Multivariate Spatial Autoregressive Model for Large Scale Social Networks

Xuening Zhu[1], Danyang Huang[2,*], Rui Pan[3], and Hansheng Wang[4]

[1] *School of Data Science, Fudan University, Shanghai, China;*

[2] *School of Statistics, Renmin University of China, Beijing, China;*

[3] *School of Statistics and Mathematics, Central University of Finance and Economics, Beijing, China;*

[4] *Guanghua School of Mangement, Peking University, Beijing, China;*

Abstract

The rapid growth of social network platforms generates a large amount of social network data, where multivariate responses are frequently collected from users. To statistically model such type of data, the multivariate spatial autoregressive (MSAR) model is studied. To estimate the model, the quasi maximum likelihood estimator (QMLE) is obtained under certain technical conditions. However, it is found that the computational cost of QMLE is expensive. To fix this problem, a least squares estimator (LSE) is developed. The corresponding identification conditions and asymptotic properties are investigated. To gauge the finite sample performance of various estimators, a number of simulation studies are conducted. Lastly, a Sina Weibo dataset is analyzed for illustration purpose.

JEL Classification: C510, C550

Key Words: Least Squares Estimator; Quasi Maximum Likelihood Estimator; Multivariate Responses; Social Network; Network Autoregression.

*Corresponding Author.
School of Statistics, Renmin University of China, Beijing 100872, P.R. China. Tel: 86-10-62511318. Fax: 86-10-62515246. Email: dyhuang@ruc.edu.cn.

1. INTRODUCTION

Spatial autoregressive (SAR) model is originally proposed for analyzing spatial data (Anselin, 2013; Banerjee et al., 2014). The SAR model assumes that observation from each spatial location is a weighted average of its spatial neighbours and a random noise. By doing so, the sophisticated spatial dependency could be modeled. Recently, this model also gains great popularity in social network analysis. This is because social network data are similar to spatial data, in the sense that observations from connected users are correlated. This makes the SAR model a candidate for network data analysis (Yang and Allenby, 2003; Chen et al., 2013; Liu, 2014; Cohen-Cole et al., 2018).

It is remarkable that the classical SAR model is designed for univariate response. In real practice, multivariate responses are typically encountered. As a result, there is a practical need to extend the SAR model to multivariate cases (Kelejian and Prucha, 2004; Liu, 2014; Cohen-Cole et al., 2018). Consider for example Sina Weibo (*www.weibo.com*), the largest Twitter-type social media in Chinese. One can treat the number of user posts as a univariate response. Then, the classical SAR model can be used to study the inter-dependency of activeness level among users. Further more, the user posts could be classified according to the contents (e.g., Finance, Economics). They naturally constitute a multivariate response for each user. There are also empirical examples showing that cross-response network effects could exist in individuals' behaviors. For example, Cohen-Cole et al. (2018) find that there are non-trivial within- and cross-choice peer effects for the students having higher grades and watching TV. To characterize this type of data, Yang and Lee (2017) study a multivariate spatial autoregressive (MSAR) model, and associate it with the simultaneous equations SAR (SESAR) models (Kelejian and Prucha, 2004; Baltagi and Bresson, 2011; De Graaff et al., 2012; Liu, 2014; Cohen-Cole et al., 2018). To estimate the model, a quasi maximum likelihood estimator (QMLE) is developed. Accordingly, the asymptotic theory is established. In this work, we re-study the QMLE but under a different set of technical conditions, which are more suitable for large scale social networks. The resulting asymptotic properties of QMLE are also re-investigated.

Although QMLE can be statistically efficient, the computational cost could be expensive. This is because the determinant of a high dimensional matrix needs to be computed, where the matrix dimension equals to the network size. In the meanwhile, it is typical for a social network to have millions (or even billions) of users. The huge network size (or matrix dimension) imposes a serious challenge on QMLE computation. On the other hand, real large scale networks are usually extremely sparse, because most network users are connected with only a very limited number of friends. As a result, more efficient algorithms can be developed for maximum likelihood estimation (Barry and Pace, 1999; Smirnov and Anselin, 2001; LeSage and Pace, 2007; Huang et al., 2016; Zhou et al., 2017). Under the condition that the network structure is sufficiently sparse, the consistent estimator can be designed and the computational complexity can be significantly reduced (Huang et al., 2016; Zhou et al., 2017). Besides the MLE-based approaches, other IV-based estimation methods, such as two stage least squares (2SLS) estimation and three stage least squares (3SLS) estimation, are also developed and widely used (Kelejian and Prucha, 2004; Baltagi and Deng, 2015; Cohen-Cole et al., 2018). Although the computational burden can be reduced, such methods could not work without exogenous variables, and can be less statistically efficient (Yang and Lee, 2017).

In this work, we propose to use a novel least squares method to estimate the MSAR model. The basic idea is to study the conditional expectation of the focal user, given the responses from the rest of the network. This leads to a least squares type objective function, which involves only two types of connections in network, and therefore can be practically optimized. As a consequence, the computational complexity can be greatly reduced. The identification issue is further investigated. Rigorous asymptotic theory is established. The theoretical findings and computational comparisons are corroborated by extensive simulation studies. A real data example about Sina Weibo is presented for illustration propose.

The rest of the article is organized as follows. Section 2 introduces the MSAR model and the parameter space. Section 3 investigates the parameter estimation and the corresponding asymptotic properties. Numerical studies are given in Section 4. The article is concluded with a brief discussion in Section 5. All technical details are left to the Appendix and a separate supplementary material.

2. MULTIVARIATE SPATIAL AUTOREGERESSION

2.1. Model and Notations

Consider a large scale network with N nodes. To describe the network structure, define an adjacency matrix $A = (a_{i_1i_2}) \in \mathbb{R}^{N\times N}$, where $a_{i_1i_2} = 1$ if the i_1th node follows the i_2th node ($i_1 \neq i_2$), and $a_{i_1i_2} = 0$ otherwise. We always assume $a_{ii} = 0$, for $1 \leq i \leq N$. In addition, we define $W = (w_{i_1i_2}) \in \mathbb{R}^{N\times N}$ as the row-normalized adjacency matrix, where $w_{i_1i_2} = n_{i_1}^{-1} a_{i_1i_2}$ and $n_{i_1} = \sum_{i_2} a_{i_1i_2}$ is called nodal out-degree. For the ith node, assume that a p-dimensional continuous response vector $(Y_{i1}, \cdots, Y_{ip})^\top \in \mathbb{R}^p$ is recorded. Accordingly, let $\mathbb{Y} = (Y_{ij}) \in \mathbb{R}^{N\times p}$ be the response matrix collected from N nodes. In addition, define $\mathbb{Y}_j = (Y_{1j}, \cdots, Y_{Nj})^\top \in \mathbb{R}^N$ to be the jth column vector of $\mathbb{Y}$, where $1 \leq j \leq p$. Correspondingly, we include exogenous variables to enhance the interpretability of the responses (LeSage, 2008; Anselin, 2013).

Specifically, let $\mathbb{X} = (\mathbb{X}_1, \cdots, \mathbb{X}_q) \in \mathbb{R}^{N\times q}$ be the q-dimensional exogenous covariates of the N nodes.

To model $\mathbb{Y}_j$, the following MSAR model is considered. For $1 \le j \le p$,

$$\mathbb{Y}_j = d_{jj} W\mathbb{Y}_j + \sum_{j'\neq j}^{p} d_{j'j} W\mathbb{Y}_{j'} + \sum_{k=1}^{q} b_{kj}\mathbb{X}_k + \mathcal{E}_j, \tag{2.1}$$

where $d_{j'j}$ and b_{kj} for $1 \le j, j' \le p, 1 \le k \le q$ are unknown parameters, and $\mathcal{E}_j = (\varepsilon_{1j}, \cdots, \varepsilon_{Nj})^\top \in \mathbb{R}^N$ is the noise vector from the jth response. Specifically, $\varepsilon_i = (\varepsilon_{i1}, \cdots, \varepsilon_{ip})^\top \in \mathbb{R}^p$ $(1 \le i \le N)$ is assumed to be identically and independently distributed with mean $\mathbf{0}_p \in \mathbb{R}^p$ and $\mathrm{cov}(\varepsilon_i) = \Sigma_e \in \mathbb{R}^{p\times p}$. In model (2.1), the first term $W\mathbb{Y}_j$ reflects the impact from connected neighbors but the response itself. Accordingly, the parameter d_{jj} is referred to as *intra-activity effect*. In addition to that, the quantity $W\mathbb{Y}_{j'}$ with $j' \neq j$ characterizes the impact due to other responses, where the corresponding parameter $d_{j'j}$ $(j' \neq j)$ is called *extra-activity effect*. It can be noted that the intra- and extra-activity effects are also referred to as *endogenous effect* and *cross-activity peer effect*, respectively. (Cohen-Cole et al., 2018). Without the second term, model (2.1) degenerates to a classical SAR model. Lastly, the term $\sum_{k=1}^{q} b_{kj}\mathbb{X}_k$ reflects the influence of the exogenous variables, where the corresponding coefficient can be referred to as *own effect* (Liu, 2014; Cohen-Cole et al., 2018). In addition, the model is referred to as pure MSAR model if the exogenous variables are not included.

Remark 1. Note that only own covariate of each node is included in (2.1). In practice, *contextual effect* from peers' influence could also be considered, which yields the following model (Liu, 2014; Cohen-Cole et al., 2018),

$$\mathbb{Y}_j = d_{jj} W\mathbb{Y}_j + \sum_{j'\neq j}^{p} d_{j'j} W\mathbb{Y}_{j'} + \sum_{k=1}^{q} b_{1,kj}\mathbb{X}_k + \sum_{k=1}^{q} b_{2,kj} W\mathbb{X}_k + \mathcal{E}_j, \tag{2.2}$$

where $\{b_{1,kj} : 1 \le k \le q, 1 \le j \le p\}$ and $\{b_{2,kj} : 1 \le k \le q, 1 \le j \le p\}$ reflect own effects and contextual effects from peers respectively. This makes the model more flexible. In fact, the estimation method of $\{b_{2,kj} : 1 \le k \le q, 1 \le j \le p\}$ is similar to $\{b_{1,kj} : 1 \le k \le q, 1 \le j \le p\}$. This is because once $W\mathbb{X}_k$s are calculated, they could also be treated as exogenous variables as $\mathbb{X}_k$s, for $1 \le k \le q$. As a result, the generalization to model (2.2) could be further made through the same estimation technique. Therefore for the sake of simplicity, we focus on model (2.1) to discuss its statistical properties in the rest of this article.

Recall that the response matrix is $\mathbb{Y} = (\mathbb{Y}_1, \cdots, \mathbb{Y}_p)$. Write $D = (d_{j'j}) \in \mathbb{R}^{p\times p}$ and $B = (b_{kj}) \in \mathbb{R}^{q\times p}$, which collect all the regression coefficients. Note that D is usually asymmetric. Then model (2.1) could be re-written in matrix form as

$$\mathbb{Y} = W\mathbb{Y}D + \mathbb{X}B + \mathbb{E}, \tag{2.3}$$

where $\mathbb{E} = (\mathcal{E}_1, \cdots, \mathcal{E}_p) \in \mathbb{R}^{N\times p}$ is the matrix of noise terms.

Remark 2. Note the MSAR model (2.3) can be also linked to the simultaneous equations SAR (SESAR) model (Cohen-Cole et al., 2018; Yang and Lee, 2017). The SESAR model takes the form as

$$\mathbb{Y}\Gamma = W\mathbb{Y}D + \mathbb{X}B + \mathbb{E}, \tag{2.4}$$

where parameter $\Gamma \in \mathbb{R}^{p\times p}$ is introduced as *simultaneity effect* (Cohen-Cole et al., 2018). Intuitively, the MSAR model can be obtained by setting $\Gamma = I_p$, where I_p is the $(p \times p)$-dimensional identity matrix. The SESAR model is empirically widely used and the estimation methods are investigated (Kelejian and Prucha, 2004; Yang and Lee, 2017). The detailed discussion about SESAR model and MSAR model can be found in Yang and Lee (2017). In this article, we concentrate on the MSAR model, which allows us to be more focus on the network effect D and own effect B.

2.2. Parameter Space

In order to study the feasible parameter space of model (2.3), we re-organize the MSAR model into a vector form as

$$\mathcal{Y} = (D^\top \otimes W)\mathcal{Y} + \widetilde{\mathbb{X}}\beta + \mathcal{E}, \tag{2.5}$$

where $\mathcal{Y} = \text{vec}(\mathbb{Y}) = (\mathbb{Y}_1^\top, \mathbb{Y}_2^\top, \cdots, \mathbb{Y}_p^\top)^\top \in \mathbb{R}^{Np}$, $\widetilde{\mathbb{X}} = I_p \otimes \mathbb{X}$, $\mathcal{E} = \text{vec}(\mathbb{E}) = (\mathcal{E}_1^\top, \mathcal{E}_2^\top, \cdots, \mathcal{E}_p^\top)^\top \in \mathbb{R}^{Np}$, $\beta = \text{vec}(B) \in \mathbb{R}^{pq}$, and $\otimes$ is the kronecker product. We define $I_n \in \mathbb{R}^{n\times n}$ to be the $(n \times n)$-dimensional identity matrix. As a result, it could be derived that $\mathcal{Y} = (I_{Np} - D^\top \otimes W)^{-1}(\widetilde{\mathbb{X}}\beta + \mathcal{E})$. Let $\lambda_j(D)$ be the jth eigenvalue of D such that $|\lambda_1(D)| \geq |\lambda_2(D)| \geq \cdots \geq |\lambda_p(D)|$. In order to ensure the matrix $(I_{Np} - D^\top \otimes W)$ to be invertible, a sufficient condition is given in Lemma 1.

Lemma 1. *Assume $|\lambda_1(D)| < 1$, then $(I_{Np} - D^\top \otimes W)$ is invertible.*

The proof of Lemma 1 is given in Section 1 in the separate supplementary material. We then assume the condition $|\lambda_1(D)| < 1$ throughout the rest of this article. By the invertibility of $(I_{Np} - D^\top \otimes W)$, the covariance of $\mathcal{Y}$ could be written as $\Sigma = \text{cov}(\mathcal{Y}) = (I_{Np} - D^\top \otimes W)^{-1}(\Sigma_e \otimes I_N)(I_{Np} - D \otimes W^\top)^{-1}$. To obtain more insights of Σ, we consider two special cases with $\Sigma_e = \sigma^2 I_p$ and $B = \mathbf{0}_{q,p}$, where $\mathbf{0}_{q,p} \in \mathbb{R}^{q\times p}$ with all elements equal to 0.

Case 1. (D is diagonal) In this case, D can be written as $D = \text{diag}\{d_{11}, d_{22}, \cdots, d_{pp}\}$. Consequently, (2.1) becomes $\mathbb{Y}_j = d_{jj} W\mathbb{Y}_j + \mathcal{E}_j$ $(1 \leq j \leq p)$, in which the jth response is only correlated to itself. As a result, the multiple responses can be modeled separately.

It can be easily obtained that Σ has a block diagonal structure, where $\Sigma = \sigma^2 \mathrm{diag}\{\Sigma_{11}, \cdots, \Sigma_{pp}\}$, where $\Sigma_{jj} = (I_N - d_{jj}W)^{-1}(I_N - d_{jj}W^\top)^{-1}$.

Case 2. (First Order Taylor's Expansion) In practice, the elements in D (i.e., $d_{j'j}$s) are usually small for a large scale social network (Chen et al., 2013; Zhu et al., 2017). This enables us to approximate Σ by its first order Taylor's expansion with respect to D as

$$\Sigma = \sigma^2(I_{Np} - D^\top \otimes W)^{-1}(I_{Np} - D \otimes W^\top)^{-1} \approx \sigma^2\big\{I_{Np} + D^\top \otimes W + D \otimes W^\top\big\}.$$

As a result, for $1 \le i_1, i_2 \le N$, $1 \le j_1, j_2 \le p$, we have

$$\mathrm{cov}(Y_{i_1j_1}, Y_{i_2j_2}) \approx \sigma^2 I(i_1 = i_2, j_1 = j_2) + \sigma^2(d_{j_2j_1}n_{i_1}^{-1}a_{i_1i_2} + d_{j_1j_2}n_{i_2}^{-1}a_{i_2i_1}), \qquad (2.6)$$

where $I(\cdot)$ is the indicator function. By (2.6), we obtain the following interesting observations. If $i_1 \neq i_2$ or $j_1 \neq j_2$, the approximation of $\mathrm{cov}(Y_{i_1j_1}, Y_{i_2j_2})$ is $\sigma^2(d_{j_2j_1}n_{i_1}^{-1}a_{i_1i_2} + d_{j_1j_2}n_{i_2}^{-1}a_{i_2i_1})$. This implies higher covariance between $Y_{i_1j_1}$ and $Y_{i_2j_2}$ if (a) the network effect between j_1th and j_2th response (i.e., $d_{j_1j_2}$ and $d_{j_2j_1}$) is large, and (b) node i_1 and i_2 are mutually connected (i.e., $a_{i_1i_2} = a_{i_2i_1} = 1$). The correlation can be stronger if i_1 and i_2 are *loyal* to each other, which implies they both have small out-degrees (i.e., n_{i_1} and n_{i_2}).

3. PARAMETER ESTIMATION

3.1. Maximum Likelihood Estimation

Let $\Omega_e = \Sigma_e^{-1}$, $S = I_{Np} - D^\top \otimes W$, and $\widetilde{\mathbb{X}} = I_p \otimes \mathbb{X}$. In addition, denote $\mathcal{D} = \mathrm{vec}(D) \in \mathbb{R}^{p^2}$ and $\xi_e = \mathrm{vec}^*(\Omega_e) \in \mathbb{R}^{p(p+1)/2}$, where $\mathrm{vec}^*(\Omega_e) \in \mathbb{R}^{p(p+1)/2}$ selects only up-triangle parameters of Ω_e, since the covariance matrix Σ_e has to be symmetric. Let $\theta = (\mathcal{D}^\top, \beta^\top, \xi_e^\top) \in \mathbb{R}^{n_{pq}}$ collect the parameters to be estimated, where $n_{pq} = p^2 + p(p+1)/2 + pq$. Given the form of MSAR model, one could write down the quasi log-likelihood function as

$$\ell(\theta) = \log |S| - (N/2) \log |\Sigma_e| - (1/2)(S\mathcal{Y} - \widetilde{\mathbb{X}}\beta)^\top (\Omega_e \otimes I_N)(S\mathcal{Y} - \widetilde{\mathbb{X}}\beta), \qquad (3.1)$$

where some constants are ignored. Denote $\hat{\theta}_M = \arg\max_\theta \ell(\theta) \in \mathbb{R}^{n_{pq}}$ to be the QMLE. The asymptotic properties of $\hat{\theta}_M$ have been studied by Yang and Lee (2017) under certain technical conditions, which fit spatial data quite well. However, some of the conditions might be too stringent for a large scale social network. For example, the column sum of the weighting matrix W can hardly be bounded for a large scale social network.

To relax the technical conditions, we re-investigate the asymptotic properties of QMLE by focusing on large scale social network data with different technical conditions.

(C1) (Network Structure)

- (C1.1) (Connectivity) Let the set of all the nodes $\{1, \cdots, N\}$ be the state space of a Markov chain, with the transition probability given by W. The Markov chain is assumed to be irreducible and aperiodic. In addition, define $\pi = (\pi_i)^\top \in \mathbb{R}^N$ to be the stationary distribution vector of the Markov chain (i.e., $\pi_i \geq 0$, $\sum_i \pi_i = 1$, and $W^\top \pi = \pi$). Assume $\sum_{i=1}^N \pi_i^2 = O(N^{-1/2-\delta})$, where $0 < \delta \leq 1/2$ is a positive constant.
- (C1.2) (Uniformity) Assume $|\lambda_1(W^*)| = O(\log N)$, where W^* is symmetric, and defined as $W^* = W + W^\top$.

(C2) (Law of Large Numbers) Assume the limits of certain network features exist, which are listed in (A.6) to (A.10) in Appendix A.1.

(C3) (COVARIATES) For an arbitrary vector $\beta \in \mathbb{R}^{pq}$, assume $N^{-1}\|\widetilde{\mathbb{X}}\beta\|^2 = O(1)$ as $N \to \infty$.

(C4) (NOISE TERM) Let Σ_e be decomposed as $\Sigma_e = (\Sigma_e^{1/2})^\top(\Sigma_e^{1/2})$. Let $\widetilde{\varepsilon}_i = (\Sigma_e^{1/2\top})^{-1}\varepsilon_i = (\widetilde{\varepsilon}_{i1}, \cdots, \widetilde{\varepsilon}_{ip})^\top \in \mathbb{R}^p$. Assume $E(\widetilde{\varepsilon}_{ij}^4) = \kappa_4$, and $E(\widetilde{\varepsilon}_{ij_1}\widetilde{\varepsilon}_{ij_2}\widetilde{\varepsilon}_{ij_3}) = 0$ for $1 \le i \le N$ and $1 \le j_1, j_2, j_3, j \le p$, where κ_4 and σ^2 are finite constants.

Condition (C1) imposes assumptions on the network structure, which can be divided into two separate conditions. Condition (C1.1) requires certain connectivity for network structure. More specifically, one can verify that both irreducibility and aperiodicity could be satisfied if the network is fully connected after a finite number of steps. This condition is satisfied if the well known six degrees of separation theory (Newman et al., 2006) holds. Condition (C1.2) imposes certain uniformity assumption on network structure. Compared with Assumption 4 of Yang and Lee (2017), which requires the row and column sums of W to be bounded, (C1.2) allows $\lambda(W^*)$ to be slowly diverging with the rate of $O(\log N)$. Next, condition (C2) is a law of large numbers type condition. This condition basically insures the convergence of certain network features as $N \to \infty$. Subsequently, conditions (C3) and (C4) set regularity conditions on exogenous covariates and noise terms respectively. The assumptions are to facilitate the asymptotic analysis and the adoption of the central limit theorem. One could generalize the exogenous covariates assumption (C4) to stochastic $\widetilde{\mathbb{X}}$ with moment conditions. Consequently, we have the following theorem.

Theorem 1. *Assume the conditions (C1)-(C4), we have the conclusion that* $\sqrt{N}(\hat{\theta}_M - \theta) \to_d N(\mathbf{0}_{n_{pq}}, (\Sigma_2^M)^{-1}\Sigma_1^M(\Sigma_2^M)^{-1})$ *as* $N \to \infty$, *where* Σ_1^M *and* Σ_2^M *are assumed to be* *positive definite matrices as*

$$\Sigma_2^M = \begin{pmatrix} \Sigma_{2d}^M & \Sigma_{2d\beta}^M & \Sigma_{2de}^M \\ \Sigma_{2d\beta}^{M\top} & \Sigma_{2\beta}^M & \mathbf{0}_{pq,\frac{p(p+1)}{2}} \\ \Sigma_{2de}^{M\top} & \mathbf{0}_{\frac{p(p+1)}{2},pq} & \Sigma_{2e}^M \end{pmatrix}, \quad \Sigma_1^M = \Sigma_2^M + \begin{pmatrix} \Delta_{1d}^M & \mathbf{0}_{p^2,pq} & \Delta_{1de}^M \\ \mathbf{0}_{pq,p^2} & \mathbf{0}_{pq,pq} & \mathbf{0}_{pq,\frac{p(p+1)}{2}} \\ \Delta_{1de}^{M\top} & \mathbf{0}_{\frac{p(p+1)}{2},pq} & \Delta_{1e}^M \end{pmatrix}. \tag{3.2}$$

The formula of the asymptotic covariance in (3.2) is given in Appendix A.1.

The proof of Theorem 1 is given in Appendix B.1. By Theorem 1, we know that QMLE is $\sqrt{N}$-consistent. Although the QMLE can be statistically efficiently in most occasions, its computational cost can be high. We then develop a novel estimator in the following section to reduce the computational complexity.

3.2. Least Squares Estimation

As we mentioned before, the computational cost of QMLE is high. This is mainly because the determinant of $(I_{Np} - D^\top \otimes W) \in \mathbb{R}^{Np\times Np}$ is involved and its computational complexity is of $O(N^3)$. To fix this problem, we develop here a novel least squares estimation method. Let $\mathbb{Y}_{-(i_1j_1)} = \{Y_{ij} : (i,j) \neq (i_1, j_1)\}$. To illustrate the method, we first discuss the conditional mean of $Y_{i_1j_1}$ given $\mathbb{Y}_{-(i_1j_1)}$ under the assumption that $\mathcal{Y}$ follows multivariate normal distribution. Then we prove the consistency and asymptotic normality under the general non-normal case.

Denote $M_{j\cdot}$ and $M_{\cdot j}$ as the jth row and column vector of M respectively. First it could be verified that the conditional mean $E\big\{Y_{i_1j_1}|\mathbb{Y}_{-(i_1j_1)}\big\}$ takes the form as

$E\{Y_{i_1j_1}|\mathbb{Y}_{-(i_1j_1)}\} = \mu_{i_1j_1} + \sum_{(i_2,j_2)\neq(i_1,j_1)}^{(N,p)} \alpha_{i_1j_1i_2j_2}(Y_{i_2j_2} - \mu_{i_2j_2})$, where

$$\alpha_{i_1j_1i_2j_2} = \frac{D_{j_1\cdot}^\top \Omega_{e,\cdot j_2} w_{i_2i_1} + \Omega_{j_2\cdot}^\top D_{\cdot j_1} w_{i_1i_2} - (D_{j_1\cdot}^\top \Omega_e D_{j_2\cdot})(\sum_{i=1}^N w_{ii_1} w_{ii_2})}{\omega_{e,j_1j_1} + (D_{j_1\cdot}^\top \Omega_e D_{j_1\cdot})(\sum_{i=1}^N w_{ii_1}^2)}, \qquad (3.3)$$

and $\mu_{ij} = E(Y_{ij})$. The verification of (3.3) is given in Section 2 in the separate supplementary material. Even though the quantity in (3.3) is analytically complicated, it is computationally feasible due to the sparsity of the network structure. Specifically, the summation involved in (3.3) only counts when: (1) nodes i_1 and i_2 are directly connected (i.e., $w_{i_1i_2} \neq 0$ or $w_{i_2i_1} \neq 0$) as shown in the left panel of Figure 1; or (2) i_1 and i_2 are indirectly connected through a third node i such that $\sum_i w_{ii_1} w_{ii_2} \neq 0$; see the 2-out-star structure in the right panel of Figure 1. Therefore, the computational cost can be greatly reduced since only the "second order friends" are involved in the computation. It is a common phenomenon for the real social network to be sufficiently sparse, which indicates the number of connected edges ($a_{i_1i_2} = 1$) can be extremely small. As a result, if the number of nodes connected with i (either directly or indirectly) are limited, the computational complexity could even be as low as $O(N)$.

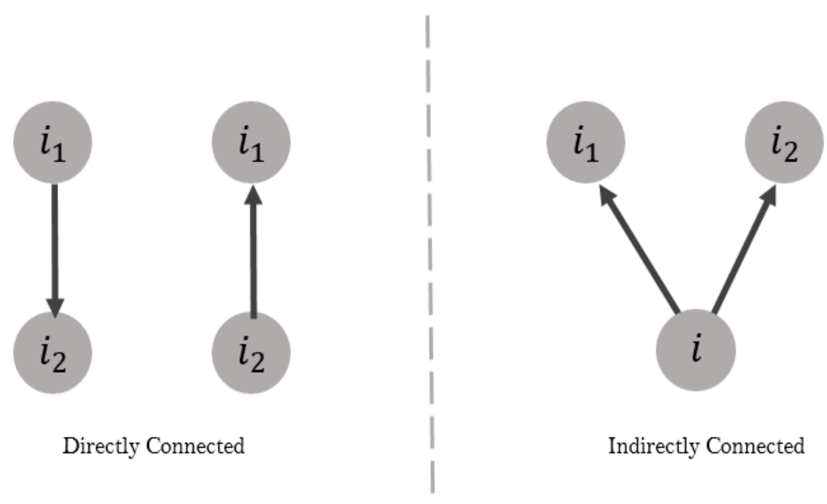

Figure 1: Two types of nodes involved in (3.3). The left panel: directly connected nodes (i.e., $w_{i_1i_2} \neq 0$ or $w_{i_2i_1} \neq 0$); The right panel: indirectly connected nodes (i.e., $\sum_i w_{ii_1} w_{ii_2} \neq 0$).

Let $\mathbb{Y}^* = \big\{E(Y_{ij}|\mathbb{Y}_{-(ij)}) : 1 \le i \le N, 1 \le j \le p\big\} \in \mathbb{R}^{N\times p}$ and $\mathcal{Y}^* = \text{vec}(\mathbb{Y}^*)$. Recall $S = I_{Np} - D^\top \otimes W$. Subsequently, define $\Omega = \Sigma^{-1} = S^\top(\Sigma_e^{-1} \otimes I_N)S$ and $m = \text{diag}^{-1}(\Omega) = \{\text{diag}(\Sigma_e^{-1}) \otimes I_N + \text{diag}(D\Sigma_e^{-1}D^\top) \otimes \text{diag}(W^\top W)\}^{-1}$. Then it can be verified that

$$\mathcal{Y}^* = U - m(\Omega - m^{-1})(\mathcal{Y} - U),$$

where $U = E(\mathcal{Y}) = S^{-1}(I_p \otimes \mathbb{X})\beta$. Accordingly, a least squares type objective function can be constructed as

$$Q(\theta) = \|\mathcal{Y} - \mathcal{Y}^*\|^2 = \Big\|mS^\top(\Sigma_e^{-1} \otimes I)\{S\mathcal{Y} - \widetilde{\mathbb{X}}\beta\}\Big\|^2. \quad (3.4)$$

Verification of (3.4) is given in Section 3 in the separate supplementary material. It is noteworthy that the objective function (3.4) only involves S, instead of its determinant and inverse form. In addition, with regards to the computation of $m = \text{diag}^{-1}(\Omega)$, one should note that although it involves an inverse of a giant matrix, it is not hard to compute. This is because $\text{diag}(\Omega)$ actually takes a diagonal form, whose inverse can be computed by taking the inverse of each diagonal element. This only requires low computational complexity. Moreover, only the previous two types of friends (i.e., Figure 1) are involved in computation. This makes the computation feasible even in large scale social network. By minimizing (3.4), the least squares type estimator (LSE) $\hat{\theta}_L = \arg\min_\theta Q(\theta)$ can be obtained.

Next, in the following two sections, we discuss the identification issue and the asymptotic properties of the proposed LSE.

3.3. Identification of LSE

In this section, we discuss the identification issue of LSE in the MSAR model (2.5). To motivate the discussion, we consider a univariate SAR model $Y = \rho_0 WY + \mathbb{X}\beta_0 + \mathcal{E}$, where $Y \in \mathbb{R}^N$ is the response vector, $\mathbb{X} \in \mathbb{R}^{N\times q}$ is the exogenous covariate matrix, ρ_0 and $\beta_0 \in \mathbb{R}^q$ are corresponding true parameters. Lee (2004) has mentioned that under this model specification, Y can be represented as

$$Y = \mathbb{X}\beta_0 + \rho_0 W(I_N - \rho_0 W)^{-1}\mathbb{X}\beta_0 + (I_N - \rho_0 W)^{-1}\mathcal{E}.$$

Intuitively, to guarantee the identification of ρ_0, $\mathbb{X}\beta_0$ and $W(I_N - \rho_0 W)^{-1}\mathbb{X}\beta_0$ should not be multicollinear. Theoretically, this statement is rigorously proved by Lee (2004) as a sufficient condition for the global identification of $(\rho_0, \beta_0^\top)^\top$ for the quasi-maximum likelihood estimation. For the multivariate response, the identification issue has also been established and frequently discussed (Kelejian and Prucha, 2004; Liu, 2014; Yang and Lee, 2017). We refer to the Section 3 of Yang and Lee (2017) for detailed discussion. As a result, we focus on the identification issue of θ for the LSE.

To tackle this problem, we follow the technique of Yang and Lee (2017) and calculate the expected least squares objective function $Q(\theta)$. For convenience, denote D_0, β_0, Σ_{e0}, and θ_0 to be the true parameter. In addition, let $S_0 = I_{Np} - D_0^\top \otimes W$ and $\Sigma_0 = S_0^{-1}(\Sigma_{e0} \otimes I_N)(S_0^\top)^{-1}$. Define $\mathbb{Q}(\theta) = E\{Q(\theta)\}$. It can be computed that,

$$\mathbb{Q}(\theta) = \Big\| mS^\top(\Sigma_e^{-1} \otimes I)(SS_0^{-1}(I_p \otimes \mathbb{X})\beta_0 - (I_p \otimes \mathbb{X})\beta)\Big\|^2 + \mathrm{tr}\Big(m\Omega\Sigma_0\Omega m\Big). \quad (3.5)$$

It suffices to show $\liminf_{n\to\infty} \min_{\theta\in\overline{B}_\epsilon(\theta_0)} 1/N\{\mathbb{Q}(\theta) - \mathbb{Q}(\theta_0)\} > 0$, where $\overline{B}_\epsilon(\theta_0)$ is the complement of an open neighborhood of θ_0 of diameter ε (White, 1996, Theorem 3.4). Define J_i $(i = 1, \cdots, p)$ to be a $1 \times p$ row vector with all zero elements except for the ith entry, which is 1. Moreover, let $\mathbb{X}_1^* = (J_1 \otimes I_N)S_0^{-1}\{I_p \otimes (W\mathbb{X})\}\beta_0, \cdots, \mathbb{X}_p^* = (J_p \otimes I_N)S_0^{-1}\{I_p \otimes (W\mathbb{X})\}\beta_0$, $\mathbb{X}^* = (\mathbb{X}_1^*, \cdots, \mathbb{X}_p^*) \in \mathbb{R}^{N\times p}$, and $\widetilde{\mathbb{X}}^* = (\mathbb{X}^*, \mathbb{X}) \in \mathbb{R}^{N\times(p+q)}$. We then make the following assumption.

(C5) (Identification Condition) Assume the limit $\lim_{N\to\infty} N^{-1}(\widetilde{\mathbb{X}}^{*\top}\widetilde{\mathbb{X}}^*)$ exists and is nonsingular.

It can be noted that the condition (C5) corroborates with the identification condition in the existing literatures (Liu, 2014; Cohen-Cole et al., 2018; Yang and Lee, 2017). Moreover, sufficient conditions under the social network setting are discussed and given by Liu (2014) and Cohen-Cole et al. (2018). They establish the identification results when the network connections have certain intransitivity properties. For the LSE, we show that the intra- and extra-activity effects (D) and the own effects (B) can be identified through the following theorem.

Theorem 2. *Assume there exists $\delta > 0$ such that*

$$\min_{|\lambda_1(D)|\le 1-\delta} \Big\{\lambda_{\min}(SS^\top)\Big\} \ge \tau, \tag{3.6}$$

where τ is a positive constant. Under condition (C5), we have that D_0 and β_0 can be identified in the parameter space $\{D : |\lambda_1(D)| \le 1-\delta\}$.

The proof of Theorem 2 is given in Section 6.1 in the supplementary material. Note that condition (C5) could only guarantee the identification of D_0 and β_0. In some cases, the condition (C5) can be violated, e.g., the pure MSAR model with no exogenous covariates. In such a situation, we establish a second identification result which could be employed to identify D_0 and the implied covariance structure Σ_{e0}^*, where $\Sigma_{e0}^* = \Sigma_{e0}/\mathrm{tr}(\Sigma_e)$. To this end, we first give the definition of the separability of the MSAR model.

Definition 1. *The MSAR model (2.3) is defined to be separable, if there exists two non-overlapped index sets* $\mathcal{I} = \{i_1, \cdots, i_{m_1}\}$ *and* $\mathcal{J} = \{j_1, \cdots, j_{m_2}\}$ *such that*

$$\mathbb{Y}^{(1)} = W\mathbb{Y}^{(1)}D^{(1)} + \mathbb{X}B^{(1)} + \mathbb{E}^{(1)} \quad \textit{and} \quad \mathbb{Y}^{(2)} = W\mathbb{Y}_1 D^{(2)} + \mathbb{X}B^{(2)} + \mathbb{E}^{(2)},$$

where $\mathbb{Y}^{(1)} = (\mathbb{Y}_i, i \in \mathcal{I}) \in \mathbb{R}^{N \times m_1}$, $D^{(1)} = (D_{i_1 i_2}, i_1, i_2 \in \mathcal{I}) \in \mathbb{R}^{m_1 \times m_1}$, $B^{(1)} = (B_{ki}, 1 \leq k \leq N, i \in \mathcal{I}) \in \mathbb{R}^{q \times m_1}$, $\mathbb{E}^{(1)} = (\mathcal{E}_j, j \in \mathcal{I}) \in \mathbb{R}^{N \times m_1}$, *and* $\mathbb{Y}^{(2)}$, $D^{(2)}$, $B^{(2)}$, $\mathbb{E}^{(2)}$ *are defined similarly by* $\mathcal{J}$. *In addition,* $\mathbb{E}^{(1)}$ *and* $\mathbb{E}^{(2)}$ *are uncorrelated with each other.*

The definition of separability essentially implies that the MSAR model in (2.3) can be segmented into at least two uncorrelated equations. Under the definition, we have the following result.

Theorem 3. *(a) Assume* I, W, $W^\top$, *and* $W^\top W$ *are linearly independent. Then* D_0 *can be identified. (b) In addition to (a), if* $p \geq 2$ *and model (2.3) is not separable, then* D_0 *and* Σ_{e0}^* *can be globally identified.*

The proof of Theorem 3 is given in Section 6.2 in the supplementary material. By Theorem 3, we know that D_0 and Σ_{e0}^* can be identified if the network structure satisfies certain conditions. Moreover, it can be noted that the identification condition in Theorem 3 is consistent with the theoretical result in existing literature, i.e., see Proposition 6 for the QMLE by Yang and Lee (2017).

3.4. Asymptotic Property of LSE

Under the identifiability of the MSAR model, we then investigate the asymptotic properties of the proposed LSE. To this end, we first introduce some notations. Let $\widetilde{S} = (\Omega_e \otimes I_N)S$. In addition, define $F = m\widetilde{S}^\top\{S\mathcal{Y} - (I_p \otimes \mathbb{X})\beta\}$, thus we have $Q(D, \beta, \Sigma_e) = F^\top F$. Denote $Q^d_{j_1j_2} = \partial Q(\theta)/\partial d_{j_1j_2}$, $Q_\beta = \partial Q(\theta)/\partial\beta$ to be the first order derivatives of $Q(\theta)$ with respect to $d_{j_1j_2}$ and β. In addition, one could define the second order derivatives of $Q(\theta)$ as $Q^d_{j_1j_2k_1k_2} = \partial^2 Q(\theta)/\partial d_{j_1j_2}\partial d_{k_1k_2}$, $Q^d_{j_1j_2\beta} = \partial^2 Q(\theta)/\partial d_{j_1j_2}\partial\beta$, $Q_{\beta\beta} = \partial Q(\theta)/\partial\beta\partial\beta^\top$. The first and second order derivatives of other matrices/vectors can be defined similarly, e.g., for F, m, Ω, respectively. Define $\gamma = (\mathcal{D}^\top, \beta^\top)^\top \in \mathbb{R}^{p^2+pq}$. In this part, we focus on the asymptotic properties of $\hat\gamma_L = (\hat{\mathcal{D}}_L^\top, \hat\beta_L^\top)^\top \in \mathbb{R}^{p^2+pq}$. For convenience, $\hat\gamma_L$ is referred to as the LSE thereafter. To this end, we require the following conditions.

(C6) (Law of Large Numbers) Assume the following limits exist, which are, $\lim_{N\to\infty} N^{-1}\text{cov}(Q^d_{j_1j_2}, Q^d_{k_1k_2}) = \Sigma^{(t_1,t_2)}_{1d}$, $\lim_{N\to\infty} N^{-1}\text{cov}(Q^d_{j_1j_2}, Q_\beta) = \Sigma^{(t_1)}_{1d\beta} \in \mathbb{R}^{pq}$, $\lim_{N\to\infty} N^{-1}\text{cov}(Q_\beta, Q_\beta) = \Sigma_{1\beta} \in \mathbb{R}^{(pq)\times(pq)}$, and $\lim_{N\to\infty} N^{-1}E(Q^d_{j_1j_2k_1k_2}) = \Sigma^{(t_1,t_2)}_{2d}$, $\lim_{N\to\infty} N^{-1}E(Q^d_{j_1j_2\beta}) = \Sigma^{(t_1)}_{2d\beta} \in \mathbb{R}^{pq}$, $\lim_{N\to\infty} N^{-1}E(Q^d_{\beta\beta}) = \Sigma_{2\beta} \in \mathbb{R}^{(pq)\times(pq)}$, with $t_1 = (j_1-1)p + j_2$ $(1 \le j_1, j_2 \le p)$ and $t_2 = (k_1-1)p + k_2$ $(1 \le k_1, k_2 \le p)$. The detailed matrix forms are given in (3.9) to (3.12) in the following.

Theorem 4. *Assume conditions (C1)–(C6) hold. In addition, let* $\Sigma_{1d} = (\Sigma^{(t_1,t_2)}_{1d} : 1 \le t_1, t_2 \le p^2)$, $\Sigma_{1d\beta} = (\Sigma^{(t_1)}_{1d\beta} : 1 \le t_1 \le p^2)^\top \in \mathbb{R}^{p^2\times(pq)}$, $\Sigma_{2d} = (\Sigma^{(t_1,t_2)}_{2d} : 1 \le t_1, t_2 \le p^2) \in \mathbb{R}^{p^2\times p^2}$, $\Sigma_{2d\beta} = (\Sigma^{(t_1)}_{2d\beta} : 1 \le t_1 \le p^2) \in \mathbb{R}^{p^2\times(pq)}$. *As* $N \to \infty$, *we have*

$$\sqrt{N}(\hat\gamma_L - \gamma) \to_d N(\mathbf{0}_{m_{pq}}, (\Sigma^L_2)^{-1}\Sigma^L_1(\Sigma^L_2)^{-1}), \tag{3.7}$$

where we have $m_{pq} = p^2 + pq$, *and*

$$\Sigma^L_1 = \begin{pmatrix} \Sigma_{1d} & \Sigma_{1d\beta} \\ \Sigma^\top_{1d\beta} & \Sigma_{1\beta} \end{pmatrix}, \quad \Sigma^L_2 = \begin{pmatrix} \Sigma_{2d} & \Sigma_{2d\beta} \\ \Sigma^\top_{2d\beta} & \Sigma_{2\beta} \end{pmatrix}, \tag{3.8}$$

where the detailed formula of (3.8) is given in the following by (3.9) to (3.12).

The proof of Theorem 4 is given in Appendix B.2. By Theorem 4, we know that the LSE $\hat{\gamma}_L$ is also $\sqrt{N}$-consistent.

We next show the detailed derivation of the asymptotic covariance in (3.8). Let $\widetilde{\mathcal{E}} = (\widetilde{\mathcal{E}}_1^\top, \widetilde{\mathcal{E}}_2^\top, \cdots, \widetilde{\mathcal{E}}_p^\top)^\top = (\Sigma_e^{-1/2} \otimes I_N)\mathcal{E}$, where $\widetilde{\mathcal{E}}_k = (\tilde{\varepsilon}_{1k}, \tilde{\varepsilon}_{2k}, \cdots, \tilde{\varepsilon}_{Nk})^\top \in \mathbb{R}^N$. Therefore we have $\mathrm{cov}(\widetilde{\mathcal{E}}) = I_{Np}$. Define $M_1 = m\widetilde{S}^\top(\Sigma_e^{1/2} \otimes I_N)$, $M_{2,j_1j_2} = (m^d_{j_1j_2}\widetilde{S}^\top + m\widetilde{S}^{d\top}_{j_1j_2} + m\widetilde{S}^\top S^d_{j_1j_2}S^{-1})\}(\Sigma_e^{1/2} \otimes I_N)$, $M_{3,j_1j_2} = m\widetilde{S}^\top S^d_{j_1j_2}S^{-1}$, and $J_{1,j_1j_2} = M_{3,j_1j_2}(I_p \otimes \mathbb{X})\beta$. We then have $F = M_1\widetilde{\mathcal{E}}$, $F^d_{j_1j_2} = M_{2,j_1j_2}\widetilde{\mathcal{E}} + J_{1,j_1j_2}$. The detailed expressions are given in Appendix A.2. Moreover, it can be verified $F_\beta = -m\widetilde{S}^\top(I_p \otimes \mathbb{X})$. In addition, define $M_{j_1j_2} = M_1^\top M_{2,j_1j_2}$, $J_{j_1j_2} = M_1^\top J_{1,j_1j_2}$, $H = 2F_\beta^\top M_1$. It can be verified that $Q^d_{j_1j_2} = 2\widetilde{\mathcal{E}}^\top M_{j_1j_2}\widetilde{\mathcal{E}} + 2\widetilde{\mathcal{E}}^\top J_{j_1j_2}$ and $Q_\beta = H\widetilde{\mathcal{E}}$. We then have

$$\begin{aligned}\Sigma_{1d}^{(t_1,t_2)} = \lim_{N\to\infty} N^{-1}\Big\{&4\mathrm{tr}(M_{j_1j_2}M_{k_1k_2}^\top) + 4\mathrm{tr}(M_{j_1j_2}M_{k_1k_2}) \\ &+ 4\mathrm{tr}\{\mathrm{diag}(M_{j_1j_2})\mathrm{diag}(M_{k_1k_2})\}(\kappa_4 - 3) + 4J_{j_1j_2}^\top J_{k_1k_2}\Big\},\end{aligned} \tag{3.9}$$

$$\Sigma_{1d\beta}^{(t_1)} = 2\lim_{N\to\infty} N^{-1}(HJ_{j_1j_2}), \quad \Sigma_{1\beta} = \lim_{N\to\infty} N^{-1}(HH^\top), \tag{3.10}$$

$$\Sigma_{2d}^{(t_1,t_2)} = 2\lim_{N\to\infty} N^{-1}\big\{\mathrm{tr}(M_{2,j_1j_2}^\top M_{2,k_1k_2}) + J_{1,j_1j_2}^\top J_{1,k_1k_2}\big\}, \tag{3.11}$$

$$\Sigma_{2d\beta}^{(t_1)} = 2\lim_{N\to\infty} N^{-1}F_\beta^\top J_{1,j_1j_2}, \quad \Sigma_{2\beta} = 2\lim_{N\to\infty} N^{-1}F_\beta^\top F_\beta. \tag{3.12}$$

The detailed verifications of (3.9) to (3.12) are given in Section 4 in the separate supplementary material.

4. NUMERICAL STUDIES

4.1. Simulation Models

To demonstrate the finite sample performance of the proposed two methods, we present three simulation examples. The main difference lies in the generating mechanism of the network structure A (i.e., W). For the noise matrix, we consider two different cases, where ε_i is generated independently: (1) multivariate normal distribution with mean $\mathbf{0}_2$ and covariance $\Sigma_e = (0.4, 0.1; 0.1, 0.6) \in \mathbb{R}^{2\times 2}$; (2) t-distribution with degree 5 and the same mean and covariance as (1). Subsequently, for each node we sample an exogenous covariate $X_i = (X_{i1}, \cdots, X_{i2})^\top \in \mathbb{R}^2$ from a multivariate normal distribution with mean $\mathbf{0}_2$ and $\Sigma_z = (\sigma_{j_1 j_2}) \in \mathbb{R}^{2\times 2}$, where $\sigma_{j_1 j_2} = 0.5^{|j_1 - j_2|}$. The corresponding network autoregression coefficient and contextual effect are fixed D as $D = (0.3, 0; -0.2, 0.1)$ and $B = (-0.5, 1.3; 1, 0.3)$. Lastly, the response $\mathcal{Y}$ is generated according to $\mathcal{Y} = (I - D^\top \otimes W)^{-1}(\widetilde{\mathbb{X}}\beta + \mathcal{E})$. We then consider the following three examples.

Example 1. (Dyad Independence Network) We follow Holland and Leinhardt (1981) to define a dyad as $\mathcal{A}_{ij} = (a_{ij}, a_{ji})$ $(1 \le i < j \le N)$ and assume different $\mathcal{A}_{ij}$s are independent. In order to reflect network sparsity, we set $P(\mathcal{A}_{ij} = (1,1)) = 20N^{-1}$. Therefore, the expected number of the mutually connected dyads (i.e., $\mathcal{A}_{ij} = (1,1)$) is $O(N)$. Next, we allow the expected degree of each node to be slowly diverging in the order of $O(N^{0.2})$ by setting $P(\mathcal{A}_{ij} = (1,0)) = P(D_{ij} = (0,1)) = 0.5N^{-0.8}$. As a result, the probability of forming a null dyad should be $P(\mathcal{A}_{i,j} = (0,0)) = 1 - 20N^{-1} - N^{-0.8}$, which is close to 1 as the network size N is large.

Example 2. (Stochastic Block Network) We next consider the stochastic block network (Wang and Wong, 1987; Nowicki and Snijders, 2001), which is another popular network structure and of particular interest for community detection (Zhao et al., 2012). To generate the block network structure, we follow Nowicki and Snijders (2001) to randomly assign a block label for each node $(k = 1, \cdots, K)$, where $K = 10, 20, 50$ is the total number of blocks. Then, let $P(a_{ij} = 1) = 0.9N^{-1}$ if i and j belong to the same block, and $P(a_{ij} = 1) = 0.3N^{-1}$ otherwise. As a result, nodes within the same block are more likely to be connected.

Example 3. (Power-Law Distribution Network) In a social network, it is commonly observed that the majority of nodes have few links but a small proportion have a large amount of links (Barabási and Albert, 1999). The number of links usually follow the power-law distribution (Clauset et al., 2009). To mimic this phenomenon, we simulate the adjacency matrix A according to Clauset et al. (2009) as follows. First, we generate the in-degree $m_i = \sum_j a_{ji}$ for node i by the discrete power-law distribution, i.e., $P(m_i = k) = ck^{-\alpha}$ with a normalizing constant c and exponent parameter $\alpha = 2.5$. Then, for the ith node, m_i nodes are randomly selected as its followers.

4.2. Performance Measurements and Simulation Results

We consider different network sizes (i.e., $N = 100, 200, 500$) for each simulation example, and then replicate the experiment for $R = 500$ times. The proposed estimators (i.e., QMLE and LSE) are compared in each example. Let $\hat{D}^{(r)} = (\hat{d}_{jk}^{(r)})$ be the estimator from the rth replication. The following measures are then employed to evaluate the finite sample performance. For a given parameter d_{jk} $(1 \le j, k \le p)$, the root mean square error (RMSE) is calculated by $\mathrm{RMSE}_{jk} = \{R^{-1}\sum_{r=1}^{R}(\hat{d}_{jk}^{(r)} - d_{jk})^2\}^{1/2}$ to gauge the estimation accuracy. Next, a 95% confidence interval is constructed for d_{jk} as $\mathrm{CI}_{jk}^{(r)} = (\hat{d}_{jk}^{(r)} - z_{0.975}N^{-1}\widehat{\mathrm{SE}}_{jk}^{(r)}, \hat{d}_{jk}^{(r)} + z_{0.975}N^{-1}\widehat{\mathrm{SE}}_{jk}^{(r)})$, where $\widehat{\mathrm{SE}}_{jk}^{(r)}$ is the $\{t = (j-1)p + k\}$th diagonal element of the asymptotic covariance in (3.2) and (3.7) by plugging into the QMLE and LSE respectively, and z_α is the αth lower quantile of a standard normal distribution. Then, the coverage probability is computed as $\mathrm{CP}_{jk} = R^{-1}\sum_{r=1}^{R} I(\hat{d}_{jk}^{(r)} \in \mathrm{CI}_{jk}^{(r)})$, where $I(\cdot)$ is the indicator function. In addition, the average CPU time and the network density (i.e., $\{N(N-1)\}^{-1}\sum_{i_1,i_2} a_{i_1 i_2}$) are also reported. Lastly, to further compare the computational efficiency of QMLE and LSE, we conduct the same experiment with larger sample size (i.e., N from 200 to 2500) for $R = 100$ replicates. The average CPU time of QMLE and LSE are given in Figure 2.

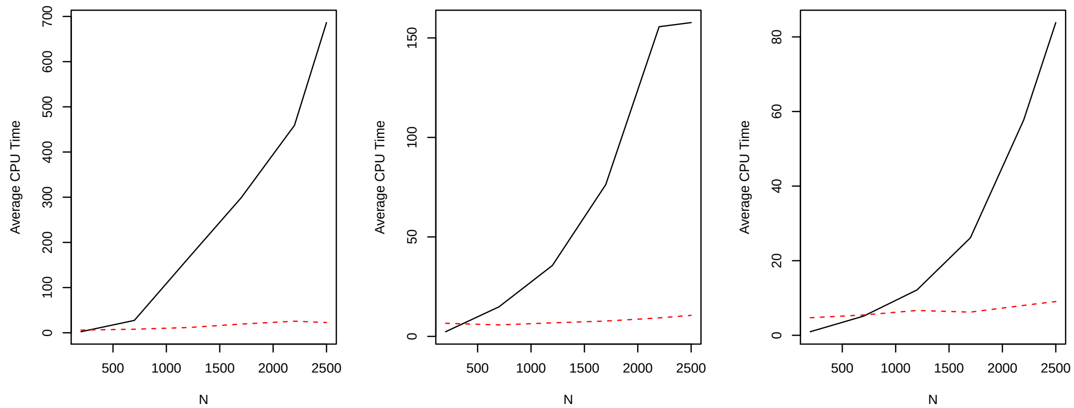

Figure 2: The average CPU time measured in second for dyad independence network (left panel), stochastic block network (middle panel), and power-law distribution network (right panel) for 100 replicates. The black solid line is the average CPU time for QMLE and the red dashed one for LSE.

The detailed results are summarized in Tables 1 to 3. As N increases, the RMSEs of both QMLE and LSE decrease. Take the dyad independence network for example, the RMSE for d_{11} of LSE drops from 0.185 to 0.128 as N increases from 200 to 500 for the normal distribution of $\mathcal{E}$. Moreover, the difference of RMSEs between QMLE and LSE is sufficiently small (e.g. the difference of RMSE for d_{12} between QMLE and LSE is 0.002 for power-law distribution network as $N = 1000$ in Case 1), which implies the estimation efficiency of LSE is almost as high as QMLE. Besides, the coverage probability of both estimators are stable at the nomial level 95%, which implies the estimated standard error $\widehat{\mathrm{SE}}_{jk}$ approximates the true standard error SE_{jk} well. Lastly, in terms of the computational time, the LSE is found to be much faster than the computation of QMLE especially when n is large. Furthermore, the polynomially increasing computational time of QMLE compared to LSE is illustrated in Figure 2.

4.3. A Sina Weibo Dataset

We next elaborate the MSAR model using a dataset collected from Sina Weibo (*www.weibo.com*), which is the largest Twitter-type social network in Chinese. The followers of an MBA official Weibo account are collected and their relationships are recorded. We investigate the users' online posting behaviours related to finance and economics. First, the posts are tagged as FINANCE or ECONOMICS if corresponding keywords are involved, where the keywords are obtained from an online public Chinese dictionary. The finance dictionary contains mostly keywords of stock markets and products. The economic dictionary contains keywords related to the macroeconomic trend and policies. Second, the log-transformed number of characters in the posts related to FINANCE and ECONOMICS topics are aggregated respectively, for each user within 75 days.

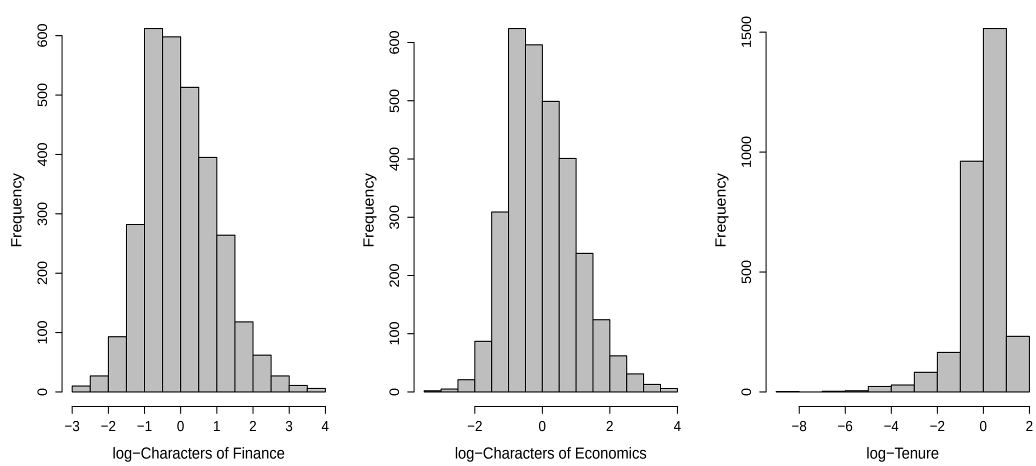

Figure 3: The left panel: histogram of log-Characters of Finance; The middle panel: histogram of log-Characters of Economics; The right panel: histogram of log-Tenure. All the continuous variables are standardized with mean 0 and variance 1.

Next, we include several personal information as the exogenous nodal covariates.

The first is the GENDER of the user, which is equal to 1 if the user is male. The second is the geographical information. Specifically, we consider two indicator variables. One is whether the user is located in BEIJING, and the other is whether the user is located in SHANGHAI. The last one is the TENURE of the user on the Sina Weibo platform, which is the time length since the user's registration with Sina Weibo.

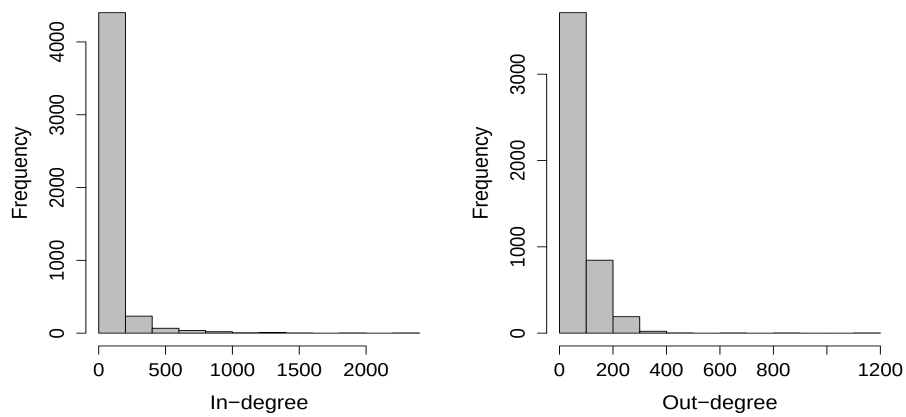

Figure 4: The Sina Weibo data analysis. The left panel: histogram of in-degrees. The highly right skewed shape indicates the existence of "super stars" in the network; The right panel: similar histogram but for out-degrees.

Lastly, we define the adjacency matrix A by the following-followee relationship among users. The histograms of in- and out-degrees are shown in Figure 4. The distribution of in-degrees is more skewed than the out-degrees. As one can see, most users have limited followers (in-degrees), but a few number of users own large number of followers. They are mostly public social media accounts and celebrities. To eliminate the "super star" effects in the dataset, the nodes with highest degrees are deleted and $N = 3,018$ nodes are involved in the analysis. The histograms of all the continuous variables are given in Figure 3, where we standardize all continuous variables with mean 0 and variance 1. The sample correlations between $\mathbb{Y}_{j_1}$ and $W\mathbb{Y}_{j_2}$ are given in

Table 4 for $1 \leq j_1, j_2 \leq p$, where a positive correlation level could be observed. The dataset includes about 68.0% male users, 61.5% users located in Beijing and 13.6% in Shanghai.

We then fit the MSAR model by QMLE and LSE respectively. The estimation results are summarized in Table 5. For each estimation method, the estimates are reported and are marked by "*" if the estimate is significant under the level $\alpha = 0.05$. A positive intra-activity effect is observed for Weibo posting contents about FINANCE and ECONOMICS. In addition, a significant positive extra-activity effect is observed from FINANCE to ECONOMICS, i.e., a user's posting activities in ECONOMICS is positively related to his/her following friends' posts in FINANCE. Lastly, with respect to the nodal covariates, it is found that the male users are more active in posting activities related to both FINANCE and ECONOMICS. No significant evidence is observed with regards to the geographical information and tenure of the user. Lastly, the computational time for the estimation of LSE is 55 seconds, which is much less than QMLE (i.e., 691 seconds).

5. CONCLUDING REMARKS

In this article, we study two estimators (i.e., QMLE and LSE) of MSAR model for large scale social networks. Specifically, the technical conditions are derived and the asymptotic properties are established. Particularly, an extent of heterogeneity for nodal in-degrees is allowed. Although the QMLE is usually statistically efficiently, it can be computationally infeasible for large scale social network. In the meanwhile, large scale social network are typically sparse. This enables us to reduce the computational cost and propose the LSE. It is proved the LSE is computationally more efficient than QMLE, by both simulation studies and extensive practical analysis.

To conclude this work, we consider here several interesting topics for future studies.

First, it is assumed the disturbances of the MSAR model exhibit no heteroskedasticity among network nodes. However, one could verify that the QMLE and LSE can be biased when heteroskedasticity exists in the covariance structure of $\mathbb{E}$. In this case, one might either assume a parametric structure on $\mathbb{E}$, or design certain estimation methods as the generalized least squares (GLS) estimation or generalized moment methods (GMM) to deal with the heteroskedasticity (Kelejian and Prucha, 2004; Lin and Lee, 2010; Baltagi and Bresson, 2011; Liu and Saraiva, 2017). It could be an interesting topic to investigate solutions under the LSE framework. Second, many responses in the real practice are observed in time series. Therefore, the time dynamics could be taken into consideration and statistically modeled. Third, it is noteworthy that the MSAR model requires the responses to be continuous. However, the discrete responses are frequently encountered in real data analysis, and needs to be further investigated. Next, one might note that there could be sample selection issue (Heckman, 1976) here since only part of the network is observed. In practice, it is very rare that the whole network is available especially the network scale is large. There is a growing interest of researches to deal with such issues in single-activity network model; see Liu (2013), Sojourner (2013), Liu et al. (2017) for more discussions. It is important to investigate the network sample selection issue under the multi-activity network model. Lastly, it is required by the MSAR model that the dimension of responses (i.e., p) is fixed. While the dimension of responses in the real practice can be sufficiently high. This could result in critical issues in both estimation and computation. How to solve this problem should be a challenging and intriguing topic for the future study.

APPENDIX A

In Appendix A, we give some basic matrix forms in Appendix A.1 and Appendix A.2. Next, we give five useful lemmas which will be employed in the rest of the proofs in Appendix A.3.

Appendix A.1: Asymptotic Covariance of QMLE in (3.2)

In this section, we give the detailed expression of asymptotic covariance matrix in (3.2). We first give the formula of the asymptotic covariance in terms of some constants. Then we give the constants definition as follows. Let $\Sigma_{2d}^M = (\Sigma_{2d}^{M(t_1,t_2)}) \in \mathbb{R}^{p^2\times p^2}$, $\Sigma_{2d\beta}^M = (\Sigma_{2d\beta}^{M(t_1)} : 1 \le t_1 \le p^2)^\top \in \mathbb{R}^{p^2\times pq}$, $\Sigma_{2\beta}^M = \kappa_\beta$, $\Sigma_{2e}^M = (\Sigma_{2e}^{M(t_1,t_2)}) \in \mathbb{R}^{p^2\times p^2}$, $\Sigma_{2de}^M = (\Sigma_{2de}^{M(t_1,t_2)}) \in \mathbb{R}^{p^2\times p^2}$. Then we have

$$\Sigma_{2d}^{M(t_1,t_2)} = \kappa^a_{j_1j_2,k_1k_2} + \kappa^b_{j_1j_2,k_1k_2} + \kappa^c_{j_1j_2,k_1k_2}, \quad \Sigma_{2d\beta}^{M(t_1)} = \alpha_{j_1j_2} \in \mathbb{R}^{pq} \tag{A.1}$$

$$\Sigma_{2e}^{M(t_1,t_2)} = \nu^a_{j_1j_2,k_1k_2}, \quad \Sigma_{2de}^{M(t_1,t_2)} = \xi^a_{j_1j_2,k_1k_2}, \tag{A.2}$$

for $t_1 = (j_1-1)p + j_2$ $(1 \le j_1, j_2 \le p)$ and $t_2 = (k_1-1)p + k_2$ $(1 \le k_1, k_2 \le p)$. In addition, we have

$$\Delta_{1d}^M = (\Sigma_{1d}^{M(t_1,t_2)}) \in \mathbb{R}^{p^2\times p^2} \text{ with } \Delta_{1d}^{M(t_1,t_2)} = \kappa^d_{j_1j_2,k_1k_2}, \tag{A.3}$$

$$\Delta_{1e}^M = (\Sigma_{1e}^{M(t_1,t_2)}) \in \mathbb{R}^{p^2\times p^2} \text{ with } \Delta_{1e}^{M(t_1,t_2)} = \nu^b_{j_1j_2,k_1k_2}, \tag{A.4}$$

$$\Delta_{1de}^M = (\Sigma_{1de}^{M(t_1,t_2)}) \in \mathbb{R}^{p^2\times p^2} \text{ with } \Delta_{1de}^{M(t_1,t_2)} = \xi^b_{j_1j_2,k_1k_2}, \tag{A.5}$$

for $t_1 = (j_1-1)p + j_2$ $(1 \le j_1, j_2 \le p)$ and $t_2 = (k_1-1)p + k_2$ $(1 \le k_1, k_2 \le p)$.

We then give the constants in (A.1) to (A.5). Recall that $\Sigma_e = \mathrm{cov}(\varepsilon_i)$ and $\Omega_e = \Sigma_e^{-1} = (\omega^e_{j_1j_2}) \in \mathbb{R}^{p\times p}$. We next define $G^d_{j_1j_2} = (\Omega_e^{1/2} \otimes I_N)(I_{j_2j_1} \otimes W)S^{-1}(\Sigma_e^{1/2} \otimes I_N)$,

$U^d_{j_1j_2} = (\Sigma_e^{1/2\top} \otimes I_N)(I_{j_2j_1} \otimes W)S^{-1}\widetilde{\mathbb{X}}\beta$, $L = \{\widetilde{\mathbb{X}}^\top(\Omega_e \otimes I_N)(\Sigma_e^{1/2} \otimes I_N)\}^\top$, and $G^e_{j_1j_2} = -1/2(\Sigma_e^{1/2\top} \otimes I_N)(\mathcal{I}_{j_1j_2} \otimes I_N)(\Sigma_e^{1/2} \otimes I_N)$, where $I_{j_1j_2} \in \mathbb{R}^{p\times p}$ is a zero matrix with only the (j_1, j_2)th element to be 1. Denote $\dot{\ell}^d_{j_1j_2}(\theta) = \partial\ell(\theta)/\partial d_{j_1j_2}$, $\dot{\ell}_\beta(\theta) = \partial\ell(\theta)/\partial\beta$, $\dot{\ell}^e_{j_1j_2}(\theta) = \partial\ell(\theta)/\partial\omega^e_{j_1j_2}$. It can be easily verified $\dot{\ell}^d_{j_1j_2}(\theta) = \widetilde{\mathcal{E}}^\top G^d_{j_1j_2}\widetilde{\mathcal{E}} + \widetilde{\mathcal{E}}^\top U^d_{j_1j_2} + c_d$, $\dot{\ell}_\beta(\theta) = L^\top\widetilde{\mathcal{E}}$, $\dot{\ell}^e_{j_1j_2}(\theta) = \widetilde{\mathcal{E}}^\top G^e_{j_1j_2}\widetilde{\mathcal{E}} + c_e$, where c_d and c_e are constants.

As assumed by Condition (C2), the following limits exist and are defined as

$$N^{-1}L^\top L \to \kappa_\beta,\ N^{-1}L^\top U^d_{j_1j_2} \to \alpha_{j_1j_2}, \tag{A.6}$$

$$N^{-1}\mathrm{tr}\big(G^d_{j_1j_2}G^d_{k_1k_2}\big) \to \kappa^a_{j_1j_2,k_1k_2},\ N^{-1}\mathrm{tr}\big(G^d_{j_1j_2}G^{d\top}_{k_1k_2}\big) \to \kappa^b_{j_1j_2,k_1k_2}, \tag{A.7}$$

$$N^{-1}U^{d\top}_{j_1j_2}U^d_{k_1k_2} \to \kappa^c_{j_1j_2,k_1k_2},\ N^{-1}\delta_4\mathrm{tr}\big\{\mathrm{diag}(G^d_{j_1j_2})\mathrm{diag}(G^d_{k_1k_2})\big\} \to \kappa^d_{j_1j_2,k_1k_2}, \tag{A.8}$$

$$2N^{-1}\mathrm{tr}(G^e_{j_1j_2}G^e_{k_1k_2}) \to \nu^a_{j_1j_2,k_1k_2},\ N^{-1}\delta_4\mathrm{tr}\big\{\mathrm{diag}(G^e_{j_1j_2})\mathrm{diag}(G^e_{k_1k_2})\big\} \to \nu^b_{j_1j_2,k_1k_2}, \tag{A.9}$$

$$2N^{-1}\mathrm{tr}(G^d_{j_1j_2}G^e_{k_1k_2}) \to \xi^a_{j_1j_2,k_1k_2},\ N^{-1}\delta_4\mathrm{tr}\{\mathrm{diag}(G^d_{j_1j_2})\mathrm{diag}(G^e_{j_1j_2})\} \to \xi^b_{j_1j_2,k_1k_2}, \tag{A.10}$$

as $N \to \infty$, where $\delta_4 = \kappa_4 - 3$.

Appendix A.2: Matrix Derivatives of LSE

We first summarize the basic matrix forms. Recall that $\Sigma_e = \mathrm{cov}(\varepsilon_i)$ and $\Omega_e = \Sigma_e^{-1}$, $S = I_{Np} - D^\top \otimes W$, $\widetilde{S} = (\Omega_e \otimes I_N)S$. Define $V = (D\Omega_e D^\top) \otimes (W^\top W)$, then we have $\Omega = \Omega_e \otimes I_N - (D\Omega_e) \otimes W^\top - (\Omega_e D^\top) \otimes W + V$. In addition, recall that $m = \mathrm{diag}^{-1}(\Omega) = \{\mathrm{diag}(\Omega_e) \otimes I_N + \mathrm{diag}(D\Omega_e D^\top) \otimes \mathrm{diag}(W^\top W)\}^{-1}$.

Next, we give the derivatives with respect to the basic matrices as follows,

$$V^d_{j_1j_2} = (I_{j_1j_2}\Omega_e D^\top + D\Omega_e I_{j_2j_1}) \otimes (W^\top W) \tag{A.11}$$

$$\Omega^d_{j_1j_2} = -(I_{j_1j_2}\Omega_e) \otimes W^\top - (\Omega_e I_{j_2j_1}) \otimes W + V^d_{j_1j_2} \tag{A.12}$$

$$m^d_{j_1j_2} = -m^2\mathrm{diag}(V^d_{j_1j_2}),\quad S^d_{j_1j_2} = -I_{j_2j_1} \otimes W,\quad \widetilde{S}^d_{j_1j_2} = -(\Omega_e I_{j_2j_1}) \otimes W. \tag{A.13}$$

$$m^d_{j_1j_2k_1k_2} = 2m^3\text{diag}(V^d_{k_1k_2})\text{diag}(V^d_{j_1j_2}) - m^2\text{diag}(V^d_{j_1j_2k_1k_2}) \tag{A.14}$$

$$V^d_{j_1j_2k_1k_2} = (I_{j_1j_2}\Omega_e I_{k_2k_1} + I_{k_1k_2}\Omega_e I_{j_2j_1}) \otimes (W^\top W) \tag{A.15}$$

$$\Omega^d_{j_1j_2k_1k_2} = V^d_{j_1j_2k_1k_2}. \tag{A.16}$$

Appendix A.3: Five Useful Lemmas

In this section, we give statements of five useful lemmas, i.e., Lemma 2 to Lemma 6. The detailed proofs are given in Section 5 in a separate supplementary material.

Lemma 2. *Let $\{V_i \in \mathbb{R}^1 : 1 \le i \le N\}$ be a set of identically distributed random variables. Assume that (a) $E(V_i) = 0$ for $1 \le i \le N$; (b) $E(V_i, V_j) = 0$ for any $i \ne j$; (c) $E(V_iV_jV_k) = 0$ for any $1 \le i, j, k \le N$; (d) $E(V_i^2) = 1$ and $E(V_i^4) = \kappa_4$, where κ_4 is a finite positive constant. Let $V = (V_1, V_2, \cdots, V_N)^\top \in \mathbb{R}^N$, $Q_1 = V^\top M_1 V + U_1^\top V$, and $Q_2 = V^\top M_2 V + U_2^\top V$, where $M_1 = (m_{1,ij}) \in \mathbb{R}^{N\times N}$ and $M_2 = (m_{2,ij}) \in \mathbb{R}^{N\times N}$ are $N \times N$ dimensional matrices, $U_1, U_2 \in \mathbb{R}^N$ are N-dimensional vectors. We then have*

$$cov(Q_1, Q_2) = tr(M_1M_2^\top) + tr(M_1M_2) + (\kappa_4 - 3)tr\big\{diag(M_1)\,diag(M_2)\big\} + U_1^\top U_2. \tag{A.17}$$

Lemma 3. *Define $|M|_e = (|m_{ij}|)$ for any arbitrary M. Further define the notation $M_1 \preccurlyeq M_2$ if $m^{(1)}_{ij} \le m^{(2)}_{ij}$, where $M_1 = (m^{(1)}_{ij}) \in \mathbb{R}^{n_1\times n_2}$ and $M_2 = (m^{(2)}_{ij}) \in \mathbb{R}^{n_1\times n_2}$ are two arbitrary matrices. Let $\mathbf{1}_n$ be the n-dimensional vector with all elements to be 1. Then we have the following results.*

(a) Assume $|\lambda_1(D)| < 1$. Then there exists a constant $c_d > 0$, such that

$$|D^m|_e \preccurlyeq c_d p^2 |\lambda_1(D)|^m \max\{m, p\}^p \mathbf{1}_p \mathbf{1}_p^\top. \tag{A.18}$$

(b) Assume condition (C1). Let $\Gamma \in \mathbb{R}^{p\times p}$ be an arbitrary $p \times p$ dimensional matrix.

Recall $S = I_{Np} - D^\top \otimes W$*, then we have*

$$|S^{-1}|_e \preccurlyeq c_0(\mathbf{1}_p\mathbf{1}_p^\top) \otimes \big(\mathbb{W}\big), \tag{A.19}$$

$$|(\Gamma \otimes W^q)S^{-1}|_e \preccurlyeq c_q(\mathbf{1}_p\mathbf{1}_p^\top) \otimes \big(\mathbb{W}_q\big), \tag{A.20}$$

where $\mathbb{W} = \sum_{m=0}^K W^m + \mathbf{1}_N\pi^\top$ *with* K *being a finite integer,* $\mathbb{W}_q = W^q\mathbb{W}$*,* π *is defined in condition (C1.1),* $c_0 = c_d p^2(K^p + c_\lambda c_w)$*,* $c_\lambda = \sum_{m=p}^\infty |\lambda_1(D)|^{m-p}m^p < \infty$*,* $c_q = c_\gamma p^q c_0$*,* $c_\gamma = \|\Gamma\|_\infty$*, and* $c_w > 1$ *is a constant.*

(c) For $0 \le q_1, \cdots, q_4 \le 1$*, and finite positive integers* r_1*,* r_2*,* r_3*, and* q*, we have*

$$\lambda_{\max}(W^\top W) = O\{(\log N)^2\}, \quad \lambda_{\max}(\mathbb{W}_q^\top \mathbb{W}_q) = O(\Delta_N), \tag{A.21}$$

$$N^{-2} tr\Big\{(W^{r_1}W^{\top r_2})^{r_3}\Big\} \to 0, \tag{A.22}$$

$$N^{-2} tr\Big\{(W^\top W)^{q_1}(\mathbb{W}_q^\top \mathbb{W}_q)^{q_2}(W^\top W)^{q_3}(\mathbb{W}_q^\top \mathbb{W}_q)^{q_4}\Big\} \to 0, \tag{A.23}$$

as $N \to \infty$*, where* $\Delta_N = (\log N)^{2(K+q)}$ *if* $\delta = 1/2$ *and* $\Delta_N = N^{1/2-\delta}$ *if* $0 < \delta < 1/2$*, and* δ *here is the positive constant defined in condition (C1.1).*

(d) For the basic matrices given in Appendix A.1, we have the upper bound as follows

$$|V_{j_1j_2}^d|_e \preccurlyeq c_{1v}(\mathbf{1}_p\mathbf{1}_p^\top) \otimes (W^\top W), \tag{A.24}$$

$$|V_{j_1j_2k_1k_2}^d|_e = |\Omega_{j_1j_2k_1k_2}^d|_e \preccurlyeq c_{2v}(\mathbf{1}_p\mathbf{1}_p^\top) \otimes (W^\top W), \tag{A.25}$$

$$|m|_e \preccurlyeq c_{0m} I_{Np}, \quad |m_{j_1j_2}^d|_e \preccurlyeq c_{1m}(\mathbf{1}_p\mathbf{1}_p^\top) \otimes (W^\top W), \tag{A.26}$$

$$|m_{j_1j_2k_1k_2}^d|_e \preccurlyeq c_{2m}(\mathbf{1}_p\mathbf{1}_p^\top) \otimes \big\{W^\top W + (W^\top W)^2\big\}, \tag{A.27}$$

$$|S|_e \preccurlyeq I_{Np} + c_{1s}(\mathbf{1}_p\mathbf{1}_p^\top) \otimes W, \quad |\widetilde{S}|_e \preccurlyeq c_{2s}(\mathbf{1}_p\mathbf{1}_p^\top) \otimes (I_N + W), \tag{A.28}$$

$$|S_{j_1j_2}^d|_e \preccurlyeq c_{3s}(\mathbf{1}_p\mathbf{1}_p^\top) \otimes W, \quad |\widetilde{S}_{j_1j_2}^d|_e \preccurlyeq c_{4s}(\mathbf{1}_p\mathbf{1}_p^\top) \otimes W. \tag{A.29}$$

In addition, we have the upper bound for the matrices defined in Section

$$|M_1|_e \preccurlyeq c_{1M}(\mathbf{1}_p\mathbf{1}_p^\top) \otimes (I_N + W^\top), \tag{A.30}$$

$$|M_{2,j_1j_2}|_e \preccurlyeq c_{2M}(\mathbf{1}_p\mathbf{1}_p^\top) \otimes \widetilde{\mathbb{W}}, \quad |M_{3,j_1j_2}|_e \preccurlyeq c_{2M}(\mathbf{1}_p\mathbf{1}_p^\top) \otimes \widetilde{\mathbb{W}}, \tag{A.31}$$

$$|M_{j_1j_2}|_e \preccurlyeq c_{3M}(\mathbf{1}_p\mathbf{1}_p^\top) \otimes (\widetilde{\mathbb{W}} + W^\top\widetilde{\mathbb{W}}), \tag{A.32}$$

$$|M_1^\top M_{3,j_1j_2}|_e \preccurlyeq c_{3M}(\mathbf{1}_p\mathbf{1}_p^\top) \otimes (\widetilde{\mathbb{W}} + W^\top\widetilde{\mathbb{W}}), \tag{A.33}$$

where $\widetilde{\mathbb{W}} = W^\top + W^\top W + W^\top W W^\top + W^\top \mathbb{W}_1$. *Furthermore, we have*

$$\lambda_{\max}(\widetilde{\mathbb{W}}^\top\widetilde{\mathbb{W}}) = O\big\{(\log N)^6 N^{1/2-\delta}\big\}. \tag{A.34}$$

Lemma 4. *Suppose* $\mathbb{E} = (\mathcal{E}_1, \cdots, \mathcal{E}_p) = (\varepsilon_{ik}) \in \mathbb{R}^{N\times p}$ *and* $E(\mathcal{E}_k, \mathcal{E}_l) = \sigma_{kl} I_N$. *In addition, let* $\mathbb{X} = (X_1, \cdots, X_q) = (X_{ik}) \in \mathbb{R}^{N\times q}$ *and* $\widetilde{\mathbb{X}} = I_p \otimes \mathbb{X}$. *Assume for any* $\beta \in \mathbb{R}^{pq}$ *and* $M \in \mathbb{R}^{(Np)\times(Np)}$, *we have* $|N^{-1}(\widetilde{\mathbb{X}}\beta)^\top M(\widetilde{\mathbb{X}}\beta)| \le c_\beta N^{-1} tr(M)$ *as* $N \to \infty$, *where* c_β *is a positive constant only related to* β. *Moreover, assume* $\max_k E(\varepsilon_{ik}^4) \le \kappa_4$, *where* κ_4 *is a finite positive constant. Let*

$$Q = \sum_{k=1}^{p}\sum_{l=1}^{p} \mathcal{E}_k^\top M_{kl}\mathcal{E}_l + \sum_{n=1}^{s}\sum_{k=1}^{p}\sum_{l=1}^{p} \mathcal{E}_k^\top U_{n,kl}(\mathbb{X}\beta_{n,l}),$$

where $M_{kl} = (M_{kl,ij}) \in \mathbb{R}^{N\times N}$, $U_{n,kl} \in \mathbb{R}^{N\times N}$, $\beta_{n,l} \in \mathbb{R}^q$, *and* $\sum_{k,l}\sigma_{kl} tr(M_{kl}) = 0$, *where* $1 \le k, l \le p$. *Let* $\mathbb{M} = (|M_{kl}|_e : 1 \le k, l \le p) \in \mathbb{R}^{(Np)\times(Np)}$ *and* $\mathbb{U}_n = (|U_{n,kl}|_e : 1 \le k \le p, 1 \le l \le q) \in \mathbb{R}^{(Np)\times(Nq)}$. *Then we have* $N^{-1/2}Q \to_d N(0, \sigma_1^2)$ *if*

$$N^{-2} tr\Big\{\mathbb{M}\mathbb{M}^\top\mathbb{M}\mathbb{M}^\top\Big\} \to 0, \tag{A.35}$$

$$N^{-1}\lambda_{\max}^2(\mathbb{U}_n\mathbb{U}_n^\top) \to 0, \tag{A.36}$$

as $N \to \infty$, *where* $\sigma_1^2 = \lim_{N\to\infty} N^{-1} var(Q)$.

Lemma 5. *Let $\ddot{\ell}(\theta) = \partial^2 \ell(\theta)/\partial\theta\partial\theta^\top$ be the second order derivative of $Q(\theta)$ with respect to θ. If we assume the same conditions in Theorem 1, then $-N^{-1}\ddot{\ell}(\theta) \to_p \Sigma_2^M$.*

Lemma 6. *Let $\ddot{Q}(\gamma) = \partial^2 Q(\gamma)/\partial\gamma\partial\gamma^\top$ be the second order derivative of $Q(\gamma)$ with respect to γ. If we assume the same conditions in Theorem 4, then $N^{-1}\ddot{Q}(\gamma) \to_p \Sigma_2^L$.*

APPENDIX B

In Appendix B, we first give the proof of Theorem 1 Then we give the proof of Theorem 4 in Appendix B.2.

Appendix B.1: Proof of Theorem 1

The proof will be accomplished in the following two steps. In the first step, we prove $\hat{\theta}_M$ is $\sqrt{N}$-consistent. Next, in the second step, we show that $\hat{\theta}_M$ is asymptotically normal.

STEP 1. To establish the consistency result, we follow Fan and Li (2001) to show that for any $\epsilon > 0$, there exists a constant $C > 0$ such that

$$\lim_{N\to\infty} P\Big\{ \sup_{\|u\|=C} \ell(\theta + N^{-1/2}u) < \ell(\theta) \Big\} > 1 - \epsilon. \tag{B.1}$$

Then (B.1) implies that with probability at least $1-\epsilon$, there exists a local optimizer $\hat{\theta}_M$ in the ball $\{\theta + N^{-1/2}uC : \|u\| \le 1\}$. As a result, we have $\|\hat{\theta}_M - \theta\| = O_p(N^{-1/2})$. Let $\dot{\ell}(\theta) = \partial\ell(\theta)/\partial\theta \in \mathbb{R}^{n_{pq}}$ and $\ddot{\ell}(\theta) = \partial^2\ell(\theta)/\partial\theta\partial\theta^\top \in \mathbb{R}^{n_{pq}\times n_{pq}}$ be the first and second order derivatives of $\ell(\theta)$ respectively, where $n_{pq} = p^2 + pq + p(p+1)/2$. To this end, we apply Taylor's expansion and obtain that

$$\sup_{\|u\|=1} \Big\{ \ell(\theta + N^{-1/2}uC) - \ell(\theta) \Big\} = \sup_{\|u\|=1} \Big\{ CN^{-1/2}\dot{\ell}(\theta)^\top u + 2^{-1}C^2N^{-1}u^\top\ddot{\ell}(\theta)u + o_p(1) \Big\}$$

$$\leq C\|N^{-1/2}\dot{\ell}(\theta)\| - 2^{-1}C^2\lambda_{\min}\big\{-N^{-1}\ddot{\ell}(\theta)\big\} + o_p(1). \tag{B.2}$$

We next prove (B.2) is asymptotically negative with probability 1. To this end, we consider $\dot{\ell}(\theta)$ and $\ddot{\ell}(\theta)$ separately.

First recall $\dot{\ell}^d_{j_1j_2}(\theta) = \partial\ell(\theta)/\partial d_{j_1j_2}$, $\dot{\ell}_\beta(\theta) = \partial\ell(\theta)/\partial\beta$, $\dot{\ell}^e_{j_1j_2}(\theta) = \partial\ell(\theta)/\partial\omega^e_{j_1j_2}$ defined in Appendix A.1. One could verify that $\dot{\ell}(\theta) = (\dot{\ell}^{d\top}(\theta), \dot{\ell}^\top_\beta(\theta), \dot{\ell}^{e\top}(\theta))^\top$, where $\dot{\ell}^d(\theta) = \partial\ell(\theta)/\partial\mathcal{D} \in \mathbb{R}^{p^2}$, $\dot{\ell}^e(\theta) = \partial\ell(\theta)/\partial\xi_e \in \mathbb{R}^{p(p+1)/2}$. It can be calculated, for $1 \leq j_1, j_2 \leq p$,

$$\begin{aligned}
N^{-1/2}\dot{\ell}^d_{j_1j_2}(\theta) =& N^{-1/2}\Big[\mathcal{E}^\top(\Omega_e \otimes I_N)(I_{j_2j_1} \otimes W)\mathcal{Y} - \mathrm{tr}\big\{S^{-1}(I_{j_2j_1} \otimes W)\big\}\Big],\\
N^{-1/2}\dot{\ell}_\beta(\theta) =& N^{-1/2}\widetilde{\mathbb{X}}^\top(\Omega_e \otimes I_N)\mathcal{E},\\
N^{-1/2}\dot{\ell}^e_{j_1j_2}(\theta) =& 2^{-1}N^{-1/2}\big\{N\Sigma_{e,j_1j_2} - \mathcal{E}^\top(\mathcal{I}_{j_1j_2} \otimes I_N)\mathcal{E}\big\},
\end{aligned}$$

where $\mathcal{E} = S\mathcal{Y} - \widetilde{\mathbb{X}}\beta$, $\mathcal{I}_{j_1j_2} = I_{j_1j_2} + I_{j_2j_1}$ if $j_1 \neq j_2$, and $\mathcal{I}_{jj} = I_{jj}$. Consequently, we have $E\{N^{-1/2}\dot{\ell}^d_{j_1j_2}(\theta)\} = 0$ due to that $E\{\mathcal{E}^\top(\Omega_e \otimes I_N)(I_{j_2j_1} \otimes W)\mathcal{Y}\} = \mathrm{tr}\{S^{-1}(I_{j_2j_1} \otimes W)\}$, $E\{\dot{\ell}_\beta(\theta)\} = 0$, and $E\{\dot{\ell}^e_{j_1j_2}(\theta)\} = 0$ due to that $E\{\mathcal{E}^\top(I_{j_1j_2} \otimes I_N)\mathcal{E}\} = N\mathrm{tr}(I_{j_1j_2}\Sigma_e) = N\Sigma_{e,j_1j_2}$. It can be verified $\dot{\ell}^d_{j_1j_2}(\theta) = \widetilde{\mathcal{E}}^\top G^d_{j_1j_2}\widetilde{\mathcal{E}} + \widetilde{\mathcal{E}}^\top U^d_{j_1j_2} + c_d$, $\dot{\ell}_\beta(\theta) = L^\top\widetilde{\mathcal{E}}$, $\dot{\ell}^e_{j_1j_2}(\theta) = -\widetilde{\mathcal{E}}^\top G^e_{j_1j_2}\widetilde{\mathcal{E}} + c_e$, where $G^d_{j_1j_2}$, $U^d_{j_1j_2}$, L, and $G^e_{j_1j_2}$ are defined in Part I of Appendix A.1, and c_d and c_e are constants. Therefore, by Lemma 2, it can be easily verified $-\lim N^{-1}\mathrm{cov}(\dot{\ell}(\theta)) \to \Sigma^M_1$, where the details are omitted here. This suggests that the coefficient of the linear term in (B.2) is $O_p(1)$. In addition, we have $N^{-1}\ddot{\ell}(\theta) \to_p \Sigma^M_2$ by Lemma 5 in supplementary material. This indicates $\lambda_{\min}\{-N^{-1}\ddot{\ell}(\theta)\} \to_p \lambda_{\min}(\Sigma^M_2) > 0$ asymptotically. Therefore, the coefficient for C^2 in (B.2) is asymptotically positive. Consequently, by choosing sufficiently large C, (B.2) is negative with probability 1 as $N \to \infty$, thus (B.2) holds.

Step 2. By the first step of proof, we know that $\hat{\theta}_M$ is $\sqrt{N}$-consistent. Therefore, the Taylor's expansion technique can be applied to obtain the following asymptotic approximation

$$\sqrt{N}(\hat{\theta}_M - \theta) = \big\{N^{-1}\ddot{\ell}(\theta^*)\big\}^{-1}\big\{N^{-1/2}\dot{\ell}(\theta)\big\}, \tag{B.3}$$

where θ^* is between θ and $\hat{\theta}_M$. By the proof in the first step, we know that $-N^{-1}\ddot{\ell}(\theta^*) \to_p \Sigma_2^M$.

We next prove that $N^{-1/2}\dot{\ell}(\theta) \to_d N(\mathbf{0}, \Sigma_1^M)$. This suffices to show that for any $\eta = (\eta_d^\top, \eta_\beta^\top, \eta_e^\top)^\top \in \mathbb{R}^{p^2+pq+p(p+1)/2}$, we have $N^{-1/2}\eta^\top\dot{\ell}(\theta) \to_d N(0, \eta^\top\Sigma_1^M\eta)$, where $\eta_d = (\eta_{d1}, \cdots, \eta_{d,p^2})^\top \in \mathbb{R}^{p^2}$, $\eta_\beta = (\eta_{\beta 1}, \cdots, \eta_{\beta,pq})^\top \in \mathbb{R}^{pq}$, $\eta_e = (\eta_{e1}, \cdots, \eta_{e,p(p+1)/2})^\top \in \mathbb{R}^{p(p+1)/2}$. It can be verified that $\eta^\top\dot{\ell}(\theta) = \widetilde{\mathcal{E}}^\top G_\eta \widetilde{\mathcal{E}} + \widetilde{\mathcal{E}}^\top U_{1\eta}(\widetilde{\mathbb{X}}\beta) + \widetilde{\mathcal{E}}^\top U_{2\eta}(\widetilde{\mathbb{X}}\eta_\beta) + c_\eta$, where c_η is a constant, $G_\eta = \sum_{j_1,j_2=1}^p \eta^d_{(j_1-1)p+j_2} G^d_{j_1j_2} + \sum_{j_2\le j_1}^p \eta^e_{(j_1-1)p+j_2} G^e_{j_1j_2}$, $U_{1\eta} = \sum_{j_1,j_2=1}^p \eta^d_{(j_1-1)p+j_2}(\Sigma_e^{1/2\top} \otimes I_N)(I_{j_2j_1} \otimes W)S^{-1}$, $U_{2\eta} = \Omega_e^{1/2} \otimes I_N$. It can be derived $|G^d_{j_1j_2}|_e \preccurlyeq c_g(\mathbf{1}_p\mathbf{1}_p^\top)\otimes\mathbb{W}_1$, $|G^e_{j_1j_2}| \preccurlyeq c_e(\mathbf{1}_p\mathbf{1}_p^\top)\otimes I_N$, $|U_{1\eta}|_e \preccurlyeq c_{1\eta}(\mathbf{1}_p\mathbf{1}_p^\top)\otimes\mathbb{W}_1$, where c_g, c_e, and $c_{1\eta}$ are finite constants. By Lemma 4, it suffices to show $N^{-2}\mathrm{tr}\big\{(\mathbb{G}^{d\top}_{j_1j_2}\mathbb{G}^d_{j_1j_2})^2\big\} \to 0$, $N^{-2}\mathrm{tr}\big\{(\mathbb{G}^{e\top}_{j_1j_2}\mathbb{G}^e_{j_1j_2})^2\big\} \to 0$, and $N^{-1}\lambda^2_{\max}\{|U_{1\eta}|_e^\top|U_{1\eta}|_e\} \to 0$, which can be implied by (A.21) and (A.23) of Lemma 3 respectively. This completes the proof of Theorem 1.

Appendix B.2: Proof of Theorem 4

Following the previous procedure for the proof of QMLE, we study the asymptotic properties in two steps. In the first step, $\hat{\gamma}_L$ is proved to be $\sqrt{N}$-consistent. Secondly, the asymptotic normality of $\hat{\gamma}_L$ is established.

Step 1. Similarly, we follow the technique of Fan and Li (2001) to prove that, for any $\epsilon > 0$, there exists a constant $0 < C < \infty$, such that

$$\lim_{N\to\infty} P\Big\{\inf_{\|u\|=C} Q(\gamma + N^{-1/2}u) > Q(\gamma)\Big\} \ge 1 - \epsilon. \tag{B.4}$$

Therefore, it is implied that with probability at least $1-\epsilon$, there is a local minimizer $\hat{\gamma}_L$ in the ball $\{\gamma + N^{-1/2}uC : \|u\| \leq 1\}$. Let $\dot{Q}(\gamma) = \partial Q(\gamma)/\partial\gamma \in \mathbb{R}^{m_{pq}}$ and $\ddot{Q}(\gamma) = \partial^2 Q(\gamma)/\partial\gamma\partial\gamma^\top \in \mathbb{R}^{m_{pq}\times m_{pq}}$ be the first and second order derivatives of $Q(\gamma)$ respectively, where $m_{pq} = p^2 + pq$. In order to obtain (B.4), we conduct Taylor's expansion as

$$\inf_{\|u\|=C} \left\{ Q(\gamma + N^{-1/2}u) - Q(\gamma) \right\} = CN^{-1/2}\dot{Q}(\gamma)^\top u + 2^{-1}C^2N^{-1}u^\top\ddot{Q}(\gamma)u + o_p(1)$$

$$\geq 2^{-1}C^2\lambda_{\min}\{N^{-1}\ddot{Q}(\gamma)\} - N^{-1/2}\|\dot{Q}(\gamma)\|C + o_p(1). \qquad \text{(B.5)}$$

We next consider the terms $\dot{Q}(\gamma)$ and $\ddot{Q}(\gamma)$ respectively. First, it can be concluded $\lim_{N\to\infty} N^{-1}\text{cov}\{\dot{Q}(\gamma), \dot{Q}(\gamma)\} = \Sigma_1^L$ by Lemma 2. This implies the coefficient for the linear term of C (i.e., $N^{-1/2}\dot{Q}(\gamma)$) is $O_p(1)$. Next, by Lemma 6, we have $N^{-1}\ddot{Q}(\gamma) \to_p \Sigma_2^L$, which indicates $\lambda_{\min}\{N^{-1}\ddot{Q}(\gamma)\} \to_p \lambda_{\min}(\Sigma_2^L) > 0$ asymptotically. Therefore, the coefficient for C^2 in (B.5) is asymptotically positive. Consequently, by choosing sufficiently large C, (B.5) is positive with probability 1 as $N \to \infty$, thus (B.4) holds.

STEP 2. It has been proved $\hat{\gamma}_L$ is $\sqrt{N}$-consistent. Then, we are enabled to apply the technique of Taylor's expansion as

$$\sqrt{N}(\hat{\gamma}_L - \gamma) = \{N^{-1}\ddot{Q}(\gamma^*)\}^{-1}\{N^{-1/2}\dot{Q}(\gamma)\}, \qquad \text{(B.6)}$$

where γ^* is between γ and $\hat{\gamma}_L$. Together with the conclusion $N^{-1}\ddot{Q}(\gamma) \to_p \Sigma_2^L$ proved in Lemma 6, we have $N^{-1}\ddot{Q}(\gamma^*) \to_p \Sigma_2^L$.

We next prove that $N^{-1/2}\dot{Q}(\gamma) \to_d N(\mathbf{0}_{m_{pq}}, \Sigma_1^L)$ as $N \to \infty$. It suffices to show that for any $\eta = (\eta_d^\top, \eta_\beta^\top)^\top \in \mathbb{R}^{p^2+pq}$, we have $N^{-1/2}\eta^\top\dot{Q}(\gamma) \to_d N(0, \eta^\top\Sigma_1^L\eta)$, where

$\eta_d = (\eta_{d1}, \cdots, \eta_{d,p^2})^\top \in \mathbb{R}^{p^2}$ and $\eta_\beta = (\eta_{\beta 1}, \cdots, \eta_{\beta,pq})^\top \in \mathbb{R}^{pq}$. It can be derived

$$\eta^\top \dot{Q}(\gamma) = \widetilde{\mathcal{E}}^\top M_{1\eta} \widetilde{\mathcal{E}} + \widetilde{\mathcal{E}}^\top M_{2\eta}(\widetilde{\mathbb{X}}\beta) + \widetilde{\mathcal{E}}^\top M_3(\widetilde{\mathbb{X}}\eta_\beta),$$

where $M_{1\eta} = 2\sum_{j_1,j_2=1}^{p} \eta_{d,(j_1-1)p+j_2} M_{j_1 j_2}$, $M_{2\eta} = 2\sum_{j_1,j_2=1}^{p} \eta_{d,(j_1-1)p+j_2} M_1^\top M_{3,j_1 j_2}$, $M_3 = -2M_1^\top m \widetilde{S}^\top$. By Lemma 4, it suffices to show

$$N^{-2}\mathrm{tr}(|M_{1\eta}|_e |M_{1\eta}|_e^\top |M_{1\eta}|_e |M_{1\eta}|_e^\top) \to 0,$$
$$N^{-1}\lambda_{\max}^2(|M_{2\eta}|_e |M_{2\eta}|_e^\top) \to 0, \quad N^{-1}\lambda_{\max}^2(|M_3|_e |M_3|_e^\top) \to 0,$$

as $N \to \infty$. Note that the upper bound of each equation is established in (A.30)–(A.33) in Lemma 3. Next, by applying (A.22)–(A.21) of Lemma 3, the desired results can be obtained. This completes the proof.

Acknowledgements

The authors thank the anonymous referees, the Associate Editor and the Co-Editor O. Linton for their very helpful comments and suggestions on the original version. Xuening Zhu[1] is supported by NIDA, NIH grants P50 DA039838, a NSF grant DMS 1512422 and National Nature Science Foundation of China (NSFC, 11690015); Danyang Huang[2*] is supported by National Natural Science Foundation of China (NSFC, 11701560), fund for building world-class universities (disciplines) of Renmin University of China, the Fundamental Research Funds for the Central Universities and the Research Funds of Renmin University of China (Grant No.16XNLF01),the Beijing Municipal Social Science Foundation (Grant No.17GLC051) and the Center for Applied Statistics, School of Statistics, Renmin University of China; Hansheng Wang[4] is partially supported by National Natural Science Foundation of China (No. 71532001, No. 11525101, No. 71332006). It is also supported in part by China's National Key Research Special Program (No. 2016YFC0207700).

References

Anselin, L. (2013), *Spatial econometrics: methods and models*, vol. 4, Springer Science & Business Media.

Baltagi, B. H. and Bresson, G. (2011), "Maximum likelihood estimation and Lagrange multiplier tests for panel seemingly unrelated regressions with spatial lag and spatial errors: An application to hedonic housing prices in Paris," *Journal of Urban Economics*, 69, 24–42.

Baltagi, B. H. and Deng, Y. (2015), "EC3SLS estimator for a simultaneous system of spatial autoregressive equations with random effects," *Econometric Reviews*, 34, 659–694.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014), *Hierarchical modeling and analysis for spatial data*, Crc Press.

Barabási, A.-L. and Albert, R. (1999), "Emergence of scaling in random networks," *science*, 286, 509–512.

Barry, R. P. and Pace, R. K. (1999), "Monte Carlo estimates of the log determinant of large sparse matrices," *Linear Algebra and its applications*, 289, 41–54.

Chen, X., Chen, Y., and Xiao, P. (2013), "The impact of sampling and network topology on the estimation of social intercorrelations," *Journal of Marketing Research*, 50, 95–110.

Clauset, A., Shalizi, C. R., and Newman, M. E. (2009), "Power-law distributions in empirical data," *SIAM review*, 51, 661–703.

Cohen-Cole, E., Liu, X., and Zenou, Y. (2018), "Multivariate choices and identification of social interactions," *Journal of Applied Econometrics*, 33, 165–178.

De Graaff, T., Van Oort, F. G., and Florax, R. J. (2012), "Regional population–employment dynamics across different sectors of the economy," *Journal of Regional Science*, 52, 60–84.

Fan, J. and Li, R. (2001), "Variable selection via nonconcave penalized likelihood and its oracle properties," *Journal of the American statistical Association*, 96, 1348–1360.

Heckman, J. J. (1976), "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models," in *Annals of Economic and Social Measurement, Volume 5, number 4*, NBER, pp. 475–492.

Holland, P. W. and Leinhardt, S. (1981), "An exponential family of probability distributions for directed graphs," *Journal of the American Statistical Association*, 76, 33–50.

Huang, D., Lan, W., Z. H., and Wang, H. (2016), "Least squares estimation of spatial autoregressive models for large-scale social networks," *Working Paper.*

Kelejian, H. H. and Prucha, I. R. (2004), "Estimation of simultaneous systems of spatially interrelated cross sectional equations," *Journal of econometrics*, 118, 27–50.

Lee, L.-f. (2004), "Asymptotic distributions of quasi-Maximum likelihood estimators for spatial autoregressive models," *Econometrica*, 72, 1899–1925.

LeSage, J. P. (2008), "An introduction to spatial econometrics," *Revue d'économie industrielle*, 19–44.

LeSage, J. P. and Pace, R. K. (2007), "A matrix exponential spatial specification," *Journal of Econometrics*, 140, 190–214.

Lin, X. and Lee, L.-f. (2010), "GMM estimation of spatial autoregressive models with unknown heteroskedasticity," *Journal of Econometrics*, 157, 34–52.

Liu, X. (2013), "Estimation of a local-aggregate network model with sampled networks," *Economics Letters*, 118, 243–246.

— (2014), "Identification and efficient estimation of simultaneous equations network models," *Journal of Business & Economic Statistics*, 32, 516–536.

Liu, X., Patacchini, E., and Rainone, E. (2017), "Peer effects in bedtime decisions among adolescents: a social network model with sampled data," *The Econometrics Journal*, 20, S103–S125.

Liu, X. and Saraiva, P. (2017), "GMM estimation of spatial autoregressive models in a system of simultaneous equations with heteroskedasticity," *Econometric Reviews.*

Newman, M., Barabasi, A.-L., and Watts, D. J. (2006), *The structure and dynamics of networks*, Princeton University Press.

Nowicki, K. and Snijders, T. A. B. (2001), "Estimation and prediction for stochastic blockstructures," *Journal of the American Statistical Association*, 96, 1077–1087.

Smirnov, O. and Anselin, L. (2001), "Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach," *Computational Statistics & Data Analysis*, 35, 301–319.

Sojourner, A. (2013), "Identification of peer effects with missing peer data: Evidence from Project STAR," *The Economic Journal*, 123, 574–605.

Wang, Y. J. and Wong, G. Y. (1987), "Stochastic blockmodels for directed graphs," *Journal of the American Statistical Association*, 82, 8–19.

White, H. (1996), *Estimation, inference and specification analysis*, no. 22, Cambridge university press.

Yang, K. and Lee, L.-f. (2017), "Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models," *Journal of Econometrics*, 196, 196–214.

Yang, S. and Allenby, G. M. (2003), "Modeling interdependent consumer preferences," *Journal of Marketing Research*, 40, 282–294.

Zhao, Y., Levina, E., and Zhu, J. (2012), "Consistency of community detection in networks under degree-corrected stochastic block models," *The Annals of Statistics*, 40, 2266–2292.

Zhou, J., Tu, Y., Chen, Y., and Wang, H. (2017), "Estimating spatial autocorrelation with sampled network data," *Journal of Business & Economic Statistics*, 35, 130–138.

Zhu, X., Pan, R., Li, G., Liu, Y., Wang, H., et al. (2017), "Network vector autoregression," *The Annals of Statistics*, 45, 1096–1123.

Table 1: Simulation Results with 500 Replications for the dyad independence model. The RMSE ($\times 10^2$) and the Coverage Probability (%) are reported for QMLE and LSE respectively. The results are displayed as $\mathcal{E}$ follows normal distribution and t-distribution in two examples. The average CPU computational time is also reported.

N	Est.	d_{11}	d_{21}	d_{12}	d_{22}	b_{11}	b_{21}	b_{12}	b_{22}	CPU
				Case 1: Normal Distribution						
200	QMLE	15.8(95.6)	13.7(95.6)	23.5(94.0)	16.0(94.2)	5.3(94.6)	4.8(95.2)	6.4(94.2)	5.8(93.6)	3.01
	LSE	18.5(97.2)	16.5(95.6)	24.6(95.2)	17.6(97.2)	5.5(94.2)	5.1(93.6)	6.5(95.0)	6.1(94.4)	7.30
500	QMLE	11.3(95.6)	8.6(94.0)	17.2(94.0)	9.4(95.2)	3.4(96.0)	3.4(94.4)	4.2(96.0)	4.3(95.0)	13.53
	LSE	12.8(94.2)	9.4(95.4)	19.0(91.6)	10.4(96.8)	3.3(95.6)	3.4(95.6)	3.9(94.6)	4.0(96.2)	8.12
1000	QMLE	7.9(95.0)	6.0(93.8)	11.8(92.8)	6.8(94.4)	2.3(95.2)	2.4(94.2)	2.8(94.4)	2.7(95.8)	54.93
	LSE	8.8(95.4)	6.0(95.6)	12.8(93.2)	6.7(96.2)	2.4(95.0)	2.4(94.6)	2.7(96.6)	2.8(95.8)	7.76
				Case 2: t-distribution						
200	QMLE	18.6(93.0)	17.5(95.2)	29.0(94.6)	17.9(95.4)	6.9(94.2)	6.2(96.2)	8.4(95.8)	7.2(96.6)	3.60
	LSE	22.8(96.0)	22.1(95.8)	33.5(94.8)	21.6(94.8)	6.8(95.2)	6.1(96.2)	8.6(93.8)	7.9(94.2)	7.79
500	QMLE	11.6(94.2)	10.7(93.4)	17.6(94.2)	10.0(96.0)	4.1(95.0)	4.1(94.2)	5.2(94.4)	5.0(94.2)	15.11
	LSE	14.2(94.4)	11.4(95.6)	20.7(94.8)	11.5(95.6)	4.3(95.0)	4.4(94.2)	5.1(95.2)	5.1(96.0)	8.43
1000	QMLE	9.1(95.0)	8.0(93.4)	13.8(96.0)	7.8(94.4)	2.9(95.0)	2.9(95.4)	3.5(95.4)	3.6(95.0)	60.79
	LSE	10.0(95.8)	8.6(93.6)	14.2(94.2)	8.8(93.0)	3.1(93.6)	3.1(94.6)	3.6(96.8)	3.8(95.6)	7.58

Table 2: Simulation Results with 500 Replications the stochastic block model. The RMSE ($\times 10^2$) and the Coverage Probability (%) are reported for QMLE and LSE respectively. The results are displayed as $\mathcal{E}$ follows normal distribution and t-distribution in two examples. The average CPU computational time is also reported.

N	Est.	d_{11}	d_{21}	d_{12}	d_{22}	b_{11}	b_{21}	b_{12}	b_{22}	CPU
Case 1: Normal Distribution										
200	QMLE	4.4(90.6)	3.1(95.4)	5.5(94.8)	3.4(95.0)	5.2(94.0)	5.1(94.2)	6.6(93.2)	6.2(94.4)	2.91
	LSE	4.6(96.6)	5.7(94.0)	6.2(93.4)	5.9(95.0)	9.4(95.6)	14.6(94.6)	11.2(95.2)	16.9(92.6)	7.29
500	QMLE	2.5(94.6)	2.1(96.8)	3.6(94.0)	2.2(96.4)	3.7(94.6)	3.3(93.4)	4.3(93.6)	3.8(95.6)	10.39
	LSE	2.8(96.2)	2.4(96.6)	3.8(94.8)	2.6(94.6)	3.8(94.8)	3.4(95.2)	4.4(94.2)	4.0(94.4)	7.77
1000	QMLE	1.9(93.6)	1.5(94.8)	2.5(96.0)	1.6(95.8)	2.4(94.0)	2.4(95.0)	3.0(94.6)	3.0(93.2)	25.21
	LSE	1.9(95.8)	1.8(94.8)	2.7(93.2)	1.8(93.8)	2.7(93.2)	2.5(95.0)	3.0(94.4)	2.8(95.8)	6.10
Case 2: t-distribution										
200	QMLE	4.5(94.2)	3.8(94.6)	6.6(95.4)	4.1(95.6)	6.8(93.2)	6.7(94.0)	8.1(95.6)	8.2(95.2)	3.17
	LSE	5.0(97.4)	4.6(96.6)	6.7(97.0)	4.4(95.6)	7.4(96.4)	6.9(97.2)	8.7(93.4)	8.5(94.2)	7.10
500	QMLE	2.7(94.8)	2.4(95.0)	4.0(94.8)	2.7(93.4)	4.4(94.2)	4.4(94.2)	5.2(94.8)	5.3(93.8)	10.58
	LSE	3.2(94.2)	3.0(94.8)	4.5(94.8)	2.8(95.0)	4.8(94.2)	4.6(94.6)	5.3(96.2)	5.0(94.8)	7.50
1000	QMLE	2.1(94.2)	1.9(94.2)	2.9(93.6)	1.9(93.2)	3.1(95.2)	2.9(97.2)	3.9(94.4)	3.6(95.6)	25.30
	LSE	2.2(94.8)	2.2(95.0)	3.0(95.0)	2.0(94.2)	3.3(95.4)	3.4(96.0)	3.7(95.2)	3.5(95.8)	6.14

Table 3: Simulation Results with 500 Replications the power-law distribution model. The RMSE ($\times 10^2$) and the Coverage Probability (%) are reported for QMLE and LSE respectively. The results are displayed as $\mathcal{E}$ follows normal distribution and t-distribution in two examples. The average CPU computational time is also reported.

N	Est.	d_{11}	d_{21}	d_{12}	d_{22}	b_{11}	b_{21}	b_{12}	b_{22}	CPU
Case 1: Normal Distribution										
200	QMLE	5.1(92.4)	3.3(95.2)	6.3(95.6)	3.8(96.0)	5.4(93.8)	5.3(96.0)	6.3(95.2)	6.5(94.6)	1.98
	LSE	6.0(94.4)	3.8(96.0)	6.9(95.4)	4.2(96.4)	6.0(93.4)	5.8(95.8)	6.6(95.8)	6.6(95.2)	8.97
500	QMLE	3.2(94.8)	2.3(96.4)	3.9(95.0)	3.0(95.2)	3.2(94.2)	3.3(94.2)	3.7(95.4)	3.9(95.6)	5.17
	LSE	3.6(95.6)	2.7(95.6)	4.1(96.2)	3.1(95.0)	3.5(94.2)	3.5(93.6)	3.8(94.8)	4.0(96.2)	9.90
1000	QMLE	2.4(94.8)	1.5(97.4)	3.1(94.2)	2.0(95.4)	2.2(96.6)	2.2(94.8)	2.8(93.6)	2.8(96.4)	15.58
	LSE	2.7(95.4)	1.7(95.6)	3.3(95.2)	2.1(94.2)	2.3(96.4)	2.3(96.2)	2.8(94.0)	2.9(94.8)	7.65
Case 2: t-distribution										
200	QMLE	5.9(95.8)	4.0(94.2)	7.4(93.4)	5.1(95.2)	6.6(95.0)	6.9(94.6)	8.4(95.8)	8.4(95.8)	2.46
	LSE	7.0(96.0)	5.0(96.0)	8.0(95.8)	5.4(95.8)	7.2(95.2)	7.8(95.4)	8.8(96.2)	8.5(95.6)	9.00
500	QMLE	4.1(96.4)	2.7(95.8)	5.2(95.0)	3.5(95.2)	4.4(94.6)	4.7(94.2)	5.0(94.8)	5.7(94.4)	5.57
	LSE	4.5(95.8)	3.3(95.0)	5.3(95.4)	3.6(94.8)	4.5(94.8)	4.6(94.6)	5.3(94.8)	5.3(95.0)	9.89
1000	QMLE	2.7(95.0)	2.0(95.2)	3.3(95.8)	2.5(95.4)	3.0(94.2)	3.0(93.2)	3.6(94.4)	3.7(95.4)	16.28
	LSE	3.4(94.0)	2.3(95.6)	3.9(94.8)	2.6(94.8)	3.2(95.0)	3.1(95.0)	3.6(95.2)	3.8(96.4)	7.73

Table 4: The sample correlation between $\mathbb{Y}_{j_1}$ and $W\mathbb{Y}_{j_2}$ for $1 \leq j_1, j_2 \leq p$.

	Finance ($\mathbb{Y}_1$)	Economics ($\mathbb{Y}_2$)
Finance ($W\mathbb{Y}_1$)	0.265	0.259
Economics ($W\mathbb{Y}_2$)	0.256	0.260

Table 5: The detailed MSAR analysis results for the Sina Weibo dataset. For each estimation method (i.e, MLE and LSE), the estimates are reported, and "*" denotes that the estimates are significant under the significance level 0.05.

	QMLE Estimation		LSE Estimation	
	Finance	Economics	Finance	Economics
Finance ($W\mathbb{Y}_1$)	0.305 *	0.209 *	0.278 *	0.194 *
Economics ($W\mathbb{Y}_2$)	0.098	0.196 *	0.110	0.196 *
Intercept	-0.291 *	-0.273 *	-0.473 *	-0.455 *
Gender	0.303 *	0.296 *	0.304 *	0.297 *
Beijing	-0.027	-0.043	-0.041	-0.056
Shanghai	-0.018	-0.040	-0.002	-0.027
Tenure	-0.021	-0.014	-0.017	-0.011