−0.046	−0.015	−0.061
	[5.062]	[5.061]	[1.215]	[3.671]	[7.532]	[7.544]	[2.099]	[5.950]
PremisePk	0.075	0.075	0.023	0.098	0.068	0.068	0.022	0.090
	[4.103]	[4.103]	[1.225]	[3.384]	[3.654]	[3.656]	[1.903]	[3.461]
AC	0.069	0.069	0.021	0.091	0.068	0.068	0.022	0.090
	[2.402]	[2.402]	[1.069]	[2.114]	[2.378]	[2.378]	[1.573]	[2.283]
TV	0.239	0.239	0.074	0.313	0.238	0.239	0.078	0.317
	[9.031]	[9.027]	[1.213]	[4.293]	[9.062]	[9.053]	[2.050]	[6.099]
Bkfst	0.086	0.087	0.027	0.113	0.083	0.083	0.027	0.110
	[2.950]	[2.950]	[1.146]	[2.564]	[2.822]	[2.821]	[1.684]	[2.667]
InsBook	0.029	0.029	0.009	0.038	0.033	0.033	0.011	0.043
	[1.578]	[1.578]	[1.051]	[1.571]	[1.773]	[1.774]	[1.450]	[1.780]
MinNights	−0.006	−0.006	−0.002	−0.007	−0.005	−0.005	−0.002	−0.007
	[2.227]	[2.226]	[1.031]	[1.958]	[2.252]	[2.252]	[1.500]	[2.146]
Reviews	−0.001	−0.001	0.000	−0.001	−0.001	−0.001	0.000	−0.001
	[1.052]	[1.052]	[0.796]	[1.027]	[0.987]	[0.987]	[0.859]	[0.968]
ReScore	−0.007	−0.007	−0.002	−0.009	−0.007	−0.007	−0.002	−0.009
	[1.476]	[1.476]	[0.909]	[1.386]	[1.541]	[1.540]	[1.187]	[1.489]
ADI(v)	1.027				1.026			
	[147.060]				[146.035]			
AII(v)	2.540				1.528			
	[2.919]				[5.135]			
ATI(v)	3.566				2.554			
	[4.070]				[8.387]			

Note: The absolute values of t-ratios are inside brackets. ADI, AII, and ATI denote respectively the average direct, indirect, and total impacts on the outcome from the observable covariates. ADI(v), AII(v), and ATI(v) denote the corresponding measures from the error innovation v.

Table 5: SARAR(1,1) and SARAR(1,2) Fitted to July 10, 2021 Airbnb Data in Asheville, NC

Variable	Est.	ADI	AII	ATI	Est.	ADI	AII	ATI
$\boldsymbol{W}_{(0,20]}\boldsymbol{y}$					0.276			
					[4.498]			
$\boldsymbol{W}_{(20,50]}\boldsymbol{y}$								
$\boldsymbol{W}_{(0,50]}\boldsymbol{y}$	0.242							
	[1.632]							
$\boldsymbol{M}_{(0,20]}\boldsymbol{u}$					0.336			
					[3.717]			
$\boldsymbol{M}_{(20,50]}\boldsymbol{u}$					0.269			
					[2.832]			
$\boldsymbol{M}_{(0,50]}\boldsymbol{u}$	0.645							
	[4.418]							
Constant	3.153				3.061			
	[3.882]				[6.085]			
Superhost	0.064	0.065	0.021	0.085	0.063	0.063	0.024	0.087
	[2.628]	[2.629]	[1.148]	[2.397]	[2.587]	[2.587]	[2.050]	[2.537]
Hostcount	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	[1.797]	[1.798]	[1.083]	[1.763]	[1.662]	[1.662]	[1.524]	[1.661]
EnHome	0.295	0.295	0.094	0.389	0.296	0.297	0.112	0.408
	[10.317]	[10.327]	[1.241]	[4.710]	[10.372]	[10.390]	[3.216]	[8.194]
Accomm	0.057	0.057	0.018	0.076	0.054	0.054	0.020	0.074
	[7.926]	[7.926]	[1.225]	[4.302]	[7.407]	[7.401]	[2.938]	[6.122]
Bdrms	0.110	0.110	0.035	0.146	0.114	0.114	0.043	0.157
	[5.595]	[5.594]	[1.209]	[3.765]	[5.770]	[5.773]	[2.917]	[5.296]
Barms	0.208	0.209	0.066	0.275	0.209	0.210	0.079	0.288
	[9.924]	[9.919]	[1.227]	[4.500]	[9.948]	[9.952]	[3.143]	[7.729]
DisCenter	−0.037	−0.038	−0.012	−0.049	−0.041	−0.041	−0.016	−0.057
	[4.549]	[4.546]	[1.167]	[3.243]	[5.454]	[5.453]	[2.784]	[4.913]
PremisePk	0.078	0.078	0.025	0.103	0.074	0.074	0.028	0.102
	[4.238]	[4.240]	[1.214]	[3.388]	[4.052]	[4.053]	[2.577]	[3.859]
AC	0.078	0.078	0.025	0.102	0.070	0.070	0.026	0.097
	[2.656]	[2.655]	[1.086]	[2.273]	[2.451]	[2.451]	[1.970]	[2.403]
TV	0.250	0.251	0.080	0.330	0.239	0.240	0.090	0.331
	[9.569]	[9.551]	[1.211]	[4.295]	[9.139]	[9.136]	[3.074]	[7.186]
Bkfst	0.090	0.090	0.029	0.119	0.086	0.087	0.033	0.119
	[3.012]	[3.011]	[1.139]	[2.573]	[2.949]	[2.949]	[2.205]	[2.869]
InsBook	0.030	0.030	0.010	0.040	0.029	0.029	0.011	0.041
	[1.609]	[1.610]	[1.055]	[1.600]	[1.592]	[1.592]	[1.497]	[1.601]
MinNights	−0.006	−0.006	−0.002	−0.007	−0.006	−0.006	−0.002	−0.008
	[2.271]	[2.270]	[1.036]	[1.987]	[2.231]	[2.230]	[1.796]	[2.169]
Reviews	−0.001	−0.001	0.000	−0.001	−0.001	−0.001	0.000	−0.001
	[0.909]	[0.909]	[0.708]	[0.884]	[1.035]	[1.035]	[0.955]	[1.020]
ReScore	−0.006	−0.006	−0.002	−0.008	−0.007	−0.007	−0.003	−0.009
	[1.406]	[1.406]	[0.881]	[1.317]	[1.475]	[1.475]	[1.298]	[1.446]
ADI(v)	1.021				1.026			
	[128.346]				[156.184]			
AII(v)	2.698				2.470			
	[2.861]				[3.680]			
ATI(v)	3.719				3.496			
	[3.910]				[5.164]			

Note: The absolute values of t-ratios are inside brackets. ADI, AII, and ATI denote respectively the average direct, indirect, and total impacts on the outcome from the observable covariates. ADI(v), AII(v), and ATI(v) denote the corresponding measures from the error innovation v.

Figure 1: Cumulative Distribution of max $\varepsilon(\mathbf{\Psi}^{-1}(\boldsymbol{\gamma}))$ under SARAR(2,2)

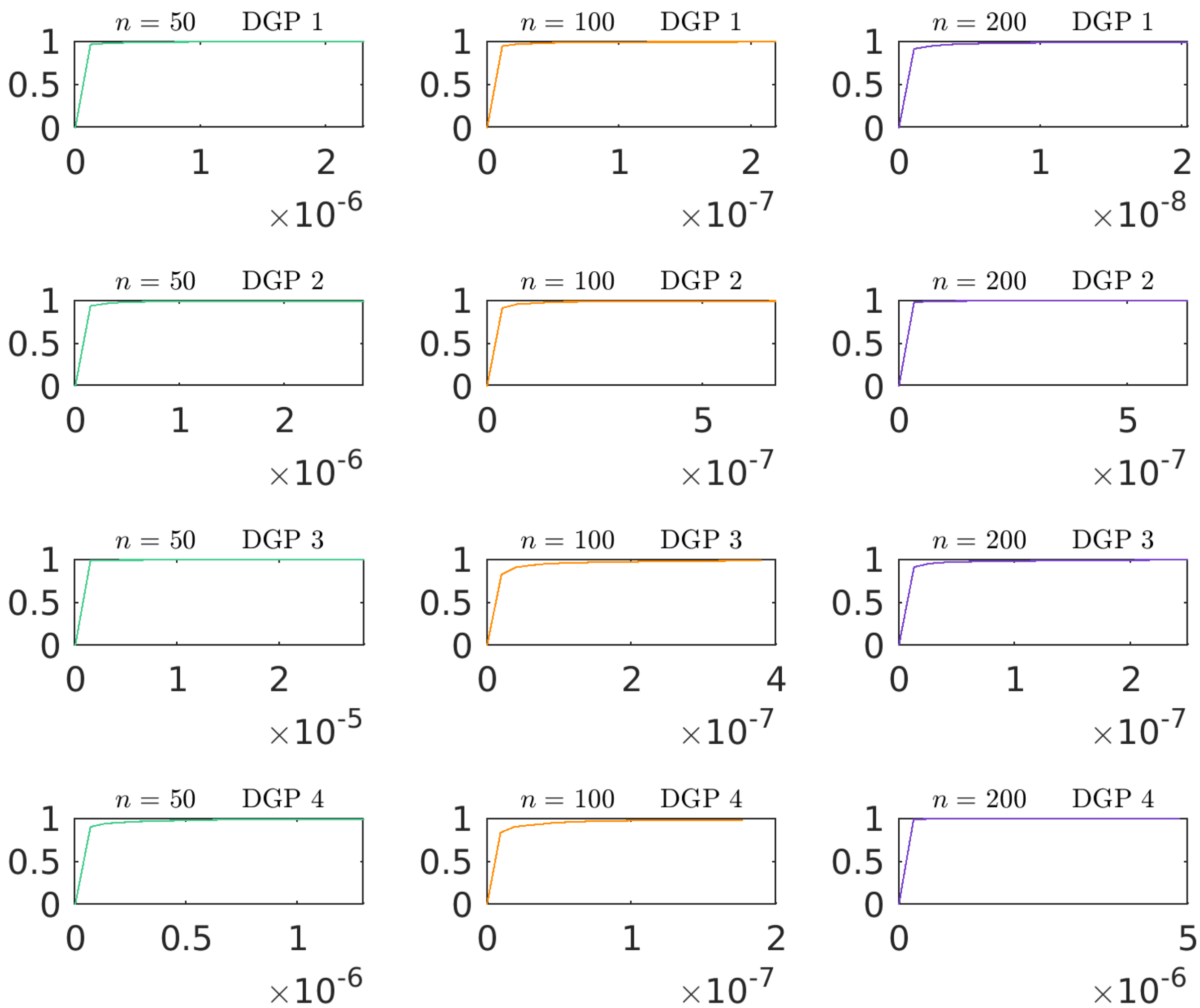


Figure 2: Time Ratio (GMM/II) in Estimating SARAR(2,2) when k_x Increases

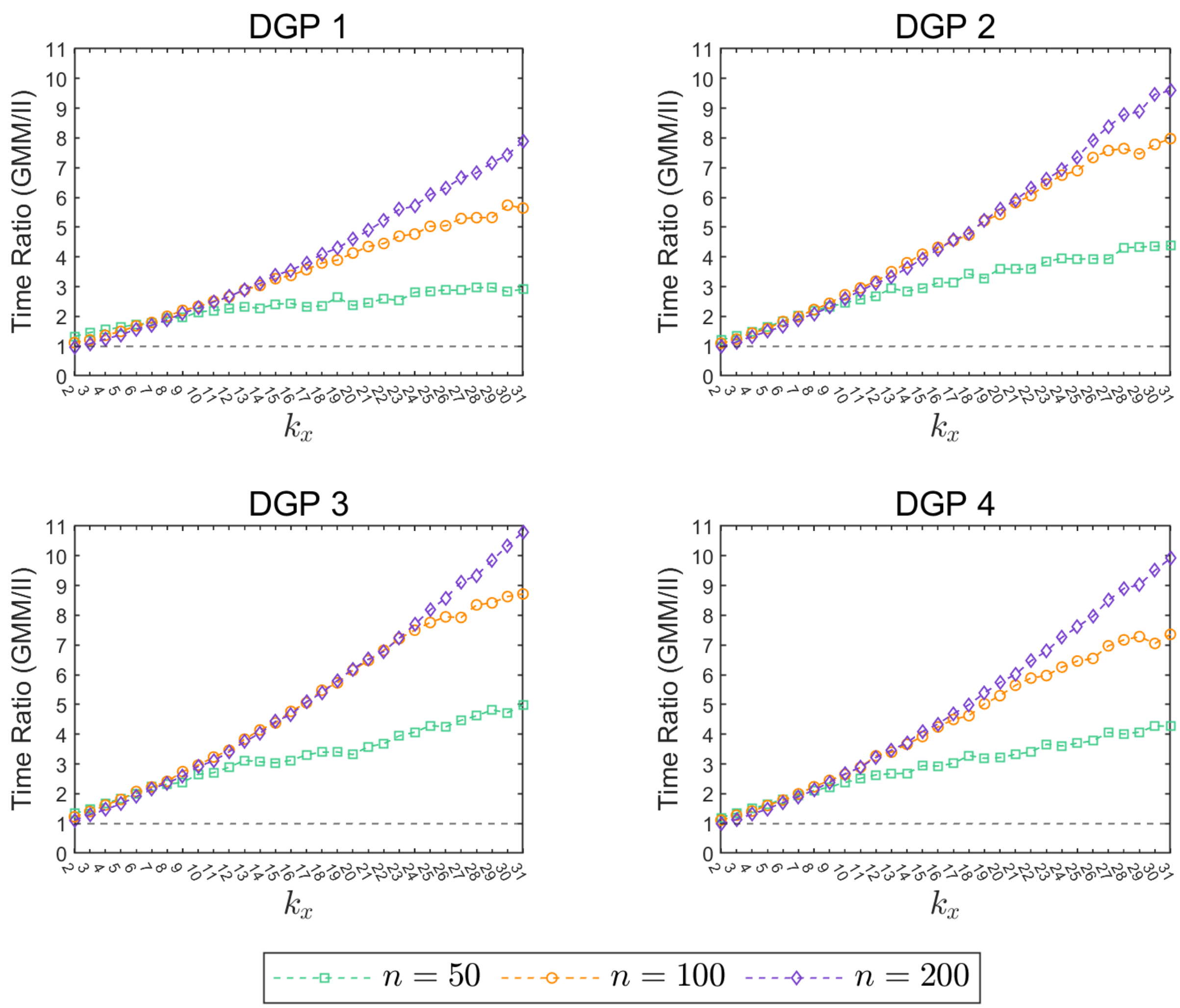
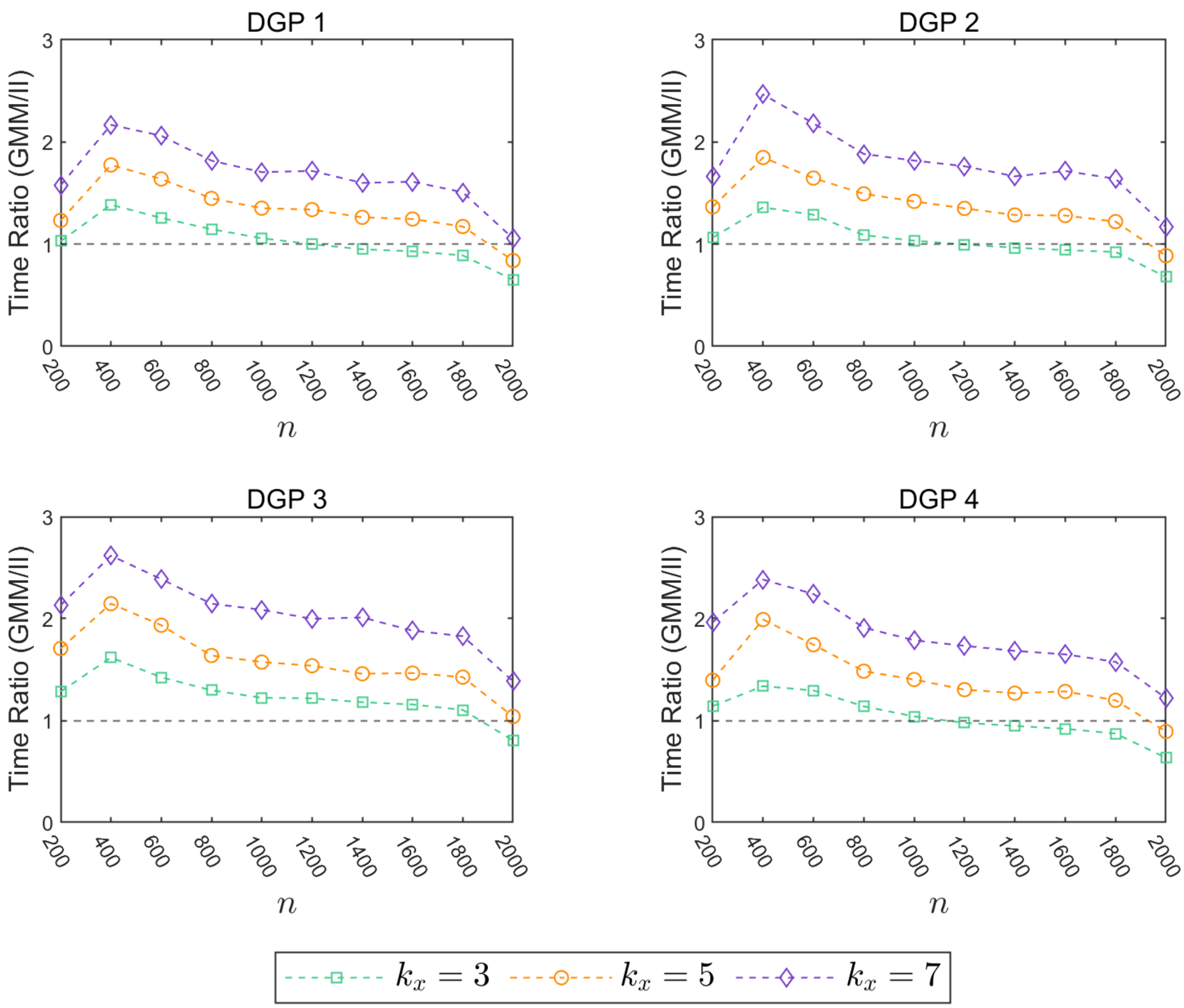


Figure 3: Time Ratio (GMM/II) in Estimating SARAR(2,2) when n Increases



Appendix

The appendix first collects some useful results that are used for the proofs of the lemmas and theorems in the paper. Lemma A.1 is given by Kelejian and Prucha (2010) and Lemma A.2 follows from Appendix A.7 of Ullah (2004). The proofs of Lemmas 2 and 3 are very similar to that of Lemma 1 and are omitted. Recall the notation $\boldsymbol{A}^{\star} = \boldsymbol{A} + \boldsymbol{A}'$.

Some Useful Results

Lemma A.1 *For $i = 1, \cdots, k$, let $\boldsymbol{A}_i$ be $n \times n$ non-random matrices with row and column sums that are bounded uniformly in absolute value, $\boldsymbol{b}_i = (b_{i1}, \cdots, b_{in})'$ be vectors of constants such that $\sup n^{-1} \sum_{j=1}^{n} |b_{ij}|^{2+\eta} < \infty$ for some $\eta > 0$, and $l_i = \boldsymbol{b}_i'\boldsymbol{v} + \boldsymbol{v}'\boldsymbol{A}_i\boldsymbol{v}$ be linear quadratic forms in $\boldsymbol{v}$, where the elements of $\boldsymbol{v}$ satisfy Assumption 3. For $\boldsymbol{l} = (l_1, \cdots, l_k)'$, if $\mathrm{Var}(\boldsymbol{l}) \succ 0$, then*

$$[\mathrm{Var}(\boldsymbol{l})]^{-1/2}[\boldsymbol{l} - \mathrm{E}(\boldsymbol{l})] \xrightarrow{d} \mathrm{N}(0, \boldsymbol{I}_k).$$

Lemma A.2 *For $\boldsymbol{v}$ with elements following Assumption 3, let $\boldsymbol{\Sigma}^{(3)} = \mathrm{Dg}(\mathrm{E}(v_1^3), ..., \mathrm{E}(v_n^3))$, $\boldsymbol{\Sigma}^{(4)} = \mathrm{Dg}(\mathrm{E}(v_1^4) - 3\sigma_1^4, ..., \mathrm{E}(v_n^4) - 3\sigma_n^4)$, and $\boldsymbol{A}$ and $\boldsymbol{B}$ be nonrandom, then*

$$\begin{aligned}
&\mathrm{E}(\boldsymbol{v}'\boldsymbol{A}\boldsymbol{v}) = \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{A}),\\
&\mathrm{E}(\boldsymbol{v}\boldsymbol{v}'\boldsymbol{A}\boldsymbol{v}) = \mathrm{dg}(\boldsymbol{\Sigma}^{(3)} \odot \boldsymbol{A}),\\
&\mathrm{E}(\boldsymbol{v}'\boldsymbol{A}\boldsymbol{v}\boldsymbol{v}'\boldsymbol{B}\boldsymbol{v}) = \mathrm{tr}(\boldsymbol{\Sigma}^{(4)} \odot \boldsymbol{A} \odot \boldsymbol{B}) + \mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{A})\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{B}) + \mathrm{tr}[\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}(\boldsymbol{B} + \boldsymbol{B}')].
\end{aligned}$$

Elements of The Jacobian and Hessian Matrices

Let $e_j(\boldsymbol{\gamma}) = \boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{\lambda})\boldsymbol{R}'(\boldsymbol{\rho})\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{F}_j'(\boldsymbol{\rho})\boldsymbol{F}_j(\boldsymbol{\rho})\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{R}(\boldsymbol{\rho})\boldsymbol{S}(\boldsymbol{\lambda})\boldsymbol{y}$ and denote

$$\begin{aligned}
\boldsymbol{G}_{i_1,i_2} &\equiv \frac{\partial \boldsymbol{G}_{i_1}(\boldsymbol{\lambda}_0)}{\partial \lambda_{i_2}} = \boldsymbol{G}_{i_1}\boldsymbol{G}_{i_2},\\
\boldsymbol{F}_{j_1,j_2} &\equiv \frac{\partial \boldsymbol{F}_{j_1}(\boldsymbol{\rho}_0)}{\partial \rho_{j_2}} = \boldsymbol{F}_{j_1}\boldsymbol{F}_{j_2},\\
\boldsymbol{H}_j &\equiv \frac{\partial \boldsymbol{H}(\boldsymbol{\rho}_0)}{\partial \rho_j} = [\boldsymbol{H}\boldsymbol{M}_j\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}']^{\star},\\
\boldsymbol{D}_{i,j} &\equiv \frac{\partial \boldsymbol{D}_i(\boldsymbol{\gamma}_0)}{\partial \rho_j} = \mathrm{Dg}[(\boldsymbol{H}_j\boldsymbol{R} - \boldsymbol{H}\boldsymbol{M}_j)\boldsymbol{G}_i\boldsymbol{R}^{-1} + \boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{F}_j],\\
d_{i,j} &\equiv \frac{\partial d_i(\boldsymbol{\rho}_0)}{\partial \rho_j} = \boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'(\boldsymbol{H}_j\boldsymbol{R} - 2\boldsymbol{H}\boldsymbol{M}_j)\boldsymbol{W}_i\boldsymbol{y},\\
e_{j_1,j_2} &\equiv \frac{\partial e_{j_1}(\boldsymbol{\gamma}_0)}{\partial \rho_{j_2}} = 2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'[\boldsymbol{H}_{j_2}\boldsymbol{F}_{j_1}'\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{R} + \boldsymbol{H}\boldsymbol{F}_{j_1}'\boldsymbol{F}_{j_1}(\boldsymbol{F}_{j_2}\boldsymbol{H}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{M}_{j_2})]\boldsymbol{S}\boldsymbol{y},\\
e_{j,(i)} &\equiv \frac{\partial e_j(\boldsymbol{\gamma}_0)}{\partial \lambda_i} = -2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{F}_j'\boldsymbol{F}_j\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}.
\end{aligned}$$

Recall that $d_i(\boldsymbol{\rho}) = \boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'(\boldsymbol{\rho})\boldsymbol{H}(\boldsymbol{\rho})\boldsymbol{R}(\boldsymbol{\rho})\boldsymbol{W}_i\boldsymbol{y}$. By observing

$$\frac{\partial \boldsymbol{S}_{(-i_1)}(\boldsymbol{\lambda}_0)}{\partial \lambda_{i_2}} = (\delta_{i_1 i_2} - 1)\boldsymbol{W}_{i_2}, \quad \frac{\partial \boldsymbol{R}_{(-j_1)}(\boldsymbol{\rho}_0)}{\partial \rho_{j_2}} = (\delta_{j_1 j_2} - 1)\boldsymbol{M}_{j_2},$$

where δ_{ij} denotes the Kronecker delta (namely, $\delta_{ij} = 1$ when $i = j$ and $\delta_{ij} = 0$ otherwise), one has

$$
\begin{aligned}
\psi_{\lambda,i_1,i_2} &= \frac{\partial \psi_{\lambda_{i_1}}(\boldsymbol{\gamma}_0)}{\partial \lambda_{i_2}} \\
&= d_{i_1}^{-1}[(\delta_{i_1 i_2}-1)\boldsymbol{y}'\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y} + 2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{D}_{i_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y} \\
&\quad - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\mathrm{Dg}(\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_{i_1,i_2}\boldsymbol{R}^{-1})\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}] - \delta_{i_1 i_2}, \\
\psi_{\lambda\rho,i,j} &= \frac{\partial \psi_{\lambda_i}(\boldsymbol{\gamma}_0)}{\partial \rho_j} \\
&= d_i^{-1}[\boldsymbol{y}'\boldsymbol{W}_i'(\boldsymbol{R}'\boldsymbol{H}_j\boldsymbol{R} - \boldsymbol{M}_j'\boldsymbol{H}\boldsymbol{R} - \boldsymbol{R}'\boldsymbol{H}\boldsymbol{M}_j)\boldsymbol{S}_{(-i)}\boldsymbol{y} \\
&\quad + \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}(2\boldsymbol{D}_i\boldsymbol{H}\boldsymbol{M}_j - 2\boldsymbol{D}_i\boldsymbol{H}_j\boldsymbol{R} - \boldsymbol{D}_{i,j}\boldsymbol{H}\boldsymbol{R})\boldsymbol{S}\boldsymbol{y}] \\
&\quad - d_i^{-2}d_{i,j}(\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}_{(-i)}\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{D}_i\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}), \\
\psi_{\rho,j_1,j_2} &= \frac{\partial \psi_{\rho_{j_1}}(\boldsymbol{\gamma}_0)}{\partial \rho_{j_2}} \\
&= e_{j_1}^{-1}\{\boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{R}'\boldsymbol{H}_{j_2} - \boldsymbol{M}_{j_2}'\boldsymbol{H})\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} \\
&\quad + \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[\boldsymbol{F}_{j_2}'\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} + (\delta_{j_1 j_2}-1)\boldsymbol{R}^{-1\prime}\boldsymbol{M}_{j_2}'\boldsymbol{F}_{j_1} + \boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1,j_2}]\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} \\
&\quad + \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}(\boldsymbol{H}_{j_2}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{M}_{j_2}) + 2\boldsymbol{K}_{j_1}\boldsymbol{H}\boldsymbol{M}_{j_2}]\boldsymbol{S}\boldsymbol{y} \\
&\quad - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[2\boldsymbol{K}_{j_1}\boldsymbol{H}_{j_2} + \mathrm{Dg}(\boldsymbol{F}_{j_1,j_2})\boldsymbol{H}]\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}\} \\
&\quad - e_{j_1}^{-2}e_{j_1,j_2}\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}(\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} - \boldsymbol{K}_{j_1})\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} - \delta_{j_1 j_2}, \\
\psi_{\rho\lambda,j,i} &= \frac{\partial \psi_{\rho_j}(\boldsymbol{\gamma}_0)}{\partial \lambda_i} \\
&= e_j^{-1}(2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{K}_j\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j)}'\boldsymbol{F}_j\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} \\
&\quad - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j)}'\boldsymbol{F}_j\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}) - e_j^{-2}e_{j,(i)}\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}(\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j)}'\boldsymbol{F}_j - \boldsymbol{K}_j)\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y},
\end{aligned}
$$

which populate, respectively, the top-left $p \times p$, top-right $p \times q$, lower-right $q \times q$, and lower-left $q \times p$ blocks of the Jacobian matrix $\boldsymbol{\Psi}(\boldsymbol{\gamma}_0)$. Essentially, all the elements of $\boldsymbol{\Psi}(\boldsymbol{\gamma}_0)$ are quadratic forms in $\boldsymbol{y} = \boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{v}$, which in term can be written as linear and quadratic forms in $\boldsymbol{v}$. Following the same strategy, one can derive all the elements of the Hessian matrix $\nabla\boldsymbol{\Psi}(\boldsymbol{\gamma}_0) = \partial\boldsymbol{\Psi}(\boldsymbol{\gamma}_0)/\partial\boldsymbol{\gamma}'$ and they are collected in the Supplementary Appendix.

Proof of Lemma 1

By substituting $\tilde{\boldsymbol{v}} = \boldsymbol{v} - \boldsymbol{R}\boldsymbol{X}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ into $\tilde{\boldsymbol{v}}'\boldsymbol{D}_i\tilde{\boldsymbol{v}}$, one has $\tilde{\boldsymbol{v}}'\boldsymbol{D}_i\tilde{\boldsymbol{v}} - \boldsymbol{v}'\boldsymbol{D}_i\boldsymbol{v} = (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{D}_i\boldsymbol{R}\boldsymbol{X}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - 2(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{D}_i\boldsymbol{v}$. Note that $\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(n^{-1/2})$, $\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{D}_i\boldsymbol{R}\boldsymbol{X} = O(n)$, $\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{D}_i\boldsymbol{v} = O_p(n^{1/2})$, and $\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y} = O_p(n)$ in view of $\boldsymbol{W}_i\boldsymbol{y} = \boldsymbol{G}_i\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{v}$ and Lemma A.2. So $\tilde{\boldsymbol{v}}'\boldsymbol{D}_i\tilde{\boldsymbol{v}} = \boldsymbol{v}'\boldsymbol{D}_i\boldsymbol{v} + O_p(1)$ and $\sqrt{n}(\hat{\lambda}_i - \lambda_{0i} - b_{\lambda_i}) = \sqrt{n}(\hat{\lambda}_i - \lambda_{0i} - \tilde{\boldsymbol{v}}'\boldsymbol{D}_i\tilde{\boldsymbol{v}}/\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}) = \sqrt{n}(\hat{\lambda}_i - \lambda_{0i} - \boldsymbol{v}'\boldsymbol{D}_i\boldsymbol{v}/\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}) + o_p(1)$. A Nagar-type (Nagar, 1959) expansion gives

$$
\begin{aligned}
\sqrt{n}\left(\hat{\lambda}_i - \lambda_{0i} - \frac{\boldsymbol{v}'\boldsymbol{D}_i\boldsymbol{v}}{\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}}\right) &= \sqrt{n}\left(\frac{r_i}{d_i} - \frac{\boldsymbol{v}'\boldsymbol{D}_i\boldsymbol{v}}{d_i}\right) \\
&= \sqrt{n}\left(\frac{r_i - \boldsymbol{v}'\boldsymbol{D}_i\boldsymbol{v}}{\mathrm{E}(d_i)}\right)\left(1 + \frac{d_i - \mathrm{E}(d_i)}{\mathrm{E}(d_i)}\right)^{-1} \\
&= \sqrt{n}\left(\frac{r_i - \boldsymbol{v}'\boldsymbol{D}_i\boldsymbol{v}}{\mathrm{E}(d_i)}\right) + o_p(1) \\
&= \frac{1}{\mathrm{E}(d_i)}\sqrt{n}(\boldsymbol{v}'\boldsymbol{E}_i\boldsymbol{v} + \boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}) + o_p(1) \\
&= \sqrt{n}s_{\lambda_i} + o_p(1),
\end{aligned}
$$

where the $o_p(1)$ term in the expansion represents

$$\underbrace{\sqrt{n}\left(\frac{r_i - \boldsymbol{v}'\boldsymbol{D}_i\boldsymbol{v}}{\mathrm{E}(d_i)}\right)}_{O_p(n^{-1/2})}\left[\underbrace{-\frac{d_i - \mathrm{E}(d_i)}{\mathrm{E}(d_i)}}_{O_p(n^{-1/2})} + \underbrace{\left(\frac{d_i - \mathrm{E}(d_i)}{\mathrm{E}(d_i)}\right)^2}_{O_p(n^{-1})} + \cdots\right].$$

These magnitudes can be verified from $d_i = \boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{v}'\boldsymbol{R}^{-1\prime}\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{v} + 2\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{v}$, and in view of Lemma A.2,

$$\begin{aligned}
\mathrm{E}(d_i) &= \mathrm{tr}(\boldsymbol{R}^{-1\prime}\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{\Sigma}) = O(n),\\
\mathrm{Var}(d_i) &= \mathrm{tr}(\boldsymbol{\Sigma}^{(4)} \odot \boldsymbol{R}^{-1\prime}\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1} \odot \boldsymbol{R}^{-1\prime}\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1})\\
&\quad + 2\mathrm{tr}[\boldsymbol{\Sigma}\boldsymbol{R}^{-1\prime}\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{\Sigma}\boldsymbol{R}^{-1\prime}\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}]\\
&\quad + 4\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{\Sigma}\boldsymbol{R}^{-1\prime}\boldsymbol{G}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{X}\boldsymbol{\beta}_0\\
&\quad + 4\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}\mathrm{dg}(\boldsymbol{\Sigma}^{(3)} \odot \boldsymbol{R}^{-1\prime}\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1})\\
&= O(n).
\end{aligned}$$

Similarly, $\mathrm{E}(r_i) = O(n)$ and $\mathrm{Var}(r_i) = O(n)$. And recall that $\mathrm{E}(r_i - \boldsymbol{v}'\boldsymbol{D}_i\boldsymbol{v}) = 0$. Thus one can claim $(r_i - \boldsymbol{v}'\boldsymbol{D}_i\boldsymbol{v})/\mathrm{E}(d_i) = O_p(n^{-1/2})$ and $[d_i - \mathrm{E}(d_i)]/\mathrm{E}(d_i) = O_p(n^{-1/2})$. Note that $\boldsymbol{E}_i$ is uniformly bounded in maximum absolute row and column sums. The elements of $\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{X}\boldsymbol{\beta}_0$ are $O(1)$ and $\boldsymbol{v}'\boldsymbol{E}_i\boldsymbol{v} + \boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}$ is in the form of linear and quadratic forms in the random vector $\boldsymbol{v}$, so is any linear combination of $\boldsymbol{v}'\boldsymbol{E}_i\boldsymbol{v} + \boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}$ and $\boldsymbol{v}'\boldsymbol{E}_j\boldsymbol{v} + \boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_j'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{v}$, $i \neq j$. Applying Lemma A.1 yields the asymptotic distribution result (3). ■

Proof of Theorem 1

Given the asymptotic distribution of $\hat{\boldsymbol{\gamma}}$ as in Lemma 3 and the definition of the II estimator, one can use the generalized delta method in Phillips (2012) to derive the asymptotic distribution of the II estimator. One needs to check a technical condition, namely, the inverse sample binding function should be asymptotically locally equicontinuous at $\boldsymbol{\gamma}_0$ almost surely. For this purpose, the following condition is sufficient: for a given $\delta > 0$, if $s \to \infty$ and $s/\sqrt{n} \to 0$,

$$\sup_{||s(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)||<\delta} \|\boldsymbol{\Psi}(\boldsymbol{\Psi}^{-1}(\boldsymbol{\gamma}) - \boldsymbol{\Psi}^{-1})\| \overset{a.s.}{\to} 0.$$

Let $\mathcal{N} = \mathcal{N}(\boldsymbol{\gamma}_0, s, \delta)$ denote the neighborhood induced by $||s(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)|| < \delta$. It is sufficient to consider $\|\boldsymbol{\Psi}(\boldsymbol{\Psi}^{-1}(\boldsymbol{\gamma}) - \boldsymbol{\Psi}^{-1})\|$, where the norm is sub-multiplicative (say, $\|\cdot\|_2$), in this neighborhood. Then

$$\begin{aligned}
\|\boldsymbol{\Psi}(\boldsymbol{\Psi}^{-1}(\boldsymbol{\gamma}) - \boldsymbol{\Psi}^{-1})\| &= \|\boldsymbol{\Psi}\boldsymbol{\Psi}^{-1}(\boldsymbol{\gamma})(\boldsymbol{\Psi} - \boldsymbol{\Psi}(\boldsymbol{\gamma}))\boldsymbol{\Psi}^{-1}\|\\
&\leq \|\boldsymbol{\Psi}\| \cdot \|\boldsymbol{\Psi}^{-1}(\boldsymbol{\gamma})\| \cdot \|\boldsymbol{\Psi} - \boldsymbol{\Psi}(\boldsymbol{\gamma})\| \cdot \|\boldsymbol{\Psi}^{-1}\|.
\end{aligned}$$

By substituting $\boldsymbol{y} = \boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{v}$ into the terms that appears in $\boldsymbol{\Psi}$, one can see that all the elements of $\boldsymbol{\Psi}$ converge to some bounded constants almost surely. To illustrate, consider ψ_{λ,i_1,i_2}:

$$\begin{aligned}
\psi_{\lambda,i_1,i_2} = (\delta_{i_1 i_2} - 1)&\frac{n^{-1}\boldsymbol{y}'\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y}}{n^{-1}d_{i_1}} + \frac{2n^{-1}\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{D}_{i_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y}}{n^{-1}d_{i_1}}\\
&- \frac{n^{-1}\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\mathrm{Dg}(\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_{i_1,i_2}\boldsymbol{R}^{-1})\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}}{n^{-1}d_{i_1}} - \delta_{i_1 i_2},
\end{aligned}$$

where the normalizing scalar n^{-1} is added. Term by term, one sees that, for instance,

$$\begin{aligned} n^{-1}\boldsymbol{y}'\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y} &= n^{-1}\boldsymbol{v}'\boldsymbol{R}^{-1\prime}\boldsymbol{S}^{-1\prime}\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{v} \\ &+ n^{-1}\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{S}^{-1\prime}\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{\beta}_0 \\ &+ 2n^{-1}\boldsymbol{v}'\boldsymbol{R}^{-1\prime}\boldsymbol{S}^{-1\prime}\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{\beta}_0. \end{aligned}$$

Notably, for the quadratic form $n^{-1}\boldsymbol{v}'\boldsymbol{R}^{-1\prime}\boldsymbol{S}^{-1\prime}\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{v}$, one can rewrite it as $(\boldsymbol{v}\boldsymbol{\Sigma}^{-1/2})'(n^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{R}^{-1\prime}\boldsymbol{S}^{-1\prime}\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{\Sigma}^{1/2})(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{v})$, where $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{v}$ is a vector of independent random variables with mean zero and variance 1. Further, one can check that the matrix $\boldsymbol{\Sigma}^{1/2}\boldsymbol{R}^{-1\prime}\boldsymbol{S}^{-1\prime}\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{\Sigma}^{1/2} \equiv \boldsymbol{C}$ is uniformly bounded in absolute row and column sums. Then $n^{-1}\boldsymbol{\Sigma}^{1/2}\boldsymbol{R}^{-1\prime}\boldsymbol{S}^{-1\prime}\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{\Sigma}^{1/2} \equiv \boldsymbol{A}$ is Hilbert-Schmidt in the sense of Varberg (1968), as $\sum_{i=1}^n\sum_{j=1}^n a_{ij}^2 \le \sum_{i=1}^n n^{-2}(\sum_{j=1}^n |c_{ij}|)^2 < \infty$. Further, $\sum_{i=1}^n |a_{ii}| < \infty$. From Theorem 1 of Varberg (1968), $(\boldsymbol{v}\boldsymbol{\Sigma}^{-1/2})'\boldsymbol{A}(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{v})$, which is $n^{-1}\boldsymbol{v}'\boldsymbol{R}^{-1\prime}\boldsymbol{S}^{-1\prime}\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{v}$, converges almost surely to a bounded constant, namely, $n^{-1}\mathrm{tr}(\boldsymbol{R}^{-1\prime}\boldsymbol{S}^{-1\prime}\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{\Sigma})$. By Kolmogorov's strong law of large numbers, the linear form $n^{-1/2}\boldsymbol{v}'\boldsymbol{R}^{-1\prime}\boldsymbol{S}^{-1\prime}\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{\beta}_0$ converges almost surely to 0 and $n^{-1}\boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{S}^{-1\prime}\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{\beta}_0 = O(1)$ by Assumption 5. So $n^{-1}\boldsymbol{y}'\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y}$ converges to a bounded constant almost surely. The same argument applies to the three terms $n^{-1}\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{D}_{i_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y}$, $n^{-1}\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\mathrm{Dg}(\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_{i_1,i_2}\boldsymbol{R}^{-1})\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}$, and $n^{-1}d_{i_1}$ that appear in ψ_{λ,i_1,i_2}. Thus one can claim that ψ_{λ,i_1,i_2} converges to a bounded constant almost surely, and similarly, so do all the other elements of $\boldsymbol{\Psi}$. It also holds for elements of $\boldsymbol{\Psi}^{-1}(\boldsymbol{\gamma})$ for $\boldsymbol{\gamma} \in \mathcal{N}$ in light of Assumption 4.(ii) and Assumption 6.(ii). Thus one can claim that $\|\boldsymbol{\Psi}\|$, $\|\boldsymbol{\Psi}^{-1}(\boldsymbol{\gamma})\|$ for $\boldsymbol{\gamma} \in \mathcal{N}$, and $\|\boldsymbol{\Psi}^{-1}\|$ are all bounded almost surely. So it is sufficient to show

$$\sup_{\boldsymbol{\gamma}\in\mathcal{N}} \|\boldsymbol{\Psi}(\boldsymbol{\gamma}) - \boldsymbol{\Psi}\| \overset{a.s.}{\to} 0.$$

Note that

$$\|(\boldsymbol{\Psi}(\boldsymbol{\gamma}) - \boldsymbol{\Psi})\| \le \left[\sup_{\boldsymbol{\gamma}^*} \|\nabla\boldsymbol{\Psi}(\boldsymbol{\gamma}^*)\|\right] \|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\|$$

where $\boldsymbol{\gamma}^* \in \mathcal{N}$ lies between $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_0$ and the Hessian matrix $\nabla\boldsymbol{\Psi}(\boldsymbol{\gamma})$ denotes the matrix derivative of $\boldsymbol{\Psi}(\boldsymbol{\gamma})$ with respect to $\boldsymbol{\gamma}$. Substituting $\boldsymbol{y} = \boldsymbol{S}^{-1}\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{S}^{-1}\boldsymbol{R}^{-1}\boldsymbol{v}$ into all the terms that appear in the Hessian matrix, one can check again that all the elements of $\nabla\boldsymbol{\Psi}$ are bounded almost surely, and for $\boldsymbol{\gamma} \in \mathcal{N}$, $\nabla\boldsymbol{\Psi}(\boldsymbol{\gamma})$ also has almost surely bounded elements in light of Assumption 4. It then follows that

$$\begin{aligned} \sup_{\boldsymbol{\gamma}\in\mathcal{N}} \|\boldsymbol{\Psi}(\boldsymbol{\gamma}) - \boldsymbol{\Psi}\| &\le \sup_{\boldsymbol{\gamma}\in\mathcal{N}} \left[\sup_{\boldsymbol{\gamma}^*} \|\nabla\boldsymbol{\Psi}(\boldsymbol{\gamma}^*)\|\right] \|\boldsymbol{\gamma} - \boldsymbol{\gamma}_0\| \\ &\le \left|\frac{\delta}{s}\right| \left[\sup_{\boldsymbol{\gamma}^*} \|\nabla\boldsymbol{\Psi}(\boldsymbol{\gamma}^*)\|\right] \\ &\overset{a.s.}{\to} 0. \end{aligned}$$

Then one can use this sufficient condition, together with Lemma A.1, to derive the asymptotic distribution (11) by following Phillips (2012). ■

Proof of Theorem 2

Substituting $\boldsymbol{R}(\hat{\boldsymbol{\rho}}_{II}) = \boldsymbol{R} - \sum_{j=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})\boldsymbol{M}_j$ and $\boldsymbol{S}(\hat{\boldsymbol{\lambda}}_{II}) = \boldsymbol{S} - \sum_{i=1}^{p}(\hat{\lambda}_{II,i} - \lambda_{0i})\boldsymbol{W}_i$, one has

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{II} = {} & [\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X} - \sum_{j=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})\boldsymbol{X}'(\boldsymbol{R}'\boldsymbol{M}_j)^{\star}\boldsymbol{X}]^{-1} \\
& \cdot [\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} - \sum_{i=1}^{p}(\hat{\lambda}_{II,i} - \lambda_{0i})\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y} \\
& - \sum_{j_2=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{M}_{j_2}\boldsymbol{S}\boldsymbol{y} \\
& - \sum_{j_1=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})\boldsymbol{X}'\boldsymbol{M}_{j_1}'\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}] + o_p(n^{-1/2}) \\
= {} & [(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1} \\
& + \sum_{j=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{R}'\boldsymbol{M}_j)^{\star}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1} \\
& + o_p(n^{-3/2})] \\
& \cdot [\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} - \sum_{i=1}^{p}(\hat{\lambda}_{II,i} - \lambda_{0i})\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y} \\
& - \sum_{j_2=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{M}_{j_2}\boldsymbol{S}\boldsymbol{y} \\
& - \sum_{j_1=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})\boldsymbol{X}'\boldsymbol{M}_{j_1}'\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}] + o_p(n^{-1/2}) \\
= {} & (\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} \\
& + \sum_{j=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{R}'\boldsymbol{M}_j)^{\star}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} \\
& - \sum_{i=1}^{p}(\hat{\lambda}_{II,i} - \lambda_{0i})(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y} \\
& - \sum_{j_2=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{M}_{j_2}\boldsymbol{S}\boldsymbol{y} \\
& - \sum_{j_1=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_{j_1}'\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} + o_p(n^{-1/2}) \\
= {} & \boldsymbol{\beta}_0 + (\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{v} \\
& + \sum_{j=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{R}'\boldsymbol{M}_j)^{\star}\boldsymbol{X}\boldsymbol{\beta}_0 \\
& + \sum_{j=1}^{q}(\hat{\rho}_{II,j} - \rho_{0j})(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{R}'\boldsymbol{M}_j)^{\star}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{v} \\
& - \sum_{i=1}^{p}(\hat{\lambda}_{II,i} - \lambda_{0i})(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}
\end{aligned}
$$

$$
\begin{aligned}
&- \sum_{j_2=1}^{q} (\hat{\rho}_{II,j} - \rho_{0j}) (\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{M}_{j_2}(\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{R}^{-1}\boldsymbol{v}) \\
&- \sum_{j_1=1}^{q} (\hat{\rho}_{II,j} - \rho_{0j}) (\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}'_{j_1}(\boldsymbol{R}\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{v}_n) + o_p(n^{-1/2}) \\
= {} & \boldsymbol{\beta}_0 + (\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{v} \\
&- \sum_{i=1}^{p} (\hat{\lambda}_{II,i} - \lambda_{0i}) (\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{G}_i(\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{R}^{-1}\boldsymbol{v}) \\
&+ o_p(n^{-1/2}) \\
= {} & \boldsymbol{\beta}_0 + (\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{v} \\
&- \sum_{i=1}^{p} (\hat{\lambda}_{II,i} - \lambda_{0i}) (\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{X}\boldsymbol{\beta}_0 + o_p(n^{-1/2}) \\
= {} & \boldsymbol{\beta}_0 + (\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{v} - \boldsymbol{J}_1(\hat{\boldsymbol{\lambda}}_{II} - \boldsymbol{\lambda}_0) + o_p(n^{-1/2}),
\end{aligned} \tag{22}
$$

where $\boldsymbol{J}_1$ stacks horizontally $(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{X}\boldsymbol{\beta}_0$, $i = 1, \cdots, p$.

From the definition of $\hat{\gamma}_{II}$, one can put $\sqrt{n}(\hat{\gamma}_{II} - \gamma_0) = \boldsymbol{\Psi}^{-1}\sqrt{n}(\hat{\gamma} - \gamma_0 - \boldsymbol{b}_{\boldsymbol{\gamma}}) + o_p(1)$. Correspondingly, $\sqrt{n}(\hat{\boldsymbol{\lambda}}_{II} - \boldsymbol{\lambda}_0) = \sqrt{n}\boldsymbol{\Psi}^{(-1)}_{\boldsymbol{\lambda\rho}}\boldsymbol{s}_{\boldsymbol{\rho}} + \sqrt{n}\boldsymbol{\Psi}^{(-1)}_{\boldsymbol{\lambda}}\boldsymbol{s}_{\boldsymbol{\lambda}} + o_p(1)$ and $\sqrt{n}(\hat{\boldsymbol{\rho}}_{II} - \boldsymbol{\rho}_0) = \sqrt{n}\boldsymbol{\Psi}^{(-1)}_{\boldsymbol{\rho}}\boldsymbol{s}_{\boldsymbol{\rho}} + \sqrt{n}\boldsymbol{\Psi}^{(-1)}_{\boldsymbol{\rho\lambda}}\boldsymbol{s}_{\boldsymbol{\lambda}} + o_p(1)$. Putting (22) and these expansions together,

$$
\hat{\boldsymbol{\theta}}_{II} - \boldsymbol{\theta}_0 = \begin{pmatrix} \boldsymbol{\Psi}^{(-1)}_{\boldsymbol{\lambda}}\boldsymbol{s}_{\boldsymbol{\lambda}} + \boldsymbol{\Psi}^{(-1)}_{\boldsymbol{\lambda\rho}}\boldsymbol{s}_{\boldsymbol{\rho}} \\ \boldsymbol{\Psi}^{(-1)}_{\boldsymbol{\rho\lambda}}\boldsymbol{s}_{\boldsymbol{\lambda}} + \boldsymbol{\Psi}^{(-1)}_{\boldsymbol{\rho}}\boldsymbol{s}_{\boldsymbol{\rho}} \\ (\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{v} - \boldsymbol{J}_1(\boldsymbol{\Psi}^{(-1)}_{\boldsymbol{\lambda}}\boldsymbol{s}_{\boldsymbol{\lambda}} + \boldsymbol{\Psi}^{(-1)}_{\boldsymbol{\lambda\rho}}\boldsymbol{s}_{\boldsymbol{\rho}}) \end{pmatrix} + o_p(n^{-1/2}). \tag{23}
$$

Recall that elements of $\boldsymbol{s}_{\boldsymbol{\lambda}}$ and $\boldsymbol{s}_{\boldsymbol{\rho}}$ are linear and quadratic forms in $\boldsymbol{v}$. This means all the components of $\hat{\boldsymbol{\theta}}_{II} - \boldsymbol{\theta}_0$ are linear and quadratic forms in $\boldsymbol{v}$. Then the asymptotic distribution of $\hat{\boldsymbol{\theta}}_{II}$ follows from Lemma A.1. ■

Scenarios of Dense Weight Matrices

Suppose that only $\boldsymbol{W}_i$ has divergent h_n. From the proof of Lemma 1, $\tilde{\boldsymbol{v}}'\boldsymbol{D}_i\tilde{\boldsymbol{v}}_i = \boldsymbol{v}'\boldsymbol{D}_i\boldsymbol{v} + (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{D}_i\boldsymbol{R}\boldsymbol{X}(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) - 2(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)'\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{D}_i\boldsymbol{v}$. Observing $\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p(1/\sqrt{n})$, $\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{D}_i\boldsymbol{R}\boldsymbol{X} = O(n/h_n)$ and $\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{D}_i\boldsymbol{v} = O_p(\sqrt{n/h_n^2})$ (since $\mathrm{Var}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{D}_i\boldsymbol{v}) = \mathrm{tr}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{D}_i\boldsymbol{\Sigma}\boldsymbol{D}_i\boldsymbol{R}\boldsymbol{X}) = O(n/h_n^2)$), one can claim that the relationship $\tilde{\boldsymbol{v}}'\boldsymbol{D}_i\tilde{\boldsymbol{v}} = \boldsymbol{v}'\boldsymbol{D}_i\boldsymbol{v}_i + O_p(1)$ still holds. Given $\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y} = O_p(n)$, one sees that $\sqrt{n}(\hat{\lambda}_i - \lambda_{0i} - b_{\lambda_i}) = \sqrt{n}(\hat{\lambda}_i - \lambda_{0i} - \boldsymbol{v}\boldsymbol{D}_i\boldsymbol{v}/d_i) + o_p(1)$ is still true and the proof of Lemma 1 goes through. It should be noted that in the special case of $\lim \sqrt{n}/h_n = 0$ (and $\lim h_n/n = 0$), i.e., h_n diverges at a rate between $\sqrt{n}$ and n, one can show that $b_{\lambda_i} = o(n^{-1/2})$ and thus the (infeasible) $\hat{\lambda}_i$ is in fact consistent. For the SAR(1) model this was pointed out by Lee (2002). Bao et al. (2020) echoed this view and emphasized that regardless of the rate of h_n, one can always recenter and it does not affect the asymptotic analysis of the resulting II estimator. It can be seen that the same conclusion holds here and Lemma 1 needs not to be modified to accommodate this special case.

If only $\boldsymbol{M}_j$ has divergent h_n, while it is still true that $\tilde{\boldsymbol{v}}'\boldsymbol{L}_j\tilde{\boldsymbol{v}} = \boldsymbol{v}'\boldsymbol{L}_j\boldsymbol{v} + O_p(1)$ and $\tilde{\boldsymbol{u}}'\boldsymbol{M}_j'\boldsymbol{M}_j\tilde{\boldsymbol{u}} = \boldsymbol{v}'\boldsymbol{F}_j'\boldsymbol{F}_j\boldsymbol{v} + O_p(1)$, now $\boldsymbol{v}'\boldsymbol{L}_j\boldsymbol{v} = O_p(\sqrt{n/h_n})$ and $\boldsymbol{v}'\boldsymbol{F}_j'\boldsymbol{F}_j\boldsymbol{v} = O_p(n/h_n)$. Then

$$
\sqrt{\frac{n}{h_n}}(\hat{\rho}_j - \rho_{0j} - b_{\rho_j}) = \sqrt{\frac{n}{h_n}}\left(\frac{\tilde{\boldsymbol{v}}'\boldsymbol{L}_j\tilde{\boldsymbol{v}}}{\tilde{\boldsymbol{u}}'\boldsymbol{M}_j'\boldsymbol{M}_j\tilde{\boldsymbol{u}}}\right)
$$

$$= \sqrt{\frac{n}{h_n}} \left(\frac{\boldsymbol{v}'\boldsymbol{L}_j\boldsymbol{v}}{\boldsymbol{v}'\boldsymbol{F}_j'\boldsymbol{F}_j\boldsymbol{v}} \right) + o_p(1)$$
$$= \sqrt{\frac{n}{h_n}} \frac{\boldsymbol{v}'\boldsymbol{L}_j\boldsymbol{v}}{\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}_j'\boldsymbol{F}_j)} + o_p(1),$$

where $s_{\rho_j} = \boldsymbol{v}'\boldsymbol{L}_j\boldsymbol{v}/\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}_j'\boldsymbol{F}_j)$ has variance $\xi_{\rho_j} = (n/h_n)\mathrm{tr}[\boldsymbol{\Sigma}\boldsymbol{L}_j\boldsymbol{\Sigma}(\boldsymbol{L}_j + \boldsymbol{L}_j')]/[\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}_j'\boldsymbol{F}_j)]^2$ with $\mathrm{tr}[\boldsymbol{\Sigma}\boldsymbol{L}_j\boldsymbol{\Sigma}(\boldsymbol{L}_j + \boldsymbol{L}_j')] = O(n/h_n)$ and $\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}_j'\boldsymbol{F}_j) = O(n/h_n)$. It follows that $\sqrt{n/h_n}\xi_{\rho_j}^{-1/2}(\hat{\rho}_j - \rho_{0j} - b_{\rho_j}) = \sqrt{n/h_n}\xi_{\rho_j}^{-1/2}s_{\rho_j} + o_p(1) \xrightarrow{d} \mathrm{N}(0,1)$. Moreover, one can show that the $\sqrt{n/h_n}$-consistent $\hat{\rho}_j - b_{\rho_j}$ is asymptotically uncorrelated with other recentered $\sqrt{n}$-consistent OLS estimator as defined in the paper. For example, consider $\sqrt{n/h_n}(\hat{\rho}_j - \rho_{0j} - b_{\rho_j})$ and $\sqrt{n}(\hat{\lambda}_i - \lambda_{0i} - b_{\lambda_i})$. Their asymptotic covariance is

$$\lim \frac{\frac{n}{\sqrt{h_n}}\mathrm{tr}[\boldsymbol{\Sigma}\boldsymbol{E}_i\boldsymbol{\Sigma}(\boldsymbol{L}_j + \boldsymbol{L}_j')]}{\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}_j'\boldsymbol{F}_j)[\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{R}^{-1\prime}\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}) + \boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{X}\boldsymbol{\beta}_0]} = 0,$$

since $\mathrm{tr}[\boldsymbol{\Sigma}\boldsymbol{E}_i\boldsymbol{\Sigma}(\boldsymbol{L}_j + \boldsymbol{L}_j')] = O(n/h_n)$, $\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}_j'\boldsymbol{F}_j) = O(n/h_n)$, and $\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{R}^{-1\prime}\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}) + \boldsymbol{\beta}_0'\boldsymbol{X}'\boldsymbol{G}_i'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{X}\boldsymbol{\beta}_0 = O(n)$. The same applies to the asymptotic covariance of $\sqrt{n/h_n}(\hat{\rho}_j - \rho_{0j} - b_{\rho_j})$ and $\sqrt{n}(\hat{\rho}_{j'} - \rho_{0j'} - b_{\rho_{j'}})$, $j' \neq j$:

$$\lim \frac{\frac{n}{\sqrt{h_n}}\mathrm{tr}[\boldsymbol{\Sigma}(\boldsymbol{L}_j + \boldsymbol{L}_j')\boldsymbol{\Sigma}(\boldsymbol{L}_{j'} + \boldsymbol{L}_{j'}')]}{\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}_j'\boldsymbol{F}_j)\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}_{j'}'\boldsymbol{F}_{j'})} = 0,$$

since $\mathrm{tr}[\boldsymbol{\Sigma}(\boldsymbol{L}_j + \boldsymbol{L}_j')\boldsymbol{\Sigma}(\boldsymbol{L}_{j'} + \boldsymbol{L}_{j'}')] = O(n/h_n)$, $\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}_j'\boldsymbol{F}_j) = O(n/h_n)$, and $\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}_{j'}'\boldsymbol{F}_{j'}) = O(n)$. (Recall that only $\boldsymbol{M}_j$ has divergent h_n.) Given all the results, one can modify Lemma 3 straightforwardly.

Suppose among the $\boldsymbol{M}$ matrices, the first t of them are dense, each characterized by $h_{n_1}, \cdots, h_{n_t}$, the other $\boldsymbol{M}$ matrices are sparse, and the $\boldsymbol{W}$ matrices may be dense or sparse, then

$$\mathrm{Dg}(\sqrt{n}, \cdots, \sqrt{n}, \sqrt{n/h_{n_1}}, \cdots, \sqrt{n/h_{n_t}}, \sqrt{n}, \cdots, \sqrt{n})\boldsymbol{\Xi}^{-1/2}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0 - \boldsymbol{b}) \xrightarrow{d} \mathrm{N}(\boldsymbol{0}, \boldsymbol{I}_{p+q}),$$

where $\boldsymbol{\Xi}$ is the same as in Lemma 3 with the following exceptions: its (j, j')-th element, $j, j' = p+1, \cdots, p+t$, is $[n/(h_{n_j}h_{n_{j'}})]\mathrm{tr}[\boldsymbol{\Sigma}\boldsymbol{L}_j\boldsymbol{\Sigma}(\boldsymbol{L}_{j'} + \boldsymbol{L}_{j'}')]/[\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}_j'\boldsymbol{F}_j)\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{F}_{j'}'\boldsymbol{F}_{j'})]$, its (i, j)-th and (j, i)-th elements, $i = 1, \cdots, p$, $j = p+1, \cdots, p+t$, are zero, and its (j, j')-th and (j', j)-th elements, $j = p+1, \cdots, p+t$,$j' = p+t+1, \cdots, p+q$, are zero. Accordingly, Theorem 1 needs to be modified, with adjusted convergence rates, by updating the matrix $\boldsymbol{V}_{\boldsymbol{\gamma}}$ with the new $\boldsymbol{\Xi}$ matrix. Finally, from the proof of Theorem 2, one sees that the asymptotic distribution of $\hat{\boldsymbol{\beta}}_{II}$ is affected by $\hat{\boldsymbol{\lambda}}_{II}$ only, so $\hat{\boldsymbol{\beta}}_{II}$ is still $\sqrt{n}$-consistent. Modification of Theorem 2 is then straightforward.

Asymptotic Distribution of GMM

In general, consider the GMM estimator based on the following set of moment conditions

$$\boldsymbol{g}(\boldsymbol{\theta}) = \frac{1}{n}(\boldsymbol{v}'(\boldsymbol{\theta})\boldsymbol{Q}, \boldsymbol{v}'(\boldsymbol{\theta})\boldsymbol{P}_1\boldsymbol{v}(\boldsymbol{\theta}), \cdots, \boldsymbol{v}'(\boldsymbol{\theta})\boldsymbol{P}_{k_p}\boldsymbol{v}(\boldsymbol{\theta}))', \tag{24}$$

where $\boldsymbol{P}_i$'s, $i = 1, \cdots, k_p$, are symmetric with $\mathrm{tr}(\boldsymbol{P}_i\boldsymbol{\Sigma}) = 0$ and $\boldsymbol{Q}$ is an $n \times k_q$ IV matrix with full column rank. Further assume that $\boldsymbol{P}_i$'s are chosen such that $\mathrm{dg}(\boldsymbol{P}_i) = \boldsymbol{0}$. (This choice obviously makes it true that $\mathrm{tr}(\boldsymbol{P}_i\boldsymbol{\Sigma}) = 0$.) In total, there are $k_g = k_q + k_p$ number of moment conditions. For the best GMM when error innovations are normal and i.i.d., $\boldsymbol{Q}$ contains

$\boldsymbol{RX}$, $\boldsymbol{RG}_1\boldsymbol{X\beta}_0$,$\cdots$,$\boldsymbol{RG}_p\boldsymbol{X\beta}_0$ and the $\boldsymbol{P}$ matrices contain $\boldsymbol{E}_1, \cdots, \boldsymbol{E}_p, \boldsymbol{L}_1, \cdots, \boldsymbol{L}_q$. Note that

$$\boldsymbol{\nabla} = \mathrm{E}\left(\frac{\partial \boldsymbol{g}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right) =$$
$$-\frac{1}{n}\left(\begin{array}{c|ccc|ccc} \boldsymbol{Q}'\boldsymbol{RX} & \boldsymbol{Q}'\boldsymbol{RG}_1\boldsymbol{X\beta}_0 & \cdots & \boldsymbol{Q}'\boldsymbol{RG}_p\boldsymbol{X\beta}_0 & \mathbf{0}_{k_q} & \cdots & \mathbf{0}_{k_q} \\ \mathbf{O}_{k_p,k_x} & 2\mathrm{tr}(\boldsymbol{P}_1\boldsymbol{A}_1) & \cdots & 2\mathrm{tr}(\boldsymbol{P}_1\boldsymbol{A}_p) & 2\mathrm{tr}(\boldsymbol{P}_1\boldsymbol{B}_1) & \cdots & 2\mathrm{tr}(\boldsymbol{P}_1\boldsymbol{B}_q) \\ & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ & 2\mathrm{tr}(\boldsymbol{P}_{k_p}\boldsymbol{A}_1) & \cdots & 2\mathrm{tr}(\boldsymbol{P}_{k_p}\boldsymbol{A}_p) & 2\mathrm{tr}(\boldsymbol{P}_{k_p}\boldsymbol{B}_1) & \cdots & 2\mathrm{tr}(\boldsymbol{P}_{k_p}\boldsymbol{B}_q) \end{array}\right),$$

where $\boldsymbol{A}_i = \boldsymbol{RG}_i\boldsymbol{R}^{-1}\boldsymbol{\Sigma}$, $i = 1, \cdots, p$, $\boldsymbol{B}_j = \boldsymbol{F}_j\boldsymbol{\Sigma}$, $j = 1, \cdots, q$, and

$$\boldsymbol{\Omega} = \mathrm{Var}(\boldsymbol{g}(\boldsymbol{\theta}_0)) = \frac{1}{n^2}\begin{pmatrix} \boldsymbol{Q}'\boldsymbol{\Sigma}\boldsymbol{Q} & \mathbf{O}_{k_q\times k_p} \\ \mathbf{O}_{k_p\times k_q} & \boldsymbol{\Omega}_{\boldsymbol{PP}} \end{pmatrix},$$

in which $\boldsymbol{\Omega}_{\boldsymbol{PP}}$ is $k_p \times k_p$ with $\mathrm{tr}(\boldsymbol{\Sigma}\boldsymbol{P}_i\boldsymbol{\Sigma}\boldsymbol{P}_j^\star)$ in its (i,j)-th position, $i, j = 1, \cdots, k_p$. Given that $\boldsymbol{P}_i$'s have zero diagonals, whether or not they involve the parameter vector $\boldsymbol{\theta}_0$ does not affect the expected gradient $\boldsymbol{\nabla}$ and the variance matrix $\boldsymbol{\Omega}$. This is because $\partial\boldsymbol{P}_i(\boldsymbol{\theta}_0)/\partial\theta_j$ also has zero diagonals and $\mathrm{E}(\boldsymbol{v}'(\partial\boldsymbol{P}_i(\boldsymbol{\theta}_0)/\partial\theta_j)\boldsymbol{v}) = \mathrm{tr}(\boldsymbol{\Sigma}(\partial\boldsymbol{P}_i(\boldsymbol{\theta}_0)/\partial\theta_j)) = 0$. For some initial consistent estimator $\tilde{\boldsymbol{\theta}}$, the feasible optimal GMM estimated, denoted by $\hat{\boldsymbol{\theta}}_{FOGMM}$, minimizes $\boldsymbol{g}'(\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}(\tilde{\boldsymbol{\theta}})\boldsymbol{g}(\boldsymbol{\theta})$ and it has the following asymptotic distribution:

$$\sqrt{n}\left(\frac{1}{n}\boldsymbol{\nabla}'\boldsymbol{\Omega}^{-1}\boldsymbol{\nabla}\right)^{1/2}(\hat{\boldsymbol{\theta}}_{FOGMM} - \boldsymbol{\theta}_0) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \boldsymbol{I}). \tag{25}$$

Partial Effects

The average partial effects of the r-th exogenous variable on the outcome variable, following LeSage and Pace (2009), can be defined as

$$\mathrm{ADI}_r = \frac{1}{n}\mathrm{tr}(\boldsymbol{S}_r), \quad \mathrm{ATI}_r = \frac{1}{n}\mathbf{1}'\boldsymbol{S}_r\mathbf{1}, \quad \mathrm{AII}_r = \mathrm{ATI}_r - \mathrm{ADI}_r, \tag{26}$$

where $\boldsymbol{S}_r = \boldsymbol{S}_r(\beta_{0r}, \boldsymbol{\lambda}_0) = \beta_{0r}\boldsymbol{S}^{-1}$. In practice, one can estimate them by replacing β_{0r} and $\boldsymbol{S}$ by $\hat{\beta}_{II,r}$ and $\boldsymbol{S}(\hat{\boldsymbol{\lambda}}_{II})$, respectively. In view of $\partial\boldsymbol{S}^{-1}(\boldsymbol{\lambda})/\partial\lambda_i = \boldsymbol{S}^{-1}(\boldsymbol{\lambda})\boldsymbol{W}_i\boldsymbol{S}^{-1}(\boldsymbol{\lambda})$, one has

$$\frac{\partial\boldsymbol{S}_r(\boldsymbol{\theta}_0)}{\partial(\boldsymbol{\lambda}', \beta_r)'} = \begin{pmatrix} \beta_{0r}\boldsymbol{S}^{-1}\boldsymbol{W}_1\boldsymbol{S}^{-1} \\ \vdots \\ \beta_{0r}\boldsymbol{S}^{-1}\boldsymbol{W}_p\boldsymbol{S}^{-1} \\ \boldsymbol{S}^{-1} \end{pmatrix} \equiv \boldsymbol{\Delta}.$$

So the standard error of ADI_r, for instance, can be calculated as

$$\frac{1}{n}\sqrt{(\mathrm{Tr}(\hat{\boldsymbol{\Delta}}))'\hat{\boldsymbol{V}}_{[1:p,p+q+r]}\mathrm{Tr}(\hat{\boldsymbol{\Delta}})},$$

where $\hat{\boldsymbol{\Delta}} = \boldsymbol{\Delta}(\hat{\boldsymbol{\lambda}}_{II}, \hat{\beta}_{II,r})$, $\mathrm{Tr}(\hat{\boldsymbol{\Delta}})$ denotes matrix trace operation applied to each $n \times n$ block of $\hat{\boldsymbol{\Delta}}$, and $\hat{\boldsymbol{V}}_{[1:p,p+q+r]}$ denotes the block of $\hat{\boldsymbol{V}}$ pertaining to positions $1, \cdots, p, p+q+r$.

Note that for SAR and SARAR, these definitions take on the same analytical form, whereas for spatial error models (namely, when $p = 0$), they are all equal to the beta parameter associated with x_r, coinciding with the typical *ceteris paribus* interpretation. They measure the effects of the observable on the outcome variable. An interesting question may also be: how would a shock to the i-th unit affect its outcome as well as the outcomes of all its neighbors? So one can define $\partial\boldsymbol{y}/\partial\boldsymbol{v}' = \boldsymbol{S}^{-1}\boldsymbol{R}^{-1}$, resembling the so-called impulse response function in time series. Then it is straightforward that the average direct impact, indirect impart, and total impact from the unobservable can be constructed and estimated, together with their standard errors. Elhorst et al. (2012) emphasized that in contrast to the first-order model, it is in general not possible to answer the question to which extent the second and higher-order effects fall on the corresponding neighborhood rings in higher-order models.

References

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht.

Badinger, H., Egger, P., 2011. Estimation of higher-order spatial autoregressive cross-section models with heteroscedastic disturbances. Papers in Regional Science. 90 (1), 213–235, 10.1111/j.1435-5957.2010.00323.x.

Bao, Y., Liu, X., 2021. Estimating a spatial autoregressive model with autoregressive disturbances based on the indirect inference principle. Spatial Economic Analysis. 16 (4), 506–529, 10.1080/17421772.2021.1902552.

Bao, Y., Liu, X., Yang, L., 2020. Indirect inference estimation of spatial autoregressions. Econometrics. 8 (3), 10.3390/econometrics8030034.

Dall'erba, S., Percoco, M., Piras, G., 2009. Service industry and cumulative growth in the regions of Europe. Entrepreneurship & Regional Development. 21 (4), 333–349, 10.1080/08985620903019815.

Elhorst, J. P., Lacombe, D. J., Piras, G., 2012. On model specification and parameter space definitions in higher order spatial econometric models. Regional Science and Urban Economics. 42 (1), 211–220, 10.1016/j.regsciurbeco.2011.09.003.

Gouriéroux, C., Monfort, A., Renault, E., 1993. Indirect inference. Journal of Applied Econometrics. 8 (S1), S85–S118, 10.1002/jae.3950080507.

Griffith, D. A., 2019. Negative spatial autocorrelation: One of the most neglected concepts in spatial statistics. Stats. 2 (3), 388–415, 10.3390/stats2030027.

Griffith, D. A., Arbia, G., 2010. Detecting negative spatial autocorrelation in georeferenced random variables. International Journal of Geographical Information Science. 24 (3), 417–437, 10.1080/13658810902832591.

Gupta, A., Robinson, P. M., 2015. Inference on higher-order spatial autoregressive models with increasingly many parameters. Journal of Econometrics. 186 (1), 19–31, 10.1016/j.jeconom.2014.12.008.

Helbich, M., Brunauer, W., Vaz, E., Nijkamp, P., 2014. Spatial heterogeneity in hedonic house price models: The case of Austria. Urban Studies. 51 (2), 390–411, 10.1177/0042098013492234.

Jin, F., Lee, L.-F., 2019. GEL estimation and tests of spatial autoregressive models. Journal of Econometrics. 208 (2), 585–612, 10.1016/j.jeconom.2018.07.007.

Kao, S. Y.-H., Bera, A. K., 2016. Spatial regression: The curious case of negative spatial dependence. Working paper, Department of Economics, University of Illinois.

Kelejian, H. H., Piras, G., 2011. An extension of Kelejian's J-test for non-nested spatial models. Regional Science and Urban Economics. 41 (3), 281–292, 10.1016/j.regsciurbeco.2011.01.002.

Kelejian, H. H., Prucha, I. R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. The Journal of Real Estate Finance and Economics. 17 (1), 99–121, 10.1023/A:1007707430416.

——— 2001. On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics. 104 (2), 219–257, 10.1016/S0304-4076(01)00064-1.

——— 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics. 157 (1), 53–67, 10.1016/j.jeconom.2009.10.025.

Kolympiris, C., Kalaitzandonakes, N., Miller, D., 2011. Spatial collocation and venture capital in the US biotechnology industry. Research Policy. 40 (9), 1188–1199, 10.1016/j.respol.2011.05.022.

Kyriacou, M., Phillips, P. C. B., Rossi, F., 2017. Indirect inference in spatial autoregression. The Econometrics Journal. 20 (2), 168–189, 10.1111/ectj.12084.

——— 2021. Continuously updated indirect inference in heteroskedastic spatial model. Econometric Theory. forthcoming, 10.1017/S0266466621000384.

Lam, C., Souza, P. C. L., 2020. Estimation and selection of spatial weight matrix in a spatial lag model. Journal of Business & Economic Statistics. 38 (3), 693–710, 10.1080/07350015.2019.1569526.

Lee, L.-F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometric Theory. 18 (2), 252–277, 10.1017/S0266466602182028.

——— 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica. 72 (6), 1899–1925, 10.1111/j.1468-0262.2004.00558.x.

——— 2007. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics. 137 (2), 489–514, 10.1016/j.jeconom.2005.10.004.

Lee, L.-F., Liu, X., 2010. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. Econometric Theory. 26 (1), 187–230, 10.1017/S0266466609090653.

LeSage, J., Pace, R. K., 2009. Introduction to Spatial Econometrics. Chapman and Hall/CRC, New York, 10.1201/9781420064254.

Lieberman, O., 1994. A Laplace approximation to the moments of a ratio of quadratic forms. Biometrika. 81 (4), 681–690, 10.1093/biomet/81.4.681.

Lin, X., Lee, L.-F., 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. Journal of Econometrics. 157 (1), 34–52, 10.1016/j.jeconom.2009.10.035.

Liu, S. F., Yang, Z., 2015. Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and nonnormality. Regional Science and Urban Economics. 52 50–70, 10.1016/j.regsciurbeco.2015.02.003.

Liu, X., Lee, L.-F., Bollinger, C. R., 2010. An efficient GMM estimator of spatial autoregressive models. Journal of Econometrics. 159 (2), 303–319, 10.1016/j.jeconom.2010.08.001.

Nagar, A. L., 1959. The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. Econometrica. 27 (4), 575–595, 10.2307/1909352.

Phillips, P. C. B., 2012. Folklore theorems, implicit maps, and indirect inference. Econometrica. 80 (1), 425–454, 10.3982/ECTA9350.

Rico-Juan, J. R., Taltavull de La Paz, P., 2021. Machine learning with explainability or spatial hedonics tools? An analysis of the asking prices in the housing market in Alicante, Spain. Expert Systems with Applications. 171, 114590, 10.1016/j.eswa.2021.114590.

Robinson, P., 2008. Correlation testing in time series, spatial and cross-sectional data. Journal of Econometrics. 147 (1), 5–16, 10.1016/j.jeconom.2008.09.001.

Rosendahl, M., 2017. iTenant: How the law should treat rental relationships in the sharing economy. William & Mary Law Review. 59 (2), 731–770, https://scholarship.law.wm.edu/wmlr/vol59/iss2/7.

Shi, P., Tsai, C.-L., 2002. Regression model selection–a residual likelihood approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 64 (2), 237–252, 10.1111/1467-9868.00335.

Smith Jr., A. A., 1993. Estimating nonlinear time-series models using simulated vector autoregressions. Journal of Applied Econometrics. 8 (S1), S63–S84, 10.1002/jae.3950080506.

Ullah, A., 2004. Finite Sample Econometrics. Oxford University Press, New York.

Varberg, D. E., 1968. Almost sure convergence of quadratic forms in independent random variable. The Annals of Mathematical Statistics. 39 (5), 1502–1506, 10.1214/aoms/1177698131.

Wilkinson, J. H., 1961. Error analysis of direct methods of matrix inversion. Journal of the ACM. 8 (3), 281–330, 10.1145/321075.321076.

Yang, C. F., 2021. Common factors and spatial dependence: An application to US house prices. Econometric Reviews. 40 (1), 14–50, 10.1080/07474938.2020.1741785.

Yang, Z., 2015. A general method for third-order bias and variance corrections on a nonlinear estimator. Journal of Econometrics. 186 178–200, 10.1016/j.jeconom.2014.07.003.

Zhang, X., Yu, J., 2018. Spatial weights matrix selection and model averaging for spatial autoregressive models. Journal of Econometrics. 203 (1), 1–18, 10.1016/j.jeconom.2017.05.021.

Supplementary Appendix to "Indirect Inference Estimation of Higher-order Spatial Autoregressive Models"

Elements of $\nabla\boldsymbol{\Psi}(\boldsymbol{\gamma}_0)$

Introduce further the following notation,

$$\boldsymbol{G}_{i_1,i_2,i_3} = \frac{\partial^2 \boldsymbol{G}_{i_1}(\boldsymbol{\lambda}_0)}{\partial\lambda_{i_2}\partial\lambda_{i_3}} = \boldsymbol{G}_{i_1}\boldsymbol{G}_{i_3}\boldsymbol{G}_{i_2} + \boldsymbol{G}_{i_1}\boldsymbol{G}_{i_2}\boldsymbol{G}_{i_3},$$

$$\boldsymbol{F}_{j_1,j_2,j_3} = \frac{\partial^2 \boldsymbol{F}_{j_1}(\boldsymbol{\rho}_0)}{\partial\rho_{j_2}\partial\rho_{j_3}} = \boldsymbol{F}_{j_1}\boldsymbol{F}_{j_3}\boldsymbol{F}_{j_2} + \boldsymbol{F}_{j_1}\boldsymbol{F}_{j_2}\boldsymbol{F}_{j_3},$$

$$\begin{aligned}
\boldsymbol{H}_{j_1,j_2} &= \frac{\partial^2 \boldsymbol{H}(\boldsymbol{\rho}_0)}{\partial\rho_{j_1}\partial\rho_{j_2}} \\
&= \{\boldsymbol{H}_{j_2}\boldsymbol{M}_{j_1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{R}' - \boldsymbol{H}\boldsymbol{M}_{j_1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}'_{j_2} \\
&\quad + \boldsymbol{H}\boldsymbol{M}_{j_1}[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}'_{j_2}\boldsymbol{R}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}\boldsymbol{X}']^\star\boldsymbol{R}'\}^\star,
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{D}_{i,j_1,j_2} &= \frac{\partial^2 \boldsymbol{D}_i(\boldsymbol{\gamma}_0)}{\partial\rho_{j_1}\partial\rho_{j_2}} \\
&= \mathrm{Dg}[(\boldsymbol{H}_{j_1,j_2}\boldsymbol{R} - \boldsymbol{H}_{j_1}\boldsymbol{M}_{j_2} - \boldsymbol{H}_{j_2}\boldsymbol{M}_{j_1})\boldsymbol{G}_i\boldsymbol{R}^{-1} \\
&\quad + (\boldsymbol{H}_{j_1}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{M}_{j_1})\boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{M}_{j_2} + (\boldsymbol{H}_{j_2}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{M}_{j_2})\boldsymbol{G}_i\boldsymbol{R}^{-1}\boldsymbol{F}_{j_1} \\
&\quad + \boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_i\boldsymbol{R}^{-1}(\boldsymbol{M}_{j_2}\boldsymbol{F}_{j_1} + \boldsymbol{F}_{j_1,j_2})],
\end{aligned}$$

$$\begin{aligned}
e_{j_1,j_2,j_3} &\equiv \frac{\partial^2 e_{j_1}(\boldsymbol{\gamma}_0)}{\partial\rho_{j_2}\partial\rho_{j_3}} = 2[\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}_{j_2,j_3}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{M}'_{j_3}\boldsymbol{H}_{j_2}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} \\
&\quad + \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}_{j_2}(\boldsymbol{F}'_{j_1,j_3}\boldsymbol{F}_{j_1})^\star\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}_{j_2}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1}\boldsymbol{H}_{j_3}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} \\
&\quad - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}_{j_2}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{M}_{j_3}\boldsymbol{S}\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{M}'_{j_3}\boldsymbol{H}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{M}_{j_2}\boldsymbol{S}\boldsymbol{y} \\
&\quad - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}_{j_3}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{M}_{j_2}\boldsymbol{S}\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}(\boldsymbol{F}'_{j_1,j_3}\boldsymbol{F}_{j_1})^\star\boldsymbol{H}\boldsymbol{M}_{j_2}\boldsymbol{S}\boldsymbol{y} \\
&\quad - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1}\boldsymbol{H}_{j_3}\boldsymbol{M}_{j_2}\boldsymbol{S}\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{M}'_{j_3}\boldsymbol{H}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1,j_2}\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} \\
&\quad + \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}_{j_3}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1,j_2}\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{F}'_{j_1,j_3}\boldsymbol{F}_{j_1,j_2}\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} \\
&\quad + \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1,j_2,j_3}\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1,j_2}\boldsymbol{H}_{j_3}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} \\
&\quad - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1,j_2}\boldsymbol{H}\boldsymbol{M}_{j_3}\boldsymbol{S}\boldsymbol{y}],
\end{aligned}$$

$$\begin{aligned}
e_{j_1,j_2,(i)} &\equiv \frac{\partial^2 e_{j_1}(\boldsymbol{\gamma}_0)}{\partial\rho_{j_2}\partial\lambda_i} = -2\boldsymbol{y}'\boldsymbol{W}'_i\boldsymbol{R}'(\boldsymbol{H}_{j_2}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{M}_{j_2} \\
&\quad + \boldsymbol{H}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1,j_2}\boldsymbol{H}\boldsymbol{R})\boldsymbol{S}\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'(\boldsymbol{H}_{j_2}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{M}_{j_2} \\
&\quad + \boldsymbol{H}\boldsymbol{F}'_{j_1}\boldsymbol{F}_{j_1,j_2}\boldsymbol{H}\boldsymbol{R})\boldsymbol{W}_i\boldsymbol{y}.
\end{aligned}$$

Upon taking another round of derivatives, one has

$$\begin{aligned}
\psi_{\lambda,i_1,i_2,i_3} &= \frac{\partial^2 \psi_{\lambda_{i_1}}(\boldsymbol{\gamma}_0)}{\partial\lambda_{i_2}\partial\lambda_{i_3}} \\
&= d_{i_1}^{-1}[2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\mathrm{Dg}(\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_{i_1,i_2}\boldsymbol{R}^{-1})\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_3}\boldsymbol{y} \\
&\quad - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\mathrm{Dg}(\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_{i_1,i_2,i_3}\boldsymbol{R}^{-1})\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{W}'_{i_3}\boldsymbol{R}'\boldsymbol{H}\boldsymbol{D}_{i_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y} \\
&\quad + 2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\mathrm{Dg}(\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_{i_1,i_3}\boldsymbol{R}^{-1})\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y}],
\end{aligned}$$

$$\begin{aligned}
\psi_{\lambda\lambda\rho,i_1,i_2,j} &= \frac{\partial^2 \psi_{\lambda_{i_1}}(\boldsymbol{\gamma}_0)}{\partial\lambda_{i_2}\partial\rho_j} \\
&= d_{i_1}^{-1}[(\delta_{i_1 i_2} - 1)(\boldsymbol{y}'\boldsymbol{W}'_{i_1}\boldsymbol{R}'\boldsymbol{H}_j\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{W}'_{i_1}\boldsymbol{M}'_j\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y} \\
&\quad - \boldsymbol{y}'\boldsymbol{W}'_{i_1}\boldsymbol{R}'\boldsymbol{H}\boldsymbol{M}_j\boldsymbol{W}_{i_2}\boldsymbol{y}) - 2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{M}'_j\boldsymbol{H}\boldsymbol{D}_{i_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y} \\
&\quad + 2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}_j\boldsymbol{D}_{i_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y} + 2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{D}_{i_1,j}\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y} \\
&\quad + 2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{D}_{i_1}\boldsymbol{H}_j\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{D}_{i_1}\boldsymbol{H}\boldsymbol{M}_j\boldsymbol{W}_{i_2}\boldsymbol{y} \\
&\quad + 2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\mathrm{Dg}(\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_{i_1,i_2}\boldsymbol{R}^{-1})\boldsymbol{H}\boldsymbol{M}_j\boldsymbol{S}\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}_j\mathrm{Dg}(\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_{i_1,i_2}\boldsymbol{R}^{-1})\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}
\end{aligned}$$

$$
\begin{aligned}
&- \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\mathrm{Dg}(\boldsymbol{H}_j\boldsymbol{R}\boldsymbol{G}_{i_1,i_2}\boldsymbol{R}^{-1} - \boldsymbol{H}\boldsymbol{M}_j\boldsymbol{G}_{i_1,i_2}\boldsymbol{R}^{-1} + \boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_{i_1,i_2}\boldsymbol{R}^{-1}\boldsymbol{M}_j)\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}]\\
&- d_{i_1}^{-2}d_{i_1,j}[(\delta_{i_1i_2}-1)\boldsymbol{y}'\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y} + 2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\boldsymbol{D}_{i_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_{i_2}\boldsymbol{y}\\
&- \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}\mathrm{Dg}(\boldsymbol{H}\boldsymbol{R}\boldsymbol{G}_{i_1,i_2}\boldsymbol{R}^{-1})\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}],\\
\psi_{\rho,j_1,j_2,j_3} &= \frac{\partial^2\psi_{\rho_{j_1}}(\boldsymbol{\gamma}_0)}{\partial\rho_{j_2}\partial\rho_{j_3}}\\
&= e_{j_1}^{-1}\{\boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{R}'\boldsymbol{H}_{j_2,j_3} - \boldsymbol{M}_{j_3}'\boldsymbol{H}_{j_2} - \boldsymbol{M}_{j_2}'\boldsymbol{H}_{j_3})\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}\\
&+ \boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{R}'\boldsymbol{H}_{j_2} - \boldsymbol{M}_{j_2}'\boldsymbol{H})[\boldsymbol{F}_{j_3}'\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}' + (\delta_{j_1j_3}-1)\boldsymbol{R}^{-1\prime}\boldsymbol{M}_{j_3}']\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}\\
&+ \boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{R}'\boldsymbol{H}_{j_2} - \boldsymbol{M}_{j_2}'\boldsymbol{H})\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1,j_3}\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}\\
&+ \boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{R}'\boldsymbol{H}_{j_2} - \boldsymbol{M}_{j_2}'\boldsymbol{H})\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}(\boldsymbol{H}_{j_3}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{M}_{j_3})\boldsymbol{S}\boldsymbol{y}\\
&+ \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[\boldsymbol{F}_{j_2,j_3}'\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} + \boldsymbol{F}_{j_2}'\boldsymbol{F}_{j_3}'\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}\\
&+ (\delta_{j_1j_3}-1)\boldsymbol{F}_{j_2}'\boldsymbol{R}^{-1\prime}\boldsymbol{M}_{j_3}'\boldsymbol{F}_{j_1} + \boldsymbol{F}_{j_2}'\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1,j_3}\\
&+ (\delta_{j_1j_2}-1)\boldsymbol{M}_{j_3}'\boldsymbol{R}^{-1\prime}\boldsymbol{M}_{j_2}'\boldsymbol{F}_{j_1} + (\delta_{j_1j_2}-1)\boldsymbol{R}^{-1\prime}\boldsymbol{M}_{j_2}'\boldsymbol{F}_{j_1,j_3}\\
&+ \boldsymbol{M}_{j_3}'\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1,j_2} + (\delta_{j_1j_3}-1)\boldsymbol{R}^{-1\prime}\boldsymbol{M}_{j_3}'\boldsymbol{F}_{j_1,j_2} + \boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1,j_2,j_3}]\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}\\
&+ \boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{R}'\boldsymbol{H}_{j_3} - \boldsymbol{M}_{j_3}'\boldsymbol{H})[\boldsymbol{F}_{j_2}'\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} + (\delta_{j_1j_2}-1)\boldsymbol{R}^{-1\prime}\boldsymbol{M}_{j_2}'\boldsymbol{F}_{j_1}\\
&+ \boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1,j_2}]\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[\boldsymbol{F}_{j_2}'\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} + (\delta_{j_1j_2}-1)\boldsymbol{R}^{-1\prime}\boldsymbol{M}_{j_2}'\boldsymbol{F}_{j_1}\\
&+ \boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1,j_2}](\boldsymbol{H}_{j_3}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{M}_{j_3})\boldsymbol{S}\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[(\boldsymbol{M}_{j_3}'\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}\\
&+ (\delta_{j_1j_3}-1)\boldsymbol{R}^{-1\prime}\boldsymbol{M}_{j_3}'\boldsymbol{F}_{j_1} + \boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1,j_3})(\boldsymbol{H}_{j_2}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{M}_{j_2})]\boldsymbol{S}\boldsymbol{y}\\
&+ \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}(\boldsymbol{H}_{j_2,j_3}\boldsymbol{R} - \boldsymbol{H}_{j_2}\boldsymbol{M}_{j_3} - \boldsymbol{H}_{j_3}\boldsymbol{M}_{j_2}) + 2\mathrm{Dg}(\boldsymbol{F}_{j_1,j_3})\boldsymbol{H}\boldsymbol{M}_{j_2}\\
&+ 2\boldsymbol{K}_{j_1}\boldsymbol{H}_{j_3}\boldsymbol{M}_{j_2}]\boldsymbol{S}\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{R}'\boldsymbol{H}_{j_3} - \boldsymbol{M}_{j_3}'\boldsymbol{H})[\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}(\boldsymbol{H}_{j_2}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{M}_{j_2})\\
&+ 2\boldsymbol{K}_{j_1}\boldsymbol{H}\boldsymbol{M}_{j_2}]\boldsymbol{S}\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[2\mathrm{Dg}(\boldsymbol{F}_{j_1,j_3})\boldsymbol{H}_{j_2} + 2\boldsymbol{K}_{j_1}\boldsymbol{H}_{j_2,j_3} + \mathrm{Dg}(\boldsymbol{F}_{j_1,j_2,j_3})\boldsymbol{H}\\
&+ \mathrm{Dg}(\boldsymbol{F}_{j_1,j_2})\boldsymbol{H}_{j_3}]\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{R}'\boldsymbol{H}_{j_3} - \boldsymbol{M}_{j_3}'\boldsymbol{H})[2\mathrm{Dg}(\boldsymbol{F}_{j_1,j_3})\boldsymbol{H}_{j_2} + 2\boldsymbol{K}_{j_1}\boldsymbol{H}_{j_2,j_3}\\
&+ \mathrm{Dg}(\boldsymbol{F}_{j_1,j_2,j_3})\boldsymbol{H} + \mathrm{Dg}(\boldsymbol{F}_{j_1,j_2})\boldsymbol{H}_{j_3}]\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} + \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[2\boldsymbol{K}_{j_1}\boldsymbol{H}_{j_2}\\
&+ \mathrm{Dg}(\boldsymbol{F}_{j_1,j_2})\boldsymbol{H}]\boldsymbol{M}_{j_3}\boldsymbol{S}\boldsymbol{y}\} - e_{j_1}^{-2}e_{j_1,j_3}\{\boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{R}'\boldsymbol{H}_{j_2} - \boldsymbol{M}_{j_2}'\boldsymbol{H})\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}\\
&+ \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[\boldsymbol{F}_{j_2}'\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} + (\delta_{j_1j_2}-1)\boldsymbol{R}^{-1\prime}\boldsymbol{M}_{j_2}'\boldsymbol{F}_{j_1} + \boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1,j_2}]\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}\\
&+ \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}(\boldsymbol{H}_{j_2}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{M}_{j_2}) + 2\boldsymbol{K}_{j_1}\boldsymbol{H}\boldsymbol{M}_{j_2}]\boldsymbol{S}\boldsymbol{y}\\
&- \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[2\boldsymbol{K}_{j_1}\boldsymbol{H}_{j_2} + \mathrm{Dg}(\boldsymbol{F}_{j_1,j_2})\boldsymbol{H}]\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}\}\\
&+ 2e_{j_1}^{-3}e_{j_1,j_2}e_{j_1,j_3}\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}(\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} - \boldsymbol{K}_{j_1})\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}\\
&- e_{j_1}^{-2}e_{j_1,j_2,j_3}\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}(\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} - \boldsymbol{K}_{j_1})\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}\\
&- e_{j_1}^{-2}e_{j_1,j_2}\boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{R}'\boldsymbol{H}_{j_2} - \boldsymbol{M}_{j_2}'\boldsymbol{H})(\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} - \boldsymbol{K}_{j_1})\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}\\
&- e_{j_1}^{-2}e_{j_1,j_2}\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[\boldsymbol{M}_{j_3}'\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} + (\delta_{j_1j_3}-1)\boldsymbol{R}^{-1\prime}\boldsymbol{M}_{j_3}'\boldsymbol{F}_{j_1} + \boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1,j_3}\\
&- \mathrm{Dg}(\boldsymbol{F}_{j_1,j_3})]\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y} - e_{j_1}^{-2}e_{j_1,j_2}\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}(\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} - \boldsymbol{K}_{j_1})(\boldsymbol{H}_{j_2}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{M}_{j_2})\boldsymbol{S}\boldsymbol{y},\\
\psi_{\rho\rho\lambda,j_1,j_2,i} &= \frac{\partial^2\psi_{\rho_{j_1}}(\boldsymbol{\gamma}_0)}{\partial\rho_{j_2}\partial\lambda_i}\\
&= e_{j_1}^{-1}\{-\boldsymbol{y}'\boldsymbol{W}_i'(\boldsymbol{R}'\boldsymbol{H}_{j_2} - \boldsymbol{M}_{j_2}'\boldsymbol{H})\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}\\
&- \boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{R}'\boldsymbol{H}_{j_2} - \boldsymbol{M}_{j_2}'\boldsymbol{H})\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}\\
&- \boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}[\boldsymbol{F}_{j_2}'\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} + (\delta_{j_1j_2}-1)\boldsymbol{R}^{-1\prime}\boldsymbol{M}_{j_2}'\boldsymbol{F}_{j_1} + \boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1,j_2}]\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}\\
&- \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[\boldsymbol{F}_{j_2}'\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} + (\delta_{j_1j_2}-1)\boldsymbol{R}^{-1\prime}\boldsymbol{M}_{j_2}'\boldsymbol{F}_{j_1} + \boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1,j_2}]\boldsymbol{H}\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}\\
&- \boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}[\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}(\boldsymbol{H}_{j_2}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{M}_{j_2}) + 2\boldsymbol{K}_{j_1}\boldsymbol{H}\boldsymbol{M}_{j_2}]\boldsymbol{S}\boldsymbol{y}\\
&- \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}(\boldsymbol{H}_{j_2}\boldsymbol{R} - \boldsymbol{H}\boldsymbol{M}_{j_2}) + 2\boldsymbol{K}_{j_1}\boldsymbol{H}\boldsymbol{M}_{j_2}]\boldsymbol{W}_i\boldsymbol{y}\\
&+ \boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}[2\boldsymbol{K}_{j_1}\boldsymbol{H}_{j_2} + \mathrm{Dg}(\boldsymbol{F}_{j_1,j_2})\boldsymbol{H}]\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}\\
&+ \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[2\boldsymbol{K}_{j_1}\boldsymbol{H}_{j_2} + \mathrm{Dg}(\boldsymbol{F}_{j_1,j_2})\boldsymbol{H}]\boldsymbol{R}\boldsymbol{W}_i\boldsymbol{y}\}\\
&- e_{j_1}^{-2}e_{j_1,(i)}\{\boldsymbol{y}'\boldsymbol{S}'(\boldsymbol{R}'\boldsymbol{H}_{j_2} - \boldsymbol{M}_{j_2}'\boldsymbol{H})\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}\boldsymbol{H}\boldsymbol{R}\boldsymbol{S}\boldsymbol{y}
\end{aligned}
$$

$$
\begin{aligned}
&+ \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[\boldsymbol{F}_{j_2}'\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} + (\delta_{j_1j_2}-1)\boldsymbol{R}^{-1\prime}\boldsymbol{M}_{j_2}'\boldsymbol{F}_{j_1} + \boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1,j_2}]\boldsymbol{HRSy}\\
&+ \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1}(\boldsymbol{H}_{j_2}\boldsymbol{R} - \boldsymbol{HM}_{j_2}) + 2\boldsymbol{K}_{j_1}\boldsymbol{HM}_{j_2}]\boldsymbol{Sy}\\
&- \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}[2\boldsymbol{K}_{j_1}\boldsymbol{H}_{j_2} + \mathrm{Dg}(\boldsymbol{F}_{j_1,j_2})\boldsymbol{H}]\boldsymbol{RSy}\}\\
&+ 2e_{j_1}^{-3}e_{j_1,(i)}e_{j_1,j_2}\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}(\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} - \boldsymbol{K}_{j_1})\boldsymbol{HRSy}\\
&- e_{j_1}^{-2}e_{j_1,j_2,(i)}\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}(\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} - \boldsymbol{K}_{j_1})\boldsymbol{HRSy}\\
&+ e_{j_1}^{-2}e_{j_1,j_2}\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}(\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} - \boldsymbol{K}_{j_1})\boldsymbol{HRSy}\\
&+ e_{j_1}^{-2}e_{j_1,j_2}\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}(\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j_1)}'\boldsymbol{F}_{j_1} - \boldsymbol{K}_{j_1})\boldsymbol{HRW}_i\boldsymbol{y},
\end{aligned}
$$

$$
\begin{aligned}
\psi_{\lambda\rho\rho,i,j_1,j_2} &= \frac{\partial^2\psi_{\lambda_i}(\boldsymbol{\gamma}_0)}{\partial\rho_{j_1}\partial\rho_{j_2}}\\
&= d_i^{-1}\{\boldsymbol{y}'\boldsymbol{W}_i'[\boldsymbol{R}'\boldsymbol{H}_{j_1,j_2}\boldsymbol{R} - (\boldsymbol{M}_{j_2}'\boldsymbol{H}_{j_1}\boldsymbol{R})^\star + (\boldsymbol{M}_{j_1}'\boldsymbol{HM}_{j_2} - \boldsymbol{M}_{j_1}'\boldsymbol{H}_{j_2}\boldsymbol{R})^\star]\boldsymbol{S}_{(-i)}\boldsymbol{y}\\
&- \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{M}_{j_2}'\boldsymbol{H}(2\boldsymbol{D}_i\boldsymbol{HM}_{j_1} - 2\boldsymbol{D}_i\boldsymbol{H}_{j_1}\boldsymbol{R} - \boldsymbol{D}_{i,j_1}\boldsymbol{HR})\boldsymbol{Sy}\\
&+ \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}_{j_2}(2\boldsymbol{D}_i\boldsymbol{HM}_{j_1} - 2\boldsymbol{D}_i\boldsymbol{H}_{j_1}\boldsymbol{R} - \boldsymbol{D}_{i,j_1}\boldsymbol{HR})\boldsymbol{Sy}\\
&+ \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}(2\boldsymbol{D}_{i,j_2}\boldsymbol{HM}_{j_1} + 2\boldsymbol{D}_i\boldsymbol{H}_{j_2}\boldsymbol{M}_{j_1} - 2\boldsymbol{D}_{i,j_2}\boldsymbol{H}_{j_1}\boldsymbol{R} - 2\boldsymbol{D}_i\boldsymbol{H}_{j_1,j_2}\boldsymbol{R} + 2\boldsymbol{D}_i\boldsymbol{H}_{j_1}\boldsymbol{M}_{j_2}\\
&- \boldsymbol{D}_{i,j_1,j_2}\boldsymbol{HR} - \boldsymbol{D}_{i,j_1}\boldsymbol{H}_{j_2}\boldsymbol{R} + \boldsymbol{D}_{i,j_1}\boldsymbol{HM}_{j_2})\boldsymbol{Sy}\}\\
&- d_i^{-2}d_{i,j_2}[\boldsymbol{y}'\boldsymbol{W}_i'(\boldsymbol{R}'\boldsymbol{H}_{j_1}\boldsymbol{R} - \boldsymbol{M}_{j_1}'\boldsymbol{HR} - \boldsymbol{R}'\boldsymbol{HM}_{j_1})\boldsymbol{S}_{(-i)}\boldsymbol{y}\\
&+ \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}(2\boldsymbol{D}_i\boldsymbol{HM}_{j_1} - 2\boldsymbol{D}_i\boldsymbol{H}_{j_1}\boldsymbol{R} - \boldsymbol{D}_{i,j_1}\boldsymbol{HR})\boldsymbol{Sy}]\\
&+ 2d_i^{-3}d_{i,j_1}d_{i,j_2}(\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{HRS}_{(-i)}\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{HD}_i\boldsymbol{HRSy})\\
&- d_i^{-2}d_{i,j_1}d_{i,j_2}(\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{HRS}_{(-i)}\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{HD}_i\boldsymbol{HRSy})\\
&- d_i^{-2}d_{i,j_1}[\boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{H}_{j_2}\boldsymbol{RS}_{(-i)}\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{M}_{j_2}'\boldsymbol{HRS}_{(-i)}\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{W}_i'\boldsymbol{R}'\boldsymbol{HM}_{j_2}\boldsymbol{S}_{(-i)}\boldsymbol{y}\\
&+ \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{M}_{j_2}'\boldsymbol{HD}_i\boldsymbol{HRSy} - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}_{j_2}\boldsymbol{D}_i\boldsymbol{HRSy} - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{HD}_{i,j_2}\boldsymbol{HRSy}\\
&- \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{HD}_i\boldsymbol{H}_{j_2}\boldsymbol{RSy} + \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{HD}_i\boldsymbol{HM}_{j_2}\boldsymbol{Sy}],
\end{aligned}
$$

$$
\begin{aligned}
\psi_{\rho\lambda\lambda,j,i_1,i_2} &= \frac{\partial^2\psi_{\rho_j}(\boldsymbol{\gamma}_0)}{\partial\lambda_{i_1}\partial\lambda_{i_2}}\\
&= e_j^{-1}(\boldsymbol{y}'\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{HR}^{-1\prime}\boldsymbol{R}_{(-j)}'\boldsymbol{F}_j\boldsymbol{HRW}_{i_2}\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{W}_{i_2}'\boldsymbol{R}'\boldsymbol{HK}_j\boldsymbol{HRW}_{i_1}\boldsymbol{y}\\
&+ \boldsymbol{y}'\boldsymbol{W}_{i_2}'\boldsymbol{R}'\boldsymbol{HR}^{-1\prime}\boldsymbol{R}_{(-j)}'\boldsymbol{F}_j\boldsymbol{HRW}_{i_1}\boldsymbol{y}) - e_j^{-2}e_{j,(i_2)}(2\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{HK}_j\boldsymbol{HRW}_{i_1}\boldsymbol{y}\\
&- \boldsymbol{y}'\boldsymbol{W}_{i_1}'\boldsymbol{R}'\boldsymbol{HR}^{-1\prime}\boldsymbol{R}_{(-j)}'\boldsymbol{F}_j\boldsymbol{HRSy} - \boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{HR}^{-1\prime}\boldsymbol{R}_{(-j)}'\boldsymbol{F}_j\boldsymbol{HRW}_{i_1}\boldsymbol{y})\\
&- 2e_j^{-2}\boldsymbol{y}'\boldsymbol{W}_{i_2}'\boldsymbol{R}'\boldsymbol{HF}_j'\boldsymbol{F}_j\boldsymbol{HRW}_{i_1}\boldsymbol{y}\cdot\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}(\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j)}'\boldsymbol{F}_j - \boldsymbol{K}_j)\boldsymbol{HRSy}\\
&+ 2e_j^{-2}e_{j,(i_1)}\boldsymbol{y}'\boldsymbol{W}_{i_2}'\boldsymbol{R}'\boldsymbol{H}(\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j)}'\boldsymbol{F}_j - \boldsymbol{K}_j)\boldsymbol{HRSy}\\
&+ 2e_j^{-2}e_{j,(i_1)}\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}(\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j)}'\boldsymbol{F}_j - \boldsymbol{K}_j)\boldsymbol{HRW}_{i_2}\boldsymbol{y}\\
&+ 4e_j^{-3}e_{j,(i_2)}e_{j,(i_1)}\boldsymbol{y}'\boldsymbol{S}'\boldsymbol{R}'\boldsymbol{H}(\boldsymbol{R}^{-1\prime}\boldsymbol{R}_{(-j)}'\boldsymbol{F}_j - \boldsymbol{K}_j)\boldsymbol{HRSy},
\end{aligned}
$$

$$
\psi_{\rho\lambda\rho,j_1,i,j_2} = \frac{\partial^2\psi_{\rho_{j_1}}(\boldsymbol{\gamma}_0)}{\partial\lambda_i\partial\rho_{j_2}} = \psi_{\rho\rho\lambda,j_1,j_2,i},
$$

$$
\psi_{\lambda\rho\lambda,i_1,j,i_2} = \frac{\partial^2\psi_{\lambda_{i_1}}(\boldsymbol{\gamma}_0)}{\partial\rho_j\partial\lambda_{i_2}} = \psi_{\lambda\lambda\rho,i_1,i_2,j}.
$$

Additional Simulation Results

Unless otherwise stated, all aspects of simulation design are the same as those in Table 1 in the main text. In Table S.1, $\boldsymbol{\beta}_0 = (2, 1.5, -1)'$. In Table S.2, $k_x = 11$, all the non-constant covariates are simulated as independent uniform random variables on the interval $[0, 1]$, and $\boldsymbol{\beta}_0 = (0.2, 0.1, -0.3, 0.1, -0.3, \cdots)'$. In Table S.3, $\boldsymbol{W}_2$ in the outcome equation is first constructed as a random symmetric matrix of zeros and ones with the number of ones restricted to be 20% of the total entries and then row-normalized.

Table S.1: Finite-Sample Performances of GMM, GS2SLS, and II in Estimating SARAR(2,2) with Strong Covariates

			GMM			GS2SLS			II		
n	DGP	$\boldsymbol{\theta}_0$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$
50	1	$\lambda_{01}=0.15$	−0.003	0.284	23.6%	0.757	1.153	14.9%	−0.027	0.193	11.0%
		$\lambda_{02}=0.1$	−0.010	0.179	17.0%	0.094	0.576	3.2%	−0.004	0.117	5.4%
		$\rho_{01}=0.5$	0.037	0.153	17.9%	−0.068	0.161	15.6%	0.022	0.123	12.3%
		$\rho_{02}=0.4$	−0.076	0.193	21.0%	−0.080	0.180	17.3%	−0.049	0.136	13.8%
	2	$\lambda_{01}=0.5$	−0.114	0.332	16.1%	0.177	0.463	6.6%	−0.102	0.260	7.6%
		$\lambda_{02}=0.4$	−0.094	0.287	17.7%	0.115	0.556	6.9%	−0.030	0.194	6.9%
		$\rho_{01}=0.15$	−0.016	0.198	17.1%	−0.031	0.158	9.9%	−0.015	0.154	9.0%
		$\rho_{02}=0.1$	−0.065	0.264	17.9%	−0.030	0.156	13.8%	−0.039	0.155	11.6%
	3	$\lambda_{01}=0.15$	−0.049	0.462	18.9%	0.603	1.012	10.8%	−0.053	0.379	11.3%
		$\lambda_{02}=0.1$	−0.064	0.355	18.7%	−0.127	0.779	4.4%	−0.016	0.236	8.3%
		$\rho_{01}=0.05$	−0.020	0.206	18.2%	−0.017	0.165	11.0%	−0.021	0.164	9.9%
		$\rho_{02}=0.02$	−0.064	0.271	21.1%	−0.034	0.160	16.6%	−0.048	0.162	13.8%
	4	$\lambda_{01}=0.5$	−0.106	0.337	19.0%	0.283	0.536	9.0%	−0.076	0.232	8.5%
		$\lambda_{02}=0.4$	−0.076	0.284	18.1%	−0.005	0.536	6.6%	−0.023	0.178	5.5%
		$\rho_{01}=0.3$	−0.007	0.187	16.0%	−0.045	0.147	11.6%	−0.014	0.138	9.8%
		$\rho_{02}=0.2$	−0.092	0.273	18.8%	−0.054	0.161	17.0%	−0.040	0.152	10.9%
100	1	$\lambda_{01}=0.15$	0.016	0.195	13.3%	0.857	1.108	14.7%	−0.001	0.133	10.0%
		$\lambda_{02}=0.1$	−0.007	0.147	7.7%	0.059	0.932	1.3%	0.000	0.103	5.8%
		$\rho_{01}=0.5$	0.018	0.101	11.7%	−0.076	0.139	23.5%	0.005	0.079	9.3%
		$\rho_{02}=0.4$	−0.040	0.118	11.7%	−0.052	0.129	14.3%	−0.017	0.085	8.2%
	2	$\lambda_{01}=0.5$	−0.052	0.241	10.0%	0.140	0.350	7.6%	−0.059	0.173	5.5%
		$\lambda_{02}=0.4$	−0.103	0.299	10.7%	0.194	0.938	5.7%	−0.023	0.196	4.6%
		$\rho_{01}=0.15$	−0.008	0.134	10.5%	−0.021	0.112	7.6%	−0.001	0.101	8.2%
		$\rho_{02}=0.1$	−0.043	0.190	11.7%	−0.021	0.115	12.1%	−0.014	0.109	10.2%
	3	$\lambda_{01}=0.15$	−0.049	0.369	10.0%	0.485	0.734	9.6%	−0.048	0.270	7.6%
		$\lambda_{02}=0.1$	−0.042	0.354	13.5%	−0.092	1.032	2.5%	0.005	0.229	7.8%
		$\rho_{01}=0.05$	−0.013	0.143	11.0%	−0.020	0.114	7.2%	−0.012	0.107	7.8%
		$\rho_{02}=0.02$	−0.043	0.192	12.3%	−0.013	0.118	11.0%	−0.015	0.103	9.9%
	4	$\lambda_{01}=0.5$	−0.065	0.235	10.0%	0.189	0.379	7.7%	−0.063	0.172	6.5%
		$\lambda_{02}=0.4$	−0.091	0.282	9.6%	0.272	0.966	5.6%	−0.020	0.175	3.5%
		$\rho_{01}=0.3$	0.007	0.118	8.9%	−0.035	0.102	7.4%	0.000	0.091	8.1%
		$\rho_{02}=0.2$	−0.045	0.174	12.2%	−0.038	0.115	11.3%	−0.015	0.098	8.8%
200	1	$\lambda_{01}=0.15$	0.008	0.129	7.4%	0.913	1.127	16.3%	−0.005	0.090	6.2%
		$\lambda_{02}=0.1$	−0.007	0.142	6.8%	0.150	1.416	3.5%	−0.003	0.099	4.9%
		$\rho_{01}=0.5$	0.011	0.068	8.8%	−0.069	0.118	26.1%	0.005	0.055	7.2%
		$\rho_{02}=0.4$	−0.020	0.077	8.9%	−0.052	0.108	17.0%	−0.011	0.059	6.6%
	2	$\lambda_{01}=0.5$	−0.031	0.168	6.3%	0.092	0.241	6.0%	−0.052	0.141	5.6%
		$\lambda_{02}=0.4$	−0.124	0.278	6.6%	0.141	1.400	3.9%	−0.023	0.197	3.6%
		$\rho_{01}=0.15$	−0.005	0.088	9.1%	−0.017	0.076	7.6%	−0.003	0.073	8.3%
		$\rho_{02}=0.1$	−0.017	0.116	8.5%	−0.018	0.081	11.7%	−0.009	0.072	7.6%
	3	$\lambda_{01}=0.15$	−0.015	0.284	8.8%	0.391	0.659	9.3%	−0.021	0.206	8.0%
		$\lambda_{02}=0.1$	−0.065	0.333	8.0%	−0.115	1.453	3.8%	−0.018	0.230	5.6%
		$\rho_{01}=0.05$	−0.006	0.090	8.4%	−0.011	0.078	7.3%	−0.004	0.072	6.9%
		$\rho_{02}=0.02$	−0.021	0.112	9.5%	−0.007	0.081	9.8%	−0.007	0.073	7.6%
	4	$\lambda_{01}=0.5$	−0.031	0.167	6.9%	0.132	0.294	6.0%	−0.045	0.136	6.3%
		$\lambda_{02}=0.4$	−0.099	0.254	6.2%	0.189	1.458	4.6%	−0.025	0.181	2.8%
		$\rho_{01}=0.3$	0.001	0.075	8.3%	−0.030	0.074	10.3%	−0.001	0.064	7.3%
		$\rho_{02}=0.2$	−0.031	0.108	7.3%	−0.025	0.079	9.8%	−0.011	0.070	8.5%

Note: All results are based on 1000 simulations and $P(5\%)$ denotes the empirical size of the two-sided 5% t-test of the specific parameter equal to its true value.

Table S.2: Finite-Sample Performances of GMM, GS2SLS, and II in Estimating SARAR(2,2) with $k_x = 11$ Covariates

			GMM			GS2SLS			II		
n	DGP	$\boldsymbol{\theta}_0$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$
50	1	$\lambda_{01} = 0.15$	−0.002	0.339	39.4%	−0.086	0.676	15.9%	−0.033	0.235	15.9%
		$\lambda_{02} = 0.1$	−0.014	0.213	29.4%	−0.151	0.450	21.9%	−0.007	0.141	10.5%
		$\rho_{01} = 0.5$	0.112	0.227	42.3%	−0.150	0.197	34.9%	0.083	0.201	20.6%
		$\rho_{02} = 0.4$	−0.133	0.249	40.8%	−0.067	0.151	17.2%	−0.090	0.209	20.4%
	2	$\lambda_{01} = 0.5$	−0.133	0.440	28.5%	0.076	0.559	17.3%	−0.114	0.326	9.2%
		$\lambda_{02} = 0.4$	−0.113	0.366	28.9%	0.142	0.419	22.3%	−0.046	0.232	10.1%
		$\rho_{01} = 0.15$	0.042	0.310	40.0%	−0.042	0.161	12.5%	0.024	0.265	13.0%
		$\rho_{02} = 0.1$	−0.115	0.337	35.0%	−0.045	0.159	16.2%	−0.070	0.251	18.3%
	3	$\lambda_{01} = 0.15$	−0.010	0.536	32.6%	−0.170	0.867	23.5%	−0.078	0.447	15.8%
		$\lambda_{02} = 0.1$	−0.062	0.410	30.1%	−0.058	0.460	20.4%	−0.023	0.269	11.4%
		$\rho_{01} = 0.05$	−0.022	0.342	44.3%	−0.045	0.155	12.6%	−0.044	0.278	11.6%
		$\rho_{02} = 0.02$	−0.102	0.334	38.5%	−0.037	0.155	16.9%	−0.080	0.243	18.5%
	4	$\lambda_{01} = 0.5$	−0.107	0.413	29.3%	0.006	0.595	16.8%	−0.096	0.308	10.8%
		$\lambda_{02} = 0.4$	−0.079	0.332	25.6%	0.187	0.428	25.5%	−0.026	0.215	9.1%
		$\rho_{01} = 0.3$	0.062	0.272	37.7%	−0.066	0.151	11.4%	0.046	0.221	14.1%
		$\rho_{02} = 0.2$	−0.122	0.336	35.5%	−0.058	0.161	15.8%	−0.047	0.229	16.2%
100	1	$\lambda_{01} = 0.15$	0.020	0.211	19.0%	0.527	0.656	31.1%	0.002	0.140	10.8%
		$\lambda_{02} = 0.1$	−0.008	0.160	13.2%	0.240	0.476	16.1%	0.003	0.112	7.6%
		$\rho_{01} = 0.5$	0.035	0.119	19.5%	−0.074	0.116	21.1%	0.021	0.099	10.1%
		$\rho_{02} = 0.4$	−0.048	0.134	20.5%	−0.066	0.118	16.7%	−0.026	0.104	9.9%
	2	$\lambda_{01} = 0.5$	−0.052	0.321	13.4%	0.596	0.665	42.2%	−0.060	0.215	7.4%
		$\lambda_{02} = 0.4$	−0.106	0.320	13.1%	0.183	0.476	14.5%	−0.033	0.201	5.0%
		$\rho_{01} = 0.15$	0.000	0.165	17.3%	−0.029	0.110	9.5%	0.014	0.123	8.9%
		$\rho_{02} = 0.1$	−0.059	0.219	18.1%	−0.026	0.112	11.2%	−0.011	0.134	12.2%
	3	$\lambda_{01} = 0.15$	−0.029	0.423	14.9%	0.644	0.809	33.1%	−0.055	0.301	8.8%
		$\lambda_{02} = 0.1$	−0.066	0.366	15.1%	0.033	0.475	11.2%	0.004	0.242	8.2%
		$\rho_{01} = 0.05$	0.002	0.170	19.8%	−0.015	0.109	7.5%	−0.002	0.130	8.9%
		$\rho_{02} = 0.02$	−0.056	0.210	20.4%	−0.008	0.110	11.9%	−0.013	0.127	10.4%
	4	$\lambda_{01} = 0.5$	−0.056	0.313	15.1%	0.550	0.629	38.3%	−0.060	0.193	6.9%
		$\lambda_{02} = 0.4$	−0.092	0.300	11.9%	0.254	0.494	16.9%	−0.020	0.189	6.2%
		$\rho_{01} = 0.3$	0.027	0.143	17.8%	−0.049	0.106	9.1%	0.025	0.116	9.9%
		$\rho_{02} = 0.2$	−0.063	0.195	15.2%	−0.046	0.118	13.8%	−0.017	0.121	11.3%
200	1	$\lambda_{01} = 0.15$	0.008	0.131	9.0%	1.047	1.100	70.1%	−0.005	0.095	7.6%
		$\lambda_{02} = 0.1$	−0.003	0.145	8.0%	−0.025	0.597	6.3%	−0.003	0.102	5.2%
		$\rho_{01} = 0.5$	0.019	0.071	11.8%	−0.060	0.093	22.6%	0.013	0.061	8.1%
		$\rho_{02} = 0.4$	−0.027	0.079	12.4%	−0.055	0.098	17.4%	−0.015	0.065	7.9%
	2	$\lambda_{01} = 0.5$	−0.034	0.266	7.7%	0.870	0.910	70.0%	−0.050	0.173	6.3%
		$\lambda_{02} = 0.4$	−0.121	0.304	8.1%	0.198	0.637	8.4%	−0.033	0.198	4.3%
		$\rho_{01} = 0.15$	−0.001	0.096	12.2%	−0.024	0.080	8.1%	0.003	0.081	9.5%
		$\rho_{02} = 0.1$	−0.030	0.128	10.7%	−0.019	0.079	9.4%	−0.010	0.081	8.5%
	3	$\lambda_{01} = 0.15$	0.002	0.318	10.9%	1.049	1.108	61.8%	−0.020	0.220	8.9%
		$\lambda_{02} = 0.1$	−0.056	0.333	7.8%	−0.025	0.633	7.3%	−0.020	0.240	6.7%
		$\rho_{01} = 0.05$	−0.002	0.111	12.6%	−0.015	0.081	5.8%	−0.003	0.079	7.0%
		$\rho_{02} = 0.02$	−0.022	0.123	11.4%	−0.007	0.079	10.5%	−0.008	0.081	7.5%
	4	$\lambda_{01} = 0.5$	−0.025	0.242	9.0%	0.880	0.916	74.4%	−0.042	0.156	6.4%
		$\lambda_{02} = 0.4$	−0.100	0.274	7.7%	0.209	0.634	10.1%	−0.028	0.183	3.4%
		$\rho_{01} = 0.3$	0.008	0.082	10.9%	−0.041	0.085	11.5%	0.009	0.073	9.2%
		$\rho_{02} = 0.2$	−0.035	0.114	10.4%	−0.026	0.082	9.9%	−0.011	0.079	9.5%

Note: All results are based on 1000 simulations and $P(5\%)$ denotes the empirical size of the two-sided 5% t-test of the specific parameter equal to its true value.

Table S.3: Finite-Sample Performances of GMM, GS2SLS, and II in Estimating SARAR(2,2) with Dense $\boldsymbol{W}_2$

			GMM			GS2SLS			II		
n	DGP	$\boldsymbol{\theta}_0$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$	Bias	RMSE	$P(5\%)$
50	1	$\lambda_{01}=0.15$	0.011	0.291	26.2%	0.954	1.328	13.3%	−0.027	0.198	11.2%
		$\lambda_{02}=0.1$	−0.030	0.253	15.4%	−0.038	1.013	3.0%	−0.011	0.176	5.3%
		$\rho_{01}=0.5$	0.034	0.153	17.6%	−0.085	0.168	17.2%	0.021	0.123	12.2%
		$\rho_{02}=0.4$	−0.071	0.190	17.5%	−0.093	0.182	17.8%	−0.048	0.136	13.1%
	2	$\lambda_{01}=0.5$	−0.093	0.407	20.3%	0.749	0.980	13.7%	−0.114	0.297	8.1%
		$\lambda_{02}=0.4$	−0.194	0.437	18.9%	0.060	1.063	5.5%	−0.065	0.302	4.9%
		$\rho_{01}=0.15$	−0.029	0.205	16.1%	−0.029	0.163	9.8%	−0.024	0.157	10.7%
		$\rho_{02}=0.1$	−0.082	0.293	19.5%	−0.038	0.170	14.6%	−0.043	0.158	12.6%
	3	$\lambda_{01}=0.15$	−0.013	0.468	18.1%	1.001	1.273	14.9%	−0.059	0.385	10.4%
		$\lambda_{02}=0.1$	−0.130	0.445	17.3%	−0.028	1.069	5.3%	−0.031	0.343	8.3%
		$\rho_{01}=0.05$	−0.021	0.210	16.7%	−0.013	0.168	9.5%	−0.023	0.164	9.8%
		$\rho_{02}=0.02$	−0.072	0.261	21.2%	−0.044	0.169	13.8%	−0.048	0.160	14.0%
	4	$\lambda_{01}=0.5$	−0.069	0.379	23.9%	0.614	0.889	13.1%	−0.096	0.287	11.2%
		$\lambda_{02}=0.4$	−0.213	0.438	20.7%	−0.101	1.092	5.0%	−0.068	0.285	5.1%
		$\rho_{01}=0.3$	−0.006	0.186	16.0%	−0.053	0.164	11.1%	−0.019	0.145	10.5%
		$\rho_{02}=0.2$	−0.091	0.287	19.9%	−0.064	0.179	16.7%	−0.043	0.156	12.9%
100	1	$\lambda_{01}=0.15$	0.016	0.197	15.2%	1.006	1.183	18.9%	−0.001	0.133	10.7%
		$\lambda_{02}=0.1$	−0.014	0.224	9.8%	0.771	1.697	5.7%	−0.000	0.163	4.8%
		$\rho_{01}=0.5$	0.013	0.094	11.0%	−0.084	0.149	23.5%	0.004	0.079	9.3%
		$\rho_{02}=0.4$	−0.033	0.110	11.8%	−0.067	0.147	17.5%	−0.017	0.085	9.8%
	2	$\lambda_{01}=0.5$	−0.040	0.306	13.8%	0.758	0.924	17.8%	−0.082	0.224	7.0%
		$\lambda_{02}=0.4$	−0.222	0.419	11.7%	0.043	1.371	2.8%	−0.054	0.286	4.8%
		$\rho_{01}=0.15$	−0.008	0.127	10.5%	−0.032	0.116	8.0%	0.002	0.104	9.2%
		$\rho_{02}=0.1$	−0.051	0.182	11.1%	−0.027	0.122	10.7%	−0.014	0.112	11.7%
	3	$\lambda_{01}=0.15$	−0.022	0.370	10.4%	1.014	1.264	14.7%	−0.051	0.278	6.5%
		$\lambda_{02}=0.1$	−0.099	0.435	10.2%	−0.052	1.494	2.8%	0.011	0.332	5.7%
		$\rho_{01}=0.05$	−0.015	0.136	11.3%	−0.022	0.124	7.5%	−0.012	0.107	8.7%
		$\rho_{02}=0.02$	−0.046	0.183	13.9%	−0.013	0.126	9.9%	−0.014	0.102	8.6%
	4	$\lambda_{01}=0.5$	−0.041	0.302	13.9%	0.641	0.814	15.8%	−0.078	0.202	7.2%
		$\lambda_{02}=0.4$	−0.183	0.393	10.6%	0.281	1.476	3.4%	−0.039	0.262	3.9%
		$\rho_{01}=0.3$	0.001	0.115	8.1%	−0.050	0.122	10.1%	0.004	0.096	9.0%
		$\rho_{02}=0.2$	−0.055	0.175	9.4%	−0.033	0.121	11.0%	−0.017	0.101	10.4%
200	1	$\lambda_{01}=0.15$	0.011	0.118	6.7%	0.954	1.176	16.5%	−0.004	0.092	6.6%
		$\lambda_{02}=0.1$	−0.016	0.199	6.8%	−0.212	2.044	3.2%	−0.004	0.158	5.5%
		$\rho_{01}=0.5$	0.010	0.062	8.2%	−0.064	0.113	24.2%	0.005	0.055	7.2%
		$\rho_{02}=0.4$	−0.020	0.070	8.5%	−0.054	0.108	15.1%	−0.011	0.059	7.2%
	2	$\lambda_{01}=0.5$	−0.024	0.233	8.7%	0.840	0.992	18.0%	−0.066	0.184	7.6%
		$\lambda_{02}=0.4$	−0.222	0.381	9.2%	−0.142	2.037	2.3%	−0.051	0.282	2.9%
		$\rho_{01}=0.15$	−0.006	0.085	9.2%	−0.027	0.089	8.2%	0.000	0.079	11.7%
		$\rho_{02}=0.1$	−0.023	0.103	7.4%	−0.020	0.086	9.7%	−0.010	0.079	10.6%
	3	$\lambda_{01}=0.15$	−0.018	0.278	9.4%	1.142	1.357	16.2%	−0.021	0.211	7.5%
		$\lambda_{02}=0.1$	−0.111	0.404	11.2%	−0.033	2.183	2.7%	−0.010	0.353	6.8%
		$\rho_{01}=0.05$	−0.003	0.081	8.1%	−0.018	0.091	5.3%	−0.003	0.071	6.6%
		$\rho_{02}=0.02$	−0.017	0.098	9.3%	−0.007	0.084	8.5%	−0.007	0.073	7.1%
	4	$\lambda_{01}=0.5$	−0.020	0.235	10.1%	0.777	0.946	17.4%	−0.058	0.168	8.3%
		$\lambda_{02}=0.4$	−0.199	0.375	9.1%	−0.100	2.062	2.4%	−0.052	0.262	2.9%
		$\rho_{01}=0.3$	−0.003	0.069	8.1%	−0.040	0.087	11.4%	0.004	0.074	11.4%
		$\rho_{02}=0.2$	−0.028	0.099	7.2%	−0.028	0.088	9.6%	−0.016	0.078	11.0%

Note: All results are based on 1000 simulations and $P(5\%)$ denotes the empirical size of the two-sided 5% t-test of the specific parameter equal to its true value.

Airbnb Estimation Results from GMM and GS2SLS

Table S.4: SARAR Fitted to Airbnb Listings on March 21, 2020 in Asheville, NC

	GMM				GS2SLS			
Variable	Est.	ADI	AII	ATI	Est.	ADI	AII	ATI
$\boldsymbol{W}_{(0,20]}\boldsymbol{y}$	0.482				0.505			
	[14.410]				[6.678]			
$\boldsymbol{W}_{(20,50]}\boldsymbol{y}$	0.026				0.019			
	[0.627]				[0.356]			
$\boldsymbol{W}_{(50,100]}\boldsymbol{y}$	0.079				0.065			
	[1.838]				[1.124]			
Constant	−0.068				−0.104			
	[0.196]				[0.249]			
Suerhost	0.006	0.006	0.008	0.013	0.006	0.006	0.009	0.015
	[0.277]	[0.277]	[0.277]	[0.277]	[0.252]	[0.252]	[0.251]	[0.251]
HostCount	0.001	0.001	0.002	0.004	0.001	0.001	0.002	0.004
	[7.901]	[7.900]	[5.839]	[7.052]	[5.895]	[5.789]	[4.349]	[5.112]
EnHome	0.329	0.335	0.464	0.799	0.330	0.336	0.466	0.802
	[16.636]	[16.668]	[7.542]	[11.130]	[14.825]	[14.615]	[6.341]	[9.340]
Accomm	0.078	0.080	0.110	0.190	0.078	0.080	0.111	0.190
	[8.671]	[8.682]	[6.040]	[7.488]	[7.002]	[7.094]	[5.198]	[6.339]
Bdrms	0.010	0.010	0.014	0.024	0.011	0.011	0.015	0.026
	[0.404]	[0.404]	[0.403]	[0.403]	[0.355]	[0.355]	[0.354]	[0.355]
Barms	0.181	0.183	0.254	0.438	0.180	0.183	0.254	0.437
	[8.544]	[8.557]	[6.142]	[7.565]	[7.976]	[7.952]	[5.425]	[6.832]
DistCenter	−0.036	−0.037	−0.051	−0.088	−0.036	−0.037	−0.051	−0.087
	[11.120]	[11.136]	[7.003]	[9.294]	[8.882]	[9.074]	[5.938]	[7.757]
PremisePk	0.005	0.005	0.007	0.011	0.005	0.005	0.007	0.012
	[0.281]	[0.281]	[0.281]	[0.281]	[0.269]	[0.269]	[0.268]	[0.268]
AC	0.209	0.212	0.294	0.506	0.207	0.210	0.292	0.502
	[5.467]	[5.464]	[4.461]	[5.027]	[4.862]	[4.851]	[3.891]	[4.418]
TV	0.136	0.138	0.191	0.329	0.133	0.136	0.188	0.324
	[6.832]	[6.845]	[5.516]	[6.403]	[6.253]	[6.259]	[4.867]	[5.747]
Bkfst	0.073	0.074	0.102	0.176	0.070	0.072	0.099	0.171
	[3.296]	[3.295]	[3.136]	[3.261]	[2.756]	[2.761]	[2.693]	[2.772]
InsBook	0.076	0.077	0.107	0.184	0.076	0.077	0.107	0.183
	[4.719]	[4.720]	[4.136]	[4.512]	[4.159]	[4.141]	[3.563]	[3.914]
MinNights	−0.001	−0.001	−0.002	−0.004	−0.001	−0.001	−0.002	−0.004
	[1.321]	[1.321]	[1.328]	[1.330]	[1.169]	[1.169]	[1.191]	[1.186]
Reviews	−0.041	−0.042	−0.058	−0.099	−0.041	−0.042	−0.058	−0.099
	[11.277]	[11.299]	[6.873]	[9.155]	[9.401]	[9.266]	[5.704]	[7.498]
ReScore	0.011	0.011	0.015	0.026	0.011	0.011	0.016	0.027
	[3.239]	[3.238]	[2.987]	[3.140]	[2.879]	[2.881]	[2.585]	[2.748]
	1.016				1.018			
ADI(v)	[447.404]				[180.703]			
	1.408				1.411			
AII(v)	[8.114]				[6.928]			
	2.424				2.428			
ATI(v)	[13.899]				[11.809]			

Note: The absolute values of t-ratios are inside brackets. ADI, AII, and ATI denote respectively the average direct, indirect, and total impacts on the outcome from the observable covariates. ADI(v), AII(v), and ATI(v) denote the corresponding measures from the error innovation v.

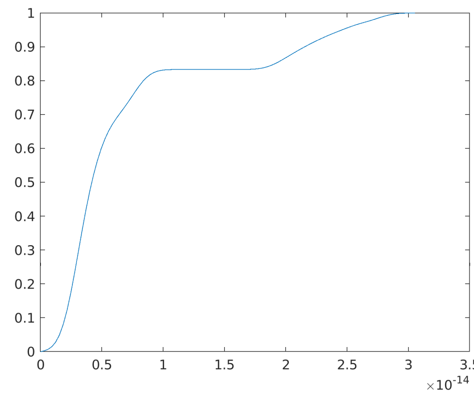
Table S.5: SARAR Fitted to Airbnb Listings on July 10, 2021 in Asheville, NC

	GMM				GS2SLS			
Variable	Est.	ADI	AII	ATI	Est.	ADI	AII	ATI
$\boldsymbol{W}_{(0,20]}\boldsymbol{y}$	0.231				0.462			
	[2.503]				[8.762]			
$\boldsymbol{M}_{(0,20]}\boldsymbol{u}$	0.650				0.288			
	[7.998]				[2.928]			
$\boldsymbol{M}_{(20,50]}\boldsymbol{u}$	0.340				0.265			
	[3.926]				[2.662]			
Constant	2.504				2.168			
	[2.300]				[3.945]			
Suerhost	0.066	0.067	0.020	0.086	0.049	0.049	0.041	0.090
	[2.823]	[2.824]	[1.659]	[2.740]	[1.944]	[1.945]	[1.852]	[1.939]
HostCount	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	[1.175]	[1.175]	[1.012]	[1.167]	[2.463]	[2.462]	[2.173]	[2.386]
EnHome	0.304	0.305	0.090	0.395	0.272	0.275	0.230	0.505
	[10.841]	[10.833]	[1.901]	[6.559]	[8.715]	[8.714]	[4.188]	[6.668]
Accomm	0.053	0.053	0.016	0.069	0.052	0.053	0.044	0.097
	[7.403]	[7.414]	[1.921]	[5.829]	[6.979]	[7.002]	[4.190]	[6.154]
Bdrms	0.118	0.119	0.035	0.154	0.095	0.096	0.080	0.177
	[6.052]	[6.047]	[1.814]	[4.768]	[4.619]	[4.616]	[3.258]	[4.159]
Barms	0.204	0.205	0.061	0.266	0.220	0.222	0.186	0.408
	[9.799]	[9.784]	[1.878]	[6.169]	[9.940]	[9.931]	[4.272]	[7.102]
DistCenter	0.049	0.049	0.015	0.064	−0.043	−0.044	−0.037	−0.081
	[1.253]	[1.254]	[1.152]	[1.273]	[5.950]	[5.968]	[4.006]	[5.503]
PremisePk	0.076	0.077	0.023	0.099	0.047	0.048	0.040	0.088
	[4.264]	[4.262]	[1.733]	[3.727]	[2.332]	[2.333]	[2.115]	[2.287]
AC	0.062	0.062	0.018	0.080	0.090	0.091	0.076	0.167
	[2.191]	[2.191]	[1.451]	[2.119]	[2.984]	[2.982]	[2.469]	[2.825]
TV	0.232	0.233	0.069	0.302	0.246	0.249	0.208	0.457
	[9.058]	[9.045]	[1.872]	[5.966]	[8.485]	[8.497]	[4.252]	[6.728]
Bkfst	0.078	0.078	0.023	0.101	0.110	0.112	0.093	0.205
	[2.686]	[2.684]	[1.524]	[2.508]	[3.391]	[3.388]	[2.662]	[3.146]
InsBook	0.024	0.024	0.007	0.031	0.042	0.042	0.035	0.077
	[1.348]	[1.348]	[1.114]	[1.334]	[2.165]	[2.165]	[1.958]	[2.111]
MinNights	−0.005	−0.005	−0.002	−0.007	−0.005	−0.005	−0.004	−0.010
	[2.641]	[2.639]	[1.501]	[2.457]	[2.179]	[2.180]	[2.038]	[2.164]
Reviews	−0.001	−0.001	0.000	−0.001	−0.001	−0.001	−0.001	−0.002
	[1.090]	[1.090]	[0.952]	[1.082]	[1.240]	[1.241]	[1.213]	[1.238]
ReScore	−0.007	−0.007	−0.002	−0.009	−0.007	−0.007	−0.006	−0.012
	[1.658]	[1.658]	[1.261]	[1.627]	[1.441]	[1.442]	[1.443]	[1.459]
ADI(v)	1.446				1.039			
	[5.909]				[2.403]			
AII(v)	128.509				3.112			
	[0.909]				[0.129]			
ATI(v)	129.955				4.151			
	[0.917]				[0.169]			

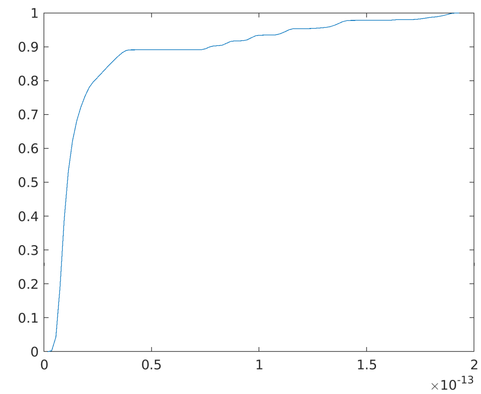
Note: The absolute values of t-ratios are inside brackets. ADI, AII, and ATI denote respectively the average direct, indirect, and total impacts on the outcome from the observable covariates. ADI(v), AII(v), and ATI(v) denote the corresponding measures from the error innovation v.

CDF's of Inverse Error Bound of $\Psi(\gamma)$ for Airbnb Data

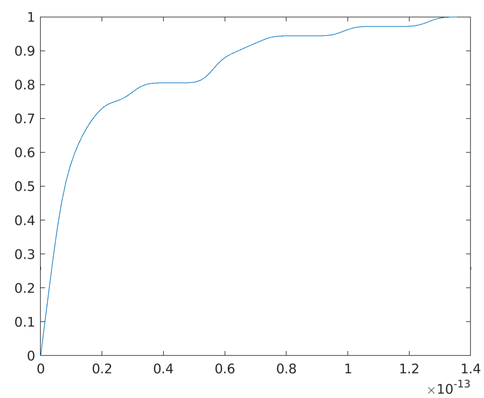
Figure S.1: Cumulative Distribution Functions of $\varepsilon(\mathbf{\Psi}^{-1}(\boldsymbol{\gamma}))$



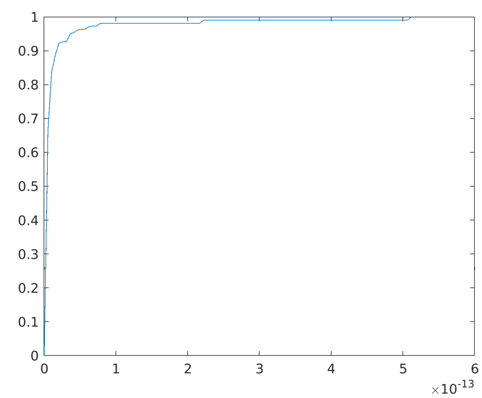
2020 Data with $\boldsymbol{W}_{(0,50]}$ and $\boldsymbol{W}_{(50,100]}$



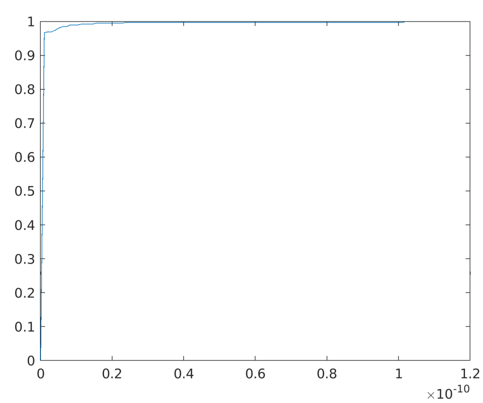
2020 Data with $\boldsymbol{W}_{(0,20]}$, $\boldsymbol{W}_{(20,50]}$, and $\boldsymbol{W}_{(50,100]}$



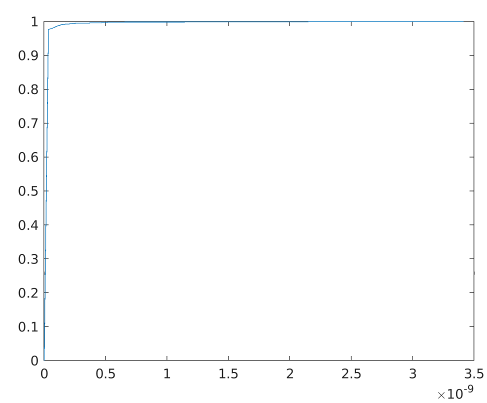
2021 Data with $\boldsymbol{W}_{(0,20]}$ and $\boldsymbol{M}_{(0,20]}$



2021 Data with $\boldsymbol{W}_{(0,50]}$ and $\boldsymbol{M}_{(0,50]}$



2021 Data with $\boldsymbol{W}_{(0,20]}$, $\boldsymbol{M}_{(0,20]}$, and $\boldsymbol{M}_{(20,50]}$



2021 Data with $\boldsymbol{W}_{(0,20]}$, $\boldsymbol{W}_{(20,50]}$, $\boldsymbol{M}_{(0,20]}$, and $\boldsymbol{M}_{(20,50]}$