SCIENCE CHINA
Mathematics

Progress of Projects Supported by NSFC
• ARTICLES •

November 2022 Vol. 65 No. 11: 2219–2242
https://doi.org/10.1007/s11425-021-1917-4

A case study on the shareholder network effect of stock market data: An SARMA approach

Rong Zhang[1], Jing Zhou[2,*], Wei Lan[3] & Hansheng Wang[4]

[1]*School of Mathematics and Statistics, Yunnan University, Kunming* 650500*, China;*
[2]*Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing* 100872*, China;*
[3]*The Center of Statistical Research, Southwestern University of Finance and Economics, Chengdu* 611130*, China;*
[4]*Guanghua School of Management, Peking University, Beijing* 100871*, China*

Email: lurong1028@mail.ynu.edu.cn, jing.zhou@ruc.edu.cn, lanwei@swufe.edu.cn, hansheng@pku.edu.cn

Received January 21, 2021; accepted September 26, 2021; published online July 8, 2022

Abstract One of the key research problems in financial markets is the investigation of inter-stock dependence. A good understanding in this regard is crucial for portfolio optimization. To this end, various econometric models have been proposed. Most of them assume that the random noise associated with each subject is independent. However, dependence might still exist within this random noise. Ignoring this valuable information might lead to biased estimations and inaccurate predictions. In this article, we study a spatial autoregressive moving average model with exogenous covariates. Spatial dependence from both response and random noise is considered simultaneously. A quasi-maximum likelihood estimator is developed, and the estimated parameters are shown to be consistent and asymptotically normal. We then conduct an extensive analysis of the proposed method by applying it to the Chinese stock market data.

Keywords spatial autoregressive moving average model, shareholder network effect, quasi-maximum likelihood estimator, stock market data

MSC(2020) 62H11, 62P05

Citation: Zhang R, Zhou J, Lan W, et al. A case study on the shareholder network effect of stock market data: An SARMA approach. Sci China Math, 2022, 65: 2219–2242, https://doi.org/10.1007/s11425-021-1917-4

1 Introduction

The stock market is often regarded as the barometer of a country's economic development. This is especially true in the Chinese stock market. The Chinese stock market is a big market with great potential and broad development prospects. After more than twenty years of development, the Chinese stock market reached a scale that took many other countries decades or even hundreds of years to achieve [13,23]. The China Securities Depository and Clearing Corporation Limited reported that by 2020, there were 2,423 companies listed on the Shenzhen Stock Exchange with a market value of 35.2 trillion *yuan*, and 1,800 companies listed on the Shanghai Stock Exchange with a market value of 45.5 trillion *yuan*. Additionally, the total circulation market value of the Shanghai and Shenzhen Stock Exchange reached 64.81 trillion *yuan* that same year. The Chinese stock market plays an important role in enterprise

* Corresponding author

financing, improvement of the financing structure, optimal allocation of social resources, and promoting social and economic growth.

For financial markets, a key research problem is investigating inter-stock dependence. A good understanding of this dependence is crucial for portfolio optimization [25]. To this end, various econometric models have been proposed. For example, Sharpe [28] developed the popularly used capital asset pricing model (CAPM). This model quantified the return of a single stock or stock portfolio by the risk-free return, risk correction coefficient β and equity market premium. As an extension of the CAPM, Fama and French [11] developed the famous three-factor model. They found that the market value, book-to-market ratio and price-to-earnings ratio were also important factors in explaining comovement in the returns of different stocks. These groundbreaking theories have inspired many follow-up studies trying to explain inter-stock dependence from various perspectives. For example, Chan et al. [4] found that the cash flow, stock size and book-to-market ratio can help explain stock returns synchronicity. They also demonstrated that the same industry effect was one of the sources of inter-stock dependence [5]. Other factors such as the leverage, size, monetary policy and information sharing are also important in explaining covariation in stock returns [10, 16, 18, 27].

It is remarkable that none of the aforementioned models have studied inter-stock dependence from the perspective of the shareholder network. A shareholder network refers to a network in which each node represents a stock and each edge represents a major common shareholder relationship between the two stocks. An accurate definition of a common shareholder is provided in the next paragraph. We argue that this is an important relationship that might be partially responsible for inter-stock dependence. In fact, various spatial autoregressive models have been successfully used to capture such a dependence effect [33, 35, 36]. Despite the usefulness of those pioneering studies, all the models suffer one common limitation. They all assume that the random noise associated with different stocks is independent. This assumption implies that the dependence structure observed in the responses can be fully captured by two components: the spatial autoregressive component and the regression component associated with exogenous covariates. Unfortunately, in reality, this assumption could be questionable. In fact, much empirical evidence suggests that nontrivial dependence might still exist in random noises. Ignoring this crucial insight might lead to biased estimations and thus inaccurate predictions.

To fix the problem, we propose a spatial autoregressive moving average (SARMA) model [1, 8, 14, 15]. It accounts for the shareholder network effect of multiple stocks. This model could be viewed as a natural extension of the spatial autoregressive model but with an additional spatial moving average (SMA) component. By including the SMA component, the previously ignored dependence in the random noise can be captured and investigated. To this end, we identify the top five or top ten shareholders for each stock. Two stocks are said to have a common shareholder relationship if there is some overlap in the top shareholders between the two stocks. Thereafter, an adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{N\times N}$ can be constructed to describe the common shareholder network. Specifically, for any two arbitrary stocks i and j, $a_{ij} = 1$ if they share at least one common shareholder; otherwise $a_{ij} = 0$. As a result, we can use the adjacency matrix A to capture the shareholder effect of different stocks. To estimate the parameters of the SARMA model, we employ a quasi-maximum likelihood estimation method. By temporarily assuming that the error term follows an independent Gaussian distribution, a quasi-log-likelihood function can be established. Its first- and second-order derivatives can then be analytically calculated. A standard Newton-Raphson type computational algorithm is then used to optimize the quasi-log-likelihood function. This leads to the quasi-maximum likelihood estimator (QMLE). Theoretically, we can prove that the QMLE is consistent and asymptotically normal under appropriate conditions, even if the random noise is not Gaussian at all. To demonstrate its practical use, we apply the proposed method to the Chinese stock market data and extensively analyze the results.

The rest of this article is organized as follows. Section 2 introduces the stock market data and network dependence, which includes the description of the data, descriptive analysis, explanation of the SARMA approach, statistical analysis and real data results. Section 3 discusses the asymptotic theory, which contains the description of the QMLE. In this section, the statistical hypotheses are also tested, and the results of the simulation and hypothesis testing are provided. A short discussion and some concluding remarks are presented in Section 4. All the technical details are relegated to the appendix.

2 Stock market data and network dependence

2.1 Data description

We present here a case study on the shareholder network effect in the Chinese stock market, where the data were collected from the China Stock Market and Accounting Research (CSMAR) Database. Specifically, a total of 509 stocks in the Shanghai A stock market, 591 stocks in the Shenzhen A stock market and 184 stocks in the growth enterprise market (GEM) were analyzed. For each stock market, the response variable Y_i for $1 \leqslant i \leqslant N$ is the excessive return (in percentage) of each stock. Here, the excessive return is defined as the daily stock return subtracted by the return of the market (e.g., as represented by the corresponding market composite index). In empirical finance, some practical evidence reveals that stocks from different industries may exhibit different performances [4]. This inspires us to collect a covariate of

$$X_i = (X_{i1}, X_{i2}, \ldots, X_{ip}) \in \mathbb{R}^p$$

for each Y_i, where $X_{ik} = 1$ if stock i belongs to the k-th industry, and $X_{ik} = 0$ otherwise. In this proposed example, we consider $p = 10$ different industries. They are, respectively, agriculture, construction, culture and sports, electricity, finance, information, manufacturing, mining, real estate and wholesale and retail. Furthermore, to investigate the inter-stock dependence, we use the adjacency matrix A described in Section 1.

2.2 Descriptive analysis

To provide a quick understanding of the three markets, various network statistics are computed. We start by calculating the in-degree and out-degree distributions. For illustration purposes, we take the Shanghai A stock market as an example. The results are shown in Figures 1–2.

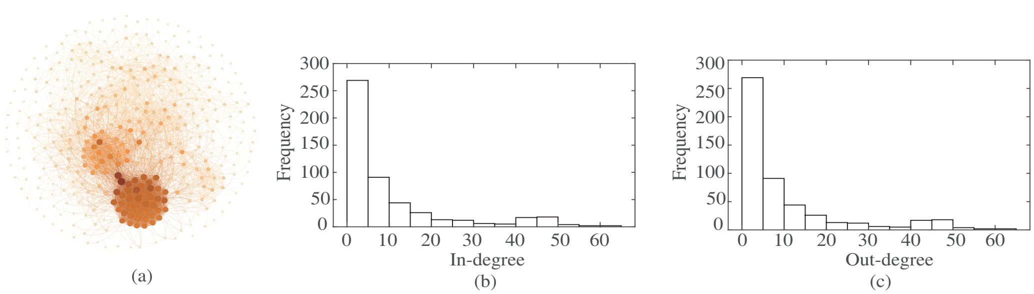


Figure 1 (Color online) Shanghai A stock market: inter-stock dependence constructed using the top five shareholders. (a) Visualization of the network structure. (b) Histogram of the in-degree distribution. (c) Histogram of the out-degree distribution

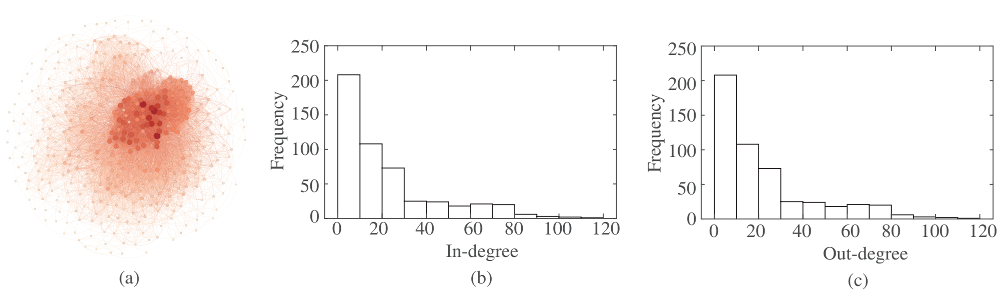


Figure 2 (Color online) Shanghai A stock market: inter-stock dependence constructed using the top ten shareholders. (a) Visualization of the network structure. (b) Histogram of the in-degree distribution. (c) Histogram of the out-degree distribution

The figures illustrate how the network structure can differ when different common shareholder information is used in the construction (e.g., top five or top ten). There are two small dense networks in Figure 1, while there is only one small network in Figure 2. We further investigate the dense clusters in Figure 1 and find that most of the nodes are large state-owned enterprises, such as the Industrial and Commercial Bank of China, Agricultural Bank of China, Bank of China and Sinopec. These enterprises, respectively, account for 6.25%, 4.84%, 3.39% and 2.7% of the total market value. Another interesting finding is that the degrees of those companies are very large. This suggests that they are correlated with each other in terms of common shareholder information. This, to some extent, validates our conjecture regarding inter-stock dependence. In addition to this, there also exists an industry effect among different stocks. To further investigate the industry effect, we visualize the results of three representative industries—manufacturing, mining and real estate—in Figures 3–5. To better display their patterns, the vertical coordinates of the in-degree and out-degree distributions are marked with logarithms. Clearly, the network structures of the three industries differ from one another. For example, the network structure of the manufacturing industry is denser than the other two industries.

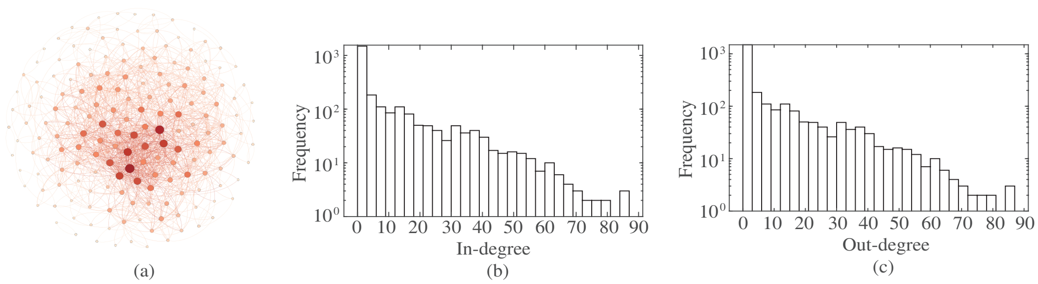


Figure 3 (Color online) Manufacturing: inter-stock dependence constructed using the top five shareholders. (a) Visualization of the network structure. (b) Histogram of the in-degree distribution. (c) Histogram of the out-degree distribution

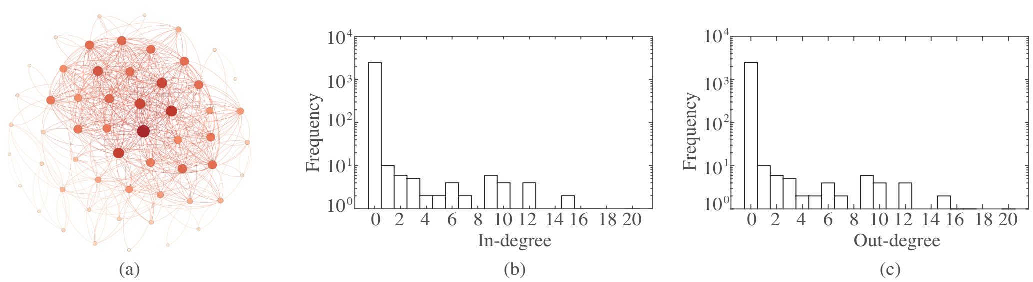


Figure 4 (Color online) Mining: inter-stock dependence constructed using the top five shareholders. (a) Visualization of the network structure. (b) Histogram of the in-degree distribution. (c) Histogram of the out-degree distribution

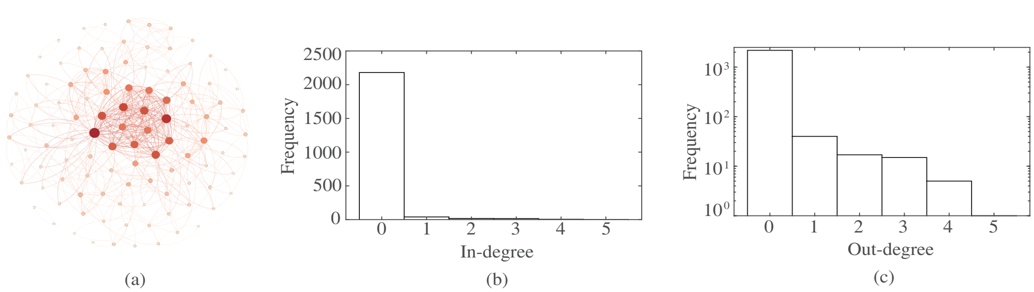


Figure 5 (Color online) Real estate: inter-stock dependence constructed using the top five shareholders. (a) Visualization of the network structure. (b) Histogram of the in-degree distribution. (c) Histogram of the out-degree distribution

2.3 The SARMA approach

To model the shareholder network effect of different stocks, we adopt the SARMA approach. It should be noted that in terms of the use of the SARMA model, there is a fundamental difference between spatial data and network data in the construction of adjacency matrices. For spatial data, the adjacency matrix is usually constructed by using spatial relationships (e.g., spatially close or neighboring). However, for network data, the adjacency matrix is constructed based on network relationships. Once the adjacency matrix is constructed, it seems that little difference exists between the two types of data. In this particular case, suppose that we have a large common shareholder network with N stocks. Its structure is then captured by the adjacency matrix A described in Section 1. Furthermore, we define $a_{ii} = 0$ and $\sum_{i=1}^N a_{ij} > 0$ for completeness.

For each stock i, we observe an excessive return Y_i and a p-dimensional exogenous covariate (i.e., the industry) $X_i = (X_{i1}, X_{i2}, \ldots, X_{ip})^{\mathrm{T}} \in \mathbb{R}^p$. To model the inter-stock relationship between Y_i's via the shareholder network, we adopt the SARMA model with X-variables (SARMAX) as follows [1]:

$$\mathbb{Y} = \rho W \mathbb{Y} + \mathbb{X}\beta + \gamma W \mathcal{E} + \mathcal{E}, \tag{2.1}$$

where $\mathbb{Y} = (Y_1, \ldots, Y_N)^{\mathrm{T}} \in \mathbb{R}^N$ is the response vector of excessive returns, $\mathbb{X} = (X_1, \ldots, X_N)^{\mathrm{T}} \in \mathbb{R}^{N\times p}$ is the design matrix of the industry related covariates, $W = (w_{ij}) \in \mathbb{R}^{N\times N}$ is a row normalized adjacency matrix with $w_{ij} = a_{ij}/\sum_{j=1}^N a_{ij}$, ρ and γ are the shareholder network effect coefficients, and β is a $p \times 1$ vector of the regression coefficients. Additionally, $\mathcal{E} = (\varepsilon_1, \ldots, \varepsilon_N)^{\mathrm{T}} \in \mathbb{R}^N$ is the random noise vector, which is assumed to follow a multivariate distribution with a mean of 0 and a covariance matrix of $\sigma^2 I \in \mathbb{R}^{N\times N}$. Here, I stands for an identity matrix with compatible dimensions. It should be noted that the disturbance process in the model (2.1) is a special case of the proposed model in [2]. According to the model (2.1), we know that

$$\mathbb{Y} = (I - \rho W)^{-1}\{\mathbb{X}\beta + (I + \gamma W)\mathcal{E}\}.$$

Accordingly, we have $\mathrm{E}(\mathbb{Y} \mid \mathbb{X}) = (I - \rho W)^{-1}\mathbb{X}\beta$ and

$$\mathrm{cov}(\mathbb{Y} \mid \mathbb{X}) = \Sigma = \sigma^2 (I - \rho W)^{-1}(I + \gamma W)(I + \gamma W^{\mathrm{T}})(I - \rho W^{\mathrm{T}})^{-1}.$$

We assume that $|\rho| < 1$ and $|\gamma| < 1$ throughout the entire article since $I - \rho W$ and $I + \gamma W$ are invertible as long as $|\rho| < 1$ and $|\gamma| < 1$, as noticed by [21]. We next consider the problem of the parameter estimation and hypothesis testing.

2.4 Statistical analysis

To estimate the shareholder network effect involved in the model (2.1), a quasi-maximum likelihood estimation method is provided. Note that $\mathrm{E}(\mathbb{Y} \mid \mathbb{X}) = (I - \rho W)^{-1}\mathbb{X}\beta$ and

$$\mathrm{cov}(\mathbb{Y} \mid \mathbb{X}) = \Sigma = \sigma^2 (I - \rho W)^{-1}(I + \gamma W)(I + \gamma W^{\mathrm{T}})(I - \rho W^{\mathrm{T}})^{-1},$$

and the normal log-likelihood function of the model (2.1) is

$$\begin{aligned}\ell(\theta) = &-\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \log|I + \gamma W| + \log|I - \rho W| \\ &- \frac{1}{2\sigma^2}\{(I - \rho W)\mathbb{Y} - \mathbb{X}\beta\}^{\mathrm{T}}(I + \gamma W^{\mathrm{T}})^{-1}(I + \gamma W)^{-1}\{(I - \rho W)\mathbb{Y} - \mathbb{X}\beta\},\end{aligned}$$

where $\theta = (\rho, \gamma, \beta^{\mathrm{T}}, \sigma^2)^{\mathrm{T}} \in \mathbb{R}^{p+3}$. We define $K(\rho) = I - \rho W$, $S(\gamma) = I + \gamma W$ and $G(\rho, \beta) = (I - \rho W)\mathbb{Y} - \mathbb{X}\beta$. Accordingly, the above log-likelihood can be rewritten as (omitting some constants)

$$\ell(\theta) = -\frac{N}{2}\log(\sigma^2) - \log|K(\rho)| + \log|S(\gamma)| - \frac{1}{2\sigma^2}G(\rho, \beta)^{\mathrm{T}}(S(\gamma)^{\mathrm{T}})^{-1}S(\gamma)^{-1}G(\rho, \beta). \tag{2.2}$$

Subsequently, the QMLE $\hat{\theta}$ can be obtained as $\hat{\theta} = \arg\max_{\theta} \ell(\theta)$. For ease of notation, we use K, S and G to represent $K(\rho)$, $S(\gamma)$ and $G(\rho, \beta)$, respectively, throughout the entire article. To ensure a good statistical inference, we further consider the problem of the hypothesis testing.

For the hypothesis testing, we propose three interesting tests, including the testing for global significance, the testing for the shareholder network effect and the testing for the industry effect. First, we write $\widetilde{\theta} = (\rho, \gamma, \beta^{\mathrm{T}})^{\mathrm{T}}$; thus, the test for global significance is

$$H_0 : \widetilde{\theta} = 0 \quad \text{vs.} \quad H_1 : \widetilde{\theta} \neq 0. \tag{2.3}$$

A quasi-likelihood ratio test is proposed to compare the associated likelihood with and without constraint $\widetilde{\theta} = 0$. When there are no constraints, the QMLE is $\hat{\theta} = (\hat{\rho}, \hat{\gamma}, \hat{\beta}^{\mathrm{T}}, \hat{\sigma}^2)^{\mathrm{T}}$. On the contrary, under the null hypothesis of $\widetilde{\theta} = 0$, we can obtain a constrained QMLE of $\hat{\theta}_g = \arg\max_{\theta:\widetilde{\theta}=0} \ell(\theta)$. The quasi-likelihood ratio test statistic is then defined as $T_g = 2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_g)\}$. We next test for the shareholder network effect using the following hypothesis test:

$$H_0 : \rho = \gamma = 0 \quad \text{vs.} \quad H_1 : \rho \neq 0 \text{ or } \gamma \neq 0. \tag{2.4}$$

Similarly, let $\hat{\theta}_s = \arg\max_{\theta:\rho=\gamma=0} \ell(\theta)$ be the QMLE obtained under constraint $\rho = \gamma = 0$. The likelihood ratio test statistic is then defined as $T_s = 2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_s)\}$. Lastly, we consider the test for the industry effect. The interested statistical hypothesis test is given as follows:

$$H_0 : \beta = 0 \quad \text{vs.} \quad H_1 : \beta \neq 0. \tag{2.5}$$

Let $\hat{\theta}_\beta = \arg\max_{\theta:\beta=0} \ell(\theta)$ be the QMLE obtained under constraint $\beta = 0$. The quasi-likelihood ratio test statistic can then be defined as $T_\beta = 2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_\beta)\}$. Having developed elegant estimation and hypothesis testing methods, we next provide a thorough empirical analysis of the Chinese stock market data.

2.5 Stock market data analysis results

In this subsection, we employ the proposed SARMA model to analyze the stock market data introduced in Subsection 2.1. The estimated results of ρ, γ and σ^2 are given in Table 1, and the results of β are displayed by using barplots in Figures 6–8. In Table 1, network 1 stands for the adjacency matrix constructed using information on the top five common shareholders, while network 2 stands for the adjacency matrix constructed using information from the top ten common shareholders. From Table 1, we can see that all the estimators are statistically significant at the 1% or 5% level. Taking the Shanghai A stock market in network 1 as an example, the estimated $\rho = 0.2127$ ($p < 0.001$) suggests that the return of a stock is positively related with the performance of its connected neighbors. We can find that if two stocks have a common shareholder, their stock returns will be highly correlated. The estimated $\gamma = -0.2352$ ($p < 0.05$) confirms that the return of a stock is negatively related with the error from its connected neighbors. Finally, from Figures 6–8, we can see that the estimated β's are very different in the three markets. For example, the coefficients of some industries, such as finance, information and manufacturing are significantly large in the Shenzhen A stock market, but not significant in other markets (e.g., GEM). This suggests an industry effect among the returns of different stocks.

Furthermore, to test the goodness-of-fit of the SARMA model, we propose here an interesting graphical strategy called a residual-residual (RR) plot. The key assumption of the SARMA model is that the observed spatial dependence in $\mathbb{Y}$ can be fully explained by the SARMA model structure. If this assumption is indeed correct, we should expect ε_i's to be mutually independent, regardless of whether they are connected. Inspired by this idea, we randomly pick one stock i and one of its connected neighbors j. We then use estimated residual $\hat{\varepsilon}_i$ as the X-coordinate and estimated $\hat{\varepsilon}_j$ as the Y-coordinate. By doing so, a total of 100 paired residuals $(\hat{\varepsilon}_i, \hat{\varepsilon}_j)$ are plotted in Figure 9. From the figure, we can see a random pattern of scatter points. This indicates that the SARMA model provides a good fit to the data.

Table 1 Estimation result for the stock market data

		Network 1		Network 2	
Market	Parameter	Estimate	p-value	Estimate	p-value
Shanghai	ρ	0.2127	<0.0010	0.3034	<0.0010
	γ	-0.2352	0.0141	-0.3075	0.0015
	σ^2	0.0004	<0.0010	0.0004	<0.0010
Shenzhen	ρ	0.7086	<0.0010	0.6908	0.0013
	γ	-0.6579	0.0004	-0.7312	0.0010
	σ^2	0.0003	<0.0010	0.0002	<0.0010
GEM	ρ	0.6467	0.0031	0.8056	0.0037
	γ	-0.8485	<0.0010	-0.7951	0.0082
	σ^2	0.0004	<0.0010	0.0005	<0.0010

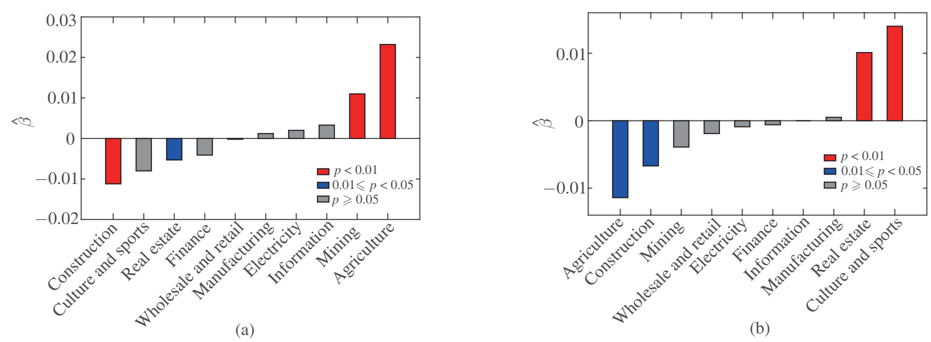


Figure 6 (Color online) The estimation results of β in the Shanghai A stock market for network 1 (a) and network 2 (b)

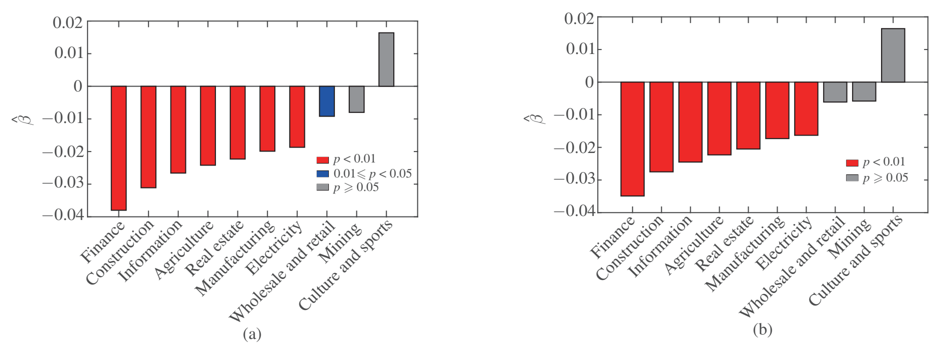


Figure 7 (Color online) The estimation results of β in the Shenzhen A stock market for network 1 (a) and network 2 (b)

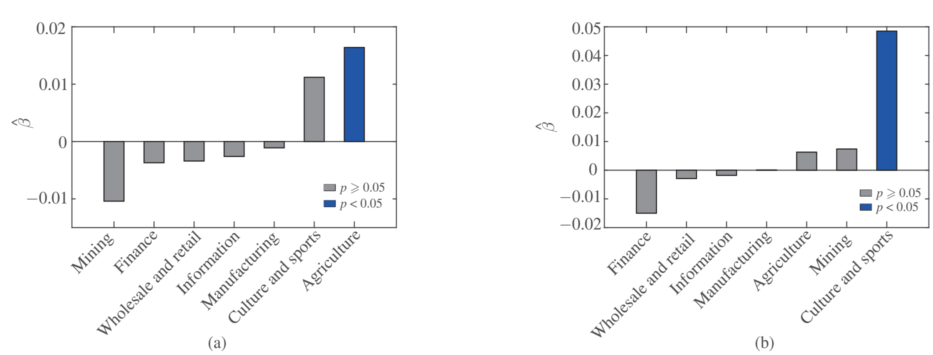


Figure 8 (Color online) The estimation results of β in the GEM for network 1 (a) and network 2 (b)

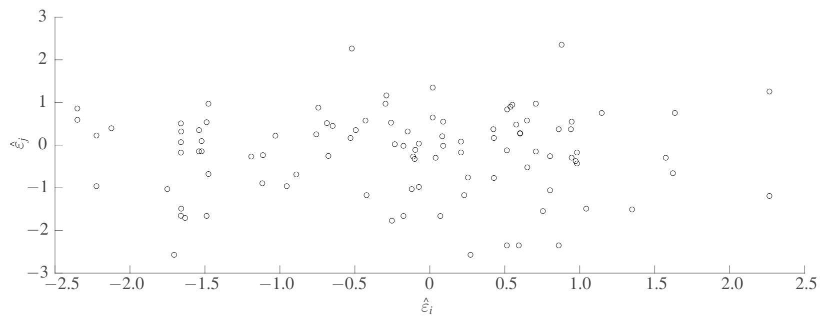


Figure 9 The RR plot

3 The asymptotic theory

3.1 The quasi-maximum likelihood estimator

To corroborate the empirical findings obtained in the previous section, we present here a relatively complicated asymptotic theory. We start by introducing a set of regularity conditions. Before stating the conditions, we first define that $\Delta_N(\theta) = -N^{-1}\mathrm{E}\{\partial^2\ell(\theta)/\partial\theta\partial\theta^{\mathrm{T}}\}$. Here,

$$\begin{aligned}\Delta_N(\theta) = [&\Delta_{N,11}, \Delta_{N,12}, \Delta_{N,13}^{\mathrm{T}}, \Delta_{N,14}; \Delta_{N,21}, \Delta_{N,22}, \Delta_{N,23}^{\mathrm{T}}, \Delta_{N,24}; \Delta_{N,31}, \Delta_{N,32}, \Delta_{N,33}, \Delta_{N,34};\\ &\Delta_{N,41}, \Delta_{N,42}, \Delta_{N,43}^{\mathrm{T}}, \Delta_{N,44}] \in \mathbb{R}^{(p+3)\times(p+3)}\end{aligned}$$

with

$$\begin{aligned}
\Delta_{N,11} &= N^{-1}\mathrm{tr}(WK^{-1}K^{-1}W) + N^{-1}\mathrm{tr}\{W^{\mathrm{T}}(K^{\mathrm{T}})^{-1}K^{-1}W\}\\
&\quad + N^{-1}\sigma^{-2}\mathrm{tr}\{(\mathbb{X}\beta)^{\mathrm{T}}(K^{\mathrm{T}})^{-1}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}WK^{-1}\mathbb{X}\beta\},\\
\Delta_{N,22} &= N^{-1}\mathrm{tr}(WS^{-1}S^{-1}W) + N^{-1}\mathrm{tr}\{W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}W\},\\
\Delta_{N,33} &= \sigma^{-2}N^{-1}\mathbb{X}^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}\mathbb{X}, \quad \Delta_{N,44} = 2^{-1}\sigma^{-4},\\
\Delta_{N,12} &= \Delta_{N,21} = N^{-1}\mathrm{tr}(WK^{-1}S^{-1}W) + N^{-1}\mathrm{tr}\{W^{\mathrm{T}}(K^{\mathrm{T}})^{-1}S^{-1}W\},\\
\Delta_{N,13} &= \Delta_{N,31} = \sigma^{-2}N^{-1}\mathbb{X}^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}WK^{-1}\mathbb{X}\beta,\\
\Delta_{N,14} &= \Delta_{N,41} = \sigma^{-2}N^{-1}\mathrm{tr}(K^{-1}W), \quad \Delta_{N,23} = \Delta_{N,32} = 0_p,\\
\Delta_{N,24} &= \Delta_{N,42} = \sigma^{-2}N^{-1}\mathrm{tr}(S^{-1}W), \quad \Delta_{N,34} = \Delta_{N,43} = 0_p.
\end{aligned}$$

We then define the matrix

$$\begin{aligned}\mathcal{J}_N(\theta, \mu^{(3)}, \mu^{(4)}) = [&\mathcal{J}_{N,11}, \mathcal{J}_{N,12}, \mathcal{J}_{N,13}^{\mathrm{T}}, \mathcal{J}_{N,14}; \mathcal{J}_{N,21}, \mathcal{J}_{N,22}, \mathcal{J}_{N,23}^{\mathrm{T}}, \mathcal{J}_{N,24}; \mathcal{J}_{N,31}, \mathcal{J}_{N,32}, \mathcal{J}_{N,33}, \mathcal{J}_{N,34};\\ &\mathcal{J}_{N,41}, \mathcal{J}_{N,42}, \mathcal{J}_{N,43}^{\mathrm{T}}, \mathcal{J}_{N,44}] \in \mathbb{R}^{(p+3)\times(p+3)},\end{aligned}$$

and we have

$$\begin{aligned}
\mathcal{J}_{N,11} &= 2\sigma^{-4}N^{-1}\mathrm{tr}\{(S^{-1}WK^{-1}\mathbb{X}\beta l_N^{\mathrm{T}}) \circ (WK^{-1})\}\mu^{(3)}\\
&\quad + \sigma^{-4}N^{-1}\mathrm{tr}\{(K^{-1}W) \circ (K^{-1}W)\}(\mu^{(4)} - 3\sigma^4),\\
\mathcal{J}_{N,22} &= \sigma^{-4}N^{-1}\mathrm{tr}\{(S^{-1}W) \circ (S^{-1}W)\}(\mu^{(4)} - 3\sigma^4),\\
\mathcal{J}_{N,33} &= 0, \quad \mathcal{J}_{N,44} = 4^{-1}\sigma^{-8}N^{-1}(\mu^{(4)} - 3\sigma^4),\\
\mathcal{J}_{N,12} &= \mathcal{J}_{N,21} = \sigma^{-4}N^{-1}\mathrm{tr}\{(S^{-1}WK^{-1}\mathbb{X}\beta l_N^{\mathrm{T}}) \circ (S^{-1}W)^{-1}\}\mu^{(3)}\\
&\quad + \sigma^{-4}N^{-1}\mathrm{tr}\{(K^{-1}W) \circ (S^{-1}W)\}(\mu^{(4)} - 3\sigma^4),\\
\mathcal{J}_{N,13} &= \mathcal{J}_{N,31} = \sigma^{-4}N^{-1}\mathbb{X}^{\mathrm{T}}(S^{\mathrm{T}})^{-1}WK^{-1}\mu^{(3)},\\
\mathcal{J}_{N,23} &= \mathcal{J}_{N,32} = \sigma^{-4}N^{-1}\mathrm{tr}\{(S^{-1}W) \circ (S^{-1}\mathbb{X}l_N^{\mathrm{T}})\}\mu^{(3)},\\
\mathcal{J}_{N,14} &= \mathcal{J}_{N,41} = 2^{-1}\sigma^{-4}N^{-1}\mathbb{X}\beta K^{-1}WS^{-1}l_N\mu^{(3)} + 2^{-1}\sigma^{-6}N^{-1}\mathrm{tr}(K^{-1}W)(\mu^{(4)} - 3\sigma^4),
\end{aligned}$$

$$\mathcal{J}_{N,24} = \mathcal{J}_{N,42} = 2^{-1}\sigma^{-6}N^{-1}\mathrm{tr}(S^{-1}W)(\mu^{(4)} - 3\sigma^4),$$
$$\mathcal{J}_{N,34} = \mathcal{J}_{N,43} = \sigma^{-6}N^{-1}l_N S^{-1}X\mu^{(3)},$$

where $\circ$ is the Hadamard product of matrices, $\mu^{(3)} = \mathrm{E}(\varepsilon_i^3)$, $\mu^{(4)} = \mathrm{E}(\varepsilon_i^4)$ and $l_N = (1, \dots, 1)^{\mathrm{T}} \in \mathbb{R}^N$. For the ease of notation, we use Δ_N and $\mathcal{J}_N$ to represent $\Delta_N(\theta)$ and $\mathcal{J}_N(\theta, \mu^{(3)}, \mu^{(4)})$, respectively, in the rest of this article. We next assume the following conditions.

(C1) (Random error) Assume that ε_i's are independent and identically distributed random variables with a mean of 0 and a variance of σ^2. Furthermore, assume $\mathrm{E}|\varepsilon_i|^{4+\lambda} < \infty$ for some positive constant $\lambda > 0$.

(C2) (Weight matrix) Assume that K is nonsingular and there exists a finite positive constant C such that $\|K^{-1}\|_\infty + \|W\|_\infty < C$ for all sufficiently large N, where $\|V\|_\infty$ is defined as $\|V\|_\infty = \max_{1 \leqslant i \leqslant N} \sum_{j=1}^N |v_{ij}|$ for any generic matrix $V = (v_{ij}) \in \mathbb{R}^{N \times N}$.

(C3) (Law of large numbers) Assume that the elements of $\mathbb{X}$ are uniformly bounded constants for all N. In addition, there exists a positive definite matrix $\Sigma \in \mathbb{R}^{p \times p}$ such that $\mathbb{X}^{\mathrm{T}}\mathbb{X}/N \to_p \Sigma$.

(C4) (Hessian matrix) Assume that there exist a positive definite matrix $\Delta \in \mathbb{R}^{(p+3)\times(p+3)}$ and $\mathcal{J} \in \mathbb{R}^{(p+3)\times(p+3)}$ such that $\Delta_N \to_p \Delta$ and $\mathcal{J}_N \to_p \mathcal{J}$ as $N \to \infty$.

These conditions are all mild and commonly used in the literature. Condition (C1) is a moment condition, which is much weaker than commonly used distribution assumptions (see, for example, the normal assumption in [24, 31]). Conditions (C2) and (C3) are standard regularity conditions used in the spatial literature [21, 30]. Condition (C4) is a law of the large number-type assumption, and it is used to show the asymptotic normality of the QMLE. A similar condition can be found in [17]. Based on the above conditions, we then have the following theorem.

Theorem 3.1. *Under Conditions* (C1)–(C4), *there exists a local optimizer* $\hat{\theta}$ *such that* $\|\hat{\theta} - \theta\| = O_p(N^{-1/2})$. *Furthermore, assume* $N \to \infty$. *Then we have* $\sqrt{N}(\hat{\theta} - \theta) \to_d N(0, \Delta^{-1} + \Delta^{-1}\mathcal{J}\Delta^{-1})$, *where* Δ *and* $\mathcal{J}$ *are defined in Condition* (C4).

By Theorem 3.1, the consistency and asymptotic normality of $\hat{\theta}$ are established by using proper conditions, and the proof of this theorem is given in Appendix B. For an asymptotically valid statistical inference, $\Delta^{-1} + \Delta^{-1}\mathcal{J}\Delta^{-1}$ needs to be estimated, which can be consistently estimated by $\Delta_N^{-1}(\hat{\theta}) + \Delta_N^{-1}(\hat{\theta})\mathcal{J}_N(\hat{\theta}, \hat{\mu}^{(3)}, \hat{\mu}^{(4)})\Delta_N^{-1}(\hat{\theta})$, where $\hat{\mu}^{(s)} = N^{-1}\sum_{i=1}^N \hat{\varepsilon}_i^s$ for $s = 3, 4$ and

$$(\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_N)^{\mathrm{T}} = \varepsilon(\hat{\rho}, \hat{\gamma}, \hat{\beta}) = S^{-1}(\hat{\gamma})G(\hat{\rho}, \hat{\beta}).$$

3.2 Testing statistical hypotheses

To demonstrate the effectiveness of the three proposed quasi-likelihood ratio tests, we investigate their asymptotic properties. First, we provide the theoretical properties under the null distributions given in Subsection 2.4.

Theorem 3.2. *Under Conditions* (C1)–(C4), *we consider the null hypotheses of* $H_0^g : \widetilde{\theta} = 0$, $H_0^s : \rho = \gamma = 0$ *and* $H_0^b : \beta = 0$, *respectively. Their quasi-likelihood ratio test statistics satisfy*

$$T_g \to_d \sum_{i=1}^{p+3} \lambda_{i,g}(\theta, \mu^{(3)}, \mu^{(4)})\chi^2_{i,1}, \quad T_s \to_d \sum_{i=1}^{p+3} \lambda_{i,s}(\theta, \mu^{(3)}, \mu^{(4)})\chi^2_{i,1}, \quad T_b \to_d \sum_{i=1}^{p+3} \lambda_{i,b}(\theta, \mu^{(3)}, \mu^{(4)})\chi^2_{i,1}$$

as $N \to \infty$, *where* $\lambda_{i,g}$, $\lambda_{i,s}$ *and* $\lambda_{i,b}$ *are the i-th largest eigenvalues of the matrices*

$$(\Delta + \mathcal{J})^{-1/2}\{\Delta^{-1} - \Delta_g^{-1}\}(\Delta + \mathcal{J})^{-1/2}, \quad (\Delta + \mathcal{J})^{-1/2}\{\Delta^{-1} - \Delta_s^{-1}\}(\Delta + \mathcal{J})^{-1/2}$$

and

$$(\Delta + \mathcal{J})^{-1/2}\{\Delta^{-1} - \Delta_b^{-1}\}(\Delta + \mathcal{J})^{-1/2}.$$

$\chi^2_{i,1}$ *represents a chi-squared distribution with degree of freedom* 1 *for* $\ell = 1, \dots, p+3$ *and* Δ_g, Δ_s *and* Δ_b *are given in Appendix* C. *Furthermore, when* $\mathcal{E}$ *is normally distributed, we have* $T_g \to_d \chi^2_{p+2}$, $T_s \to_d \chi^2_2$ *and* $T_b \to_d \chi^2_p$, *respectively.*

In practice, $\lambda_{i,g}(\theta,\mu^{(3)},\mu^{(4)})$ is unknown and can be estimated by $\lambda_{N,i,g}(\hat{\theta},\hat{\mu}^{(3)},\hat{\mu}^{(4)})$, where $\lambda_{N,i,g}(\hat{\theta},\hat{\mu}^{(3)},\hat{\mu}^{(4)})$ is the i-th largest eigenvalue of the $(p+3)\times(p+3)$ matrix

$$\{\Delta_N(\hat{\theta})+\mathcal{J}_N(\hat{\theta},\hat{\mu}^{(3)},\hat{\mu}^{(4)})\}^{-1/2}\{\Delta_N^{-1}(\hat{\theta})-\Delta_g^{-1}(\hat{\theta})\}\{\Delta_N(\hat{\theta})+\mathcal{J}_N(\hat{\theta},\hat{\mu}^{(3)},\hat{\mu}^{(4)})\}^{-1/2}.$$

A similar definition can be used in the other two cases.

Theorem 3.2 addresses the asymptotic properties of the proposed statistical tests under the three null hypotheses. We consider the power or limiting distribution under local and fixed alternatives. The fixed alternative is simple in that the powers of the three tests are close to 1. We then focus on the local alternative. To this end, we set the true values under the local alternatives for the three tests, respectively, as $H_1^g:\widetilde{\theta}=\widetilde{\theta}_g/\sqrt{N}=(\rho_g/\sqrt{N},\gamma_g/\sqrt{N},\beta_g^{\mathrm{T}}/\sqrt{N})^{\mathrm{T}}$, $H_1^s:(\rho,\gamma)=(\rho_s/\sqrt{N},\gamma_s/\sqrt{N})$ and $H_1^b:\beta=\beta_b/\sqrt{N}$, where $\widetilde{\theta}_g$ is a constant vector and $\|\widetilde{\theta}_g\|<\infty$. Additionally, ρ_s, γ_s and β_b are the non-zero constants and $|\rho_s|<\infty$, $|\gamma_s|<\infty$ and $|\beta_b|<\infty$. The associated power functions are defined as

$$\mathcal{B}_g=\mathrm{P}(T_g>\chi^2_{\mathrm{weight},g}(1-\alpha)\mid H_1^g),\quad \mathcal{B}_s=\mathrm{P}(T_s>\chi^2_{\mathrm{weight},s}(1-\alpha)\mid H_1^s)$$

and

$$\mathcal{B}_b=\mathrm{P}(T_b>\chi^2_{\mathrm{weight},b}(1-\alpha)\mid H_1^b),$$

where $\chi^2_{\mathrm{weight},g}(1-\alpha)$, $\chi^2_{\mathrm{weight},s}(1-\alpha)$ and $\chi^2_{\mathrm{weight},b}(1-\alpha)$ are the $1-\alpha$ theoretical quantiles of the weighted chi-square distributions $\sum_{i=1}^{p+3}\lambda_{i,g}\chi^2_{i,1}$, $\sum_{i=1}^{p+3}\lambda_{i,s}\chi^2_{i,1}$ and $\sum_{i=1}^{p+3}\lambda_{i,b}\chi^2_{i,1}$. The asymptotic properties of the three power functions are given below.

Theorem 3.3. *Under Conditions* (C1)–(C4), *we can obtain*

$$\lim_{N\to\infty}(\mathcal{B}_g)=1-F_g\{\chi^2_{\mathrm{weight},g}(1-\alpha)-\theta_g^{\mathrm{T}}\Delta\theta_g\},\quad \lim_{N\to\infty}(\mathcal{B}_s)=1-F_s\{\chi^2_{\mathrm{weight},s}(1-\alpha)-\theta_s^{\mathrm{T}}\Delta\theta_s\}$$

and

$$\lim_{N\to\infty}(\mathcal{B}_b)=1-F_b\{\chi^2_{\mathrm{weight},b}(1-\alpha)-\theta_b^{\mathrm{T}}\Delta\theta_b\},$$

where $F_g(\cdot)$, $F_s(\cdot)$ and $F_b(\cdot)$ are the cumulative distribution functions of the weighted chi-square distributions $\sum_{i=1}^{p+3}\lambda_{i,g}\chi^2_{i,1}$, $\sum_{i=1}^{p+3}\lambda_{i,s}\chi^2_{i,1}$ and $\sum_{i=1}^{p+3}\lambda_{i,b}\chi^2_{i,1}$, where $\theta_g=(\rho_g,\gamma_g,\beta_g^{\mathrm{T}},0)^{\mathrm{T}}$, $\theta_s=(\rho_s,\gamma_s,\mathbf{0}^{\mathrm{T}},0)^{\mathrm{T}}$ and $\theta_b=(0,0,\beta_l^{\mathrm{T}},0)^{\mathrm{T}}$.

The above theorem reveals that the three tests T_g, T_s and T_b are consistent as long as $\theta_g^{\mathrm{T}}\Delta\theta_g\to\infty$, $\theta_s^{\mathrm{T}}\Delta\theta_s\to\infty$ and $\theta_b^{\mathrm{T}}\Delta\theta_b\to\infty$.

3.3 Simulation studies

To demonstrate the finite sample performance of the proposed method, we present a total of five simulation studies. These simulation studies are similar to each other except for the generating mechanism of the adjacency matrix A. For illustration purposes, we consider $X_i=(X_{i1},X_{i2})^{\mathrm{T}}\in\mathbb{R}^2$ with $p=2$, and X_i is generated from a two-dimensional normal distribution with a mean of 0 and a variance matrix of $2I$. The true value of $\theta=(\rho,\gamma,\beta^{\mathrm{T}},\sigma^2)^{\mathrm{T}}$ is set as $\theta_0=(0.1,0.1,0.5,0.5,1)^{\mathrm{T}}$. The response variable $\mathbb{Y}$ is then generated according to the model (2.1). We consider various network sizes N equaling 100, 200, 500, 1,000 and 2,000. For each simulation sample and network size, the experiment is randomly replicated $M=1{,}000$ times. We use $\hat{\tau}^{(m)}$ to represent one particular estimator (e.g., $\hat{\rho}$) obtained in the m-th replication. We assume the true parameter to be τ_0, and the root-mean-square error (RMSE) is evaluated as

$$\mathrm{RMSE}=\left\{M^{-1}\sum_{m=1}^{M}(\hat{\tau}^{(m)}-\tau_0)^2\right\}^{1/2}.$$

In addition to that, a 95% confidence interval (CI) is constructed as

$$\mathrm{CI}^{(m)}=(\hat{\tau}^{(m)}-z_{0.975}\widehat{\mathrm{SE}}^{(m)},\hat{\tau}^{(m)}+z_{0.975}\widehat{\mathrm{SE}}^{(m)}),$$

where $\widehat{\mathrm{SE}}^{(m)}$ is computed according to the asymptotic covariance formula given in Theorem 3.1 by replacing the unknown parameters with their estimates. Here, z_α is the α-th quantile of a standard normal distribution. Accordingly, the empirical coverage probability (ECP) is computed as

$$\mathrm{ECP} = M^{-1} \sum_{m=1}^{M} I(\tau_0 \in \mathrm{CI}^{(r)}),$$

where $I(\cdot)$ is the indicator function. Next, we present the generating mechanism of the five specific network models and their simulation results.

Example 3.1 (Erdös-Rényi (ER) network model)**.** We first present a simple network structure called the ER network [9]. It is a random network with different a_{ij}'s $(i \neq j)$ independently generated with a fixed probability p. In this example, we set $p = 5/N$ so that the resulting network structure is reasonably sparse. One typical simulated network structure of this type and its degree distributions are displayed in Figure 10. For this network structure, the histograms of both the in-degree and out-degree distributions are somewhat symmetric. Simulation results are summarized in Table 2, where we find that the RMSE values steadily decrease towards 0 and the ECP values are close to the nominal level of 95% as $N \to \infty$. This suggests that the estimators are asymptotically normal and the estimated standard error (i.e., $\widehat{\mathrm{SE}}$) can approximate the true standard error well.

Example 3.2 (Dyad independence model)**.** We next consider a network with a dyad independence structure [19]. Compared with the ER network, a dyad independence model allows a_{ij} to be dependent on a_{ji} for any $i < j$. However, for any $(i_1, j_1) \neq (i_2, j_2)$ with $i_1 < j_1$ and $i_2 < j_2$, the two dyads $(a_{i_1 j_1}, a_{j_1 i_1})$ and $(a_{i_2 j_2}, a_{j_2 i_2})$ are assumed to be independent. Specifically, we generate A as follows. Define $Z_{ij} = (a_{ij}, a_{ji})$ for any $1 \leqslant i < j \leqslant N$. Next, set $\mathrm{P}\{Z_{ij} = (1,0)\} = \mathrm{P}\{Z_{ij} = (0,1)\} = 5N^{-1}$ and $\mathrm{P}\{Z_{ij} = (1,1)\} = 0.5N^{-1}$. This leads to $\mathrm{P}\{Z_{ij} = (0,0)\} = 1 - 5.5N^{-1}$, which is close to 1 for a large N. In this way, we generate Z_{ij} independently, leading to the final A and W. One typical simulated network structure of this type and its degree distributions are displayed in Figure 11. For this network structure, the histograms of both in-degree and out-degree distributions are approximately normal. The detailed results are given in Table 3, and they are qualitatively similar to those in Table 2.

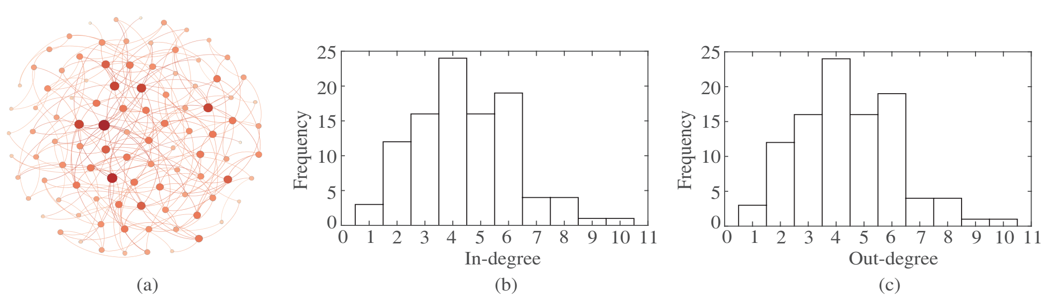


Figure 10 (Color online) One particular random realization of the ER network model. (a) Visualization of the network structure. (b) Histogram of the in-degree distribution. (c) Histogram of the out-degree distribution

Table 2 Simulation results for the ER network. The RMSE values are reported for every estimator. The ECPs (in %) are given in parentheses

N	$\hat{\rho}$	$\hat{\gamma}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
100	0.3218 (90.4)	0.2378 (89.4)	0.0505 (93.9)	0.0532 (93.9)	0.6799 (90.8)
200	0.1124 (91.8)	0.1627 (91.6)	0.0353 (95.1)	0.0345 (95.1)	0.1037 (93.5)
500	0.0628 (93.4)	0.0934 (93.6)	0.0215 (96.1)	0.0227 (96.1)	0.0630 (94.6)
1,000	0.0442 (95.3)	0.0641 (94.5)	0.0158 (95.1)	0.0161 (95.1)	0.0453 (93.8)
2,000	0.0305 (95.1)	0.0430 (95.0)	0.0117 (93.3)	0.0111 (93.3)	0.0314 (95.6)

Example 3.3 (Stochastic block model). The third network model is a stochastic block structure [26,29]. Specifically, let $K = 5$ be the total number of blocks. We then follow Zhu et al. [34], and randomly assign a block label to each node with an equal probability of $1/K$. Next, we set $\mathrm{P}(a_{ij} = 1) = 20N^{-1}$ if i and j are in the same block, and $\mathrm{P}(a_{ij} = 1) = 0.02N^{-1}$ otherwise. Correspondingly, the nodes within the same block are more likely to be connected to each other, while those from different blocks are less likely to be connected. One typical simulated network structure of this type and its degree distributions are displayed in Figure 12. For this network structure, the histograms of both the in-degree and out-degree distributions are approximately normal. The results of the model are shown in Table 4; they are qualitatively similar to those in Table 2.

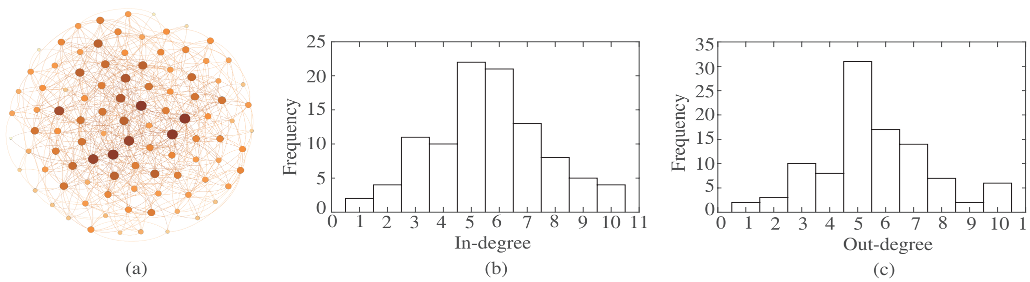


Figure 11 (Color online) One particular random realization of the dyad independence model. (a) Visualization of the network structure. (b) Histogram of the in-degree distribution. (c) Histogram of the out-degree distribution

Table 3 Simulation results for the dyad independence model. The RMSE values are reported for every estimator. The ECPs (in %) are given in parentheses

N	$\hat{\rho}$	$\hat{\gamma}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
100	0.1704 (90.7)	0.2361 (89.5)	0.0541 (94.1)	0.0595 (94.1)	0.2210 (91.3)
200	0.1077 (92.6)	0.1587 (92.3)	0.0352 (94.9)	0.0349 (94.9)	0.1031 (92.3)
500	0.0651 (94.0)	0.0970 (93.3)	0.0216 (96.1)	0.0228 (96.1)	0.0632 (93.9)
1,000	0.0441 (95.8)	0.0664 (94.1)	0.0158 (94.9)	0.0162 (94.9)	0.0453 (93.5)
2,000	0.0329 (94.4)	0.0459 (95.6)	0.0118 (93.4)	0.0111 (93.4)	0.0316 (94.8)

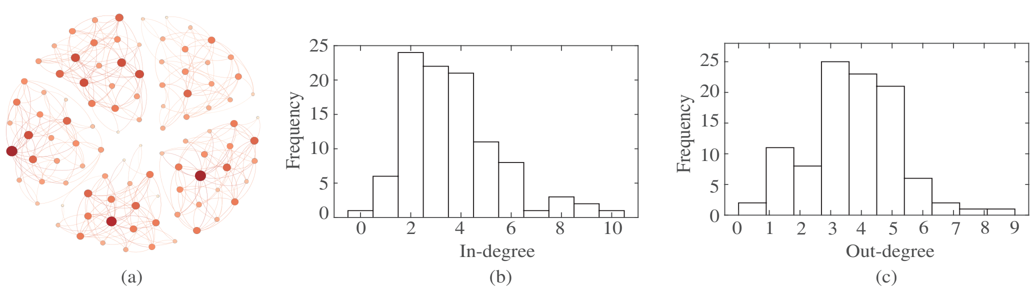


Figure 12 (Color online) One particular random realization of the stochastic block model. (a) Visualization of the network structure. (b) Histogram of the in-degree distribution. (c) Histogram of the out-degree distribution

Table 4 Simulation results for the stochastic block model. The RMSE values are reported for every estimator. The ECPs (in %) are given in parentheses

N	$\hat{\rho}$	$\hat{\gamma}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
100	0.1451 (91.3)	0.2341 (91.3)	0.0506 (94.0)	0.0520 (94.0)	0.2687 (91.8)
200	0.0854 (94.1)	0.1511 (93.2)	0.0345 (95.0)	0.0334 (95.0)	0.1015 (92.6)
500	0.0568 (93.6)	0.0967 (94.7)	0.0214 (96.2)	0.0225 (96.2)	0.0630 (94.0)
1,000	0.0370 (95.7)	0.0669 (94.9)	0.0159 (95.3)	0.0162 (95.3)	0.0453 (93.4)
2,000	0.0276 (94.7)	0.0489 (94.4)	0.0118 (93.4)	0.0111 (93.4)	0.0314 (94.7)

Example 3.4 (Power-law distributed model). The power-law distributed network structure is another important network structure in practice [3, 7]. The main feature of a power-law distributed network structure is that the majority of nodes (e.g., normal people) have very few connections, but a small amount of nodes (e.g., movie stars) have large numbers of connections. To simulate such a network structure, we follow Zhou et al. [32] and generate A as follows. First, for each node i, we generate its in-degree, $d_i = \sum_j a_{ji}$, according to the discrete power-law distribution with $\mathrm{P}(d_i = k) = sk^{-\alpha}$, where s is a normalizing constant and the exponent parameter is set as $\alpha = 2$. Next, for each i, we randomly select d_i nodes as i's potential followers. One typical simulated network structure of this type and its degree distributions are displayed in Figure 13. For this network structure, the histograms of both the in-degree and out-degree distributions are lightly skewed. The detailed simulated results are showed in Table 5, and they are qualitatively similar to those in Table 2.

Example 3.5 (Popularity scaled latent space model). The last network structure is generated according to the popularity scaled latent space model (PSLSM) of [6]. They assume a position for each node in a hypothetically assumed latent space. Two nodes therefore are more likely to be connected with each other if they stay close to each other. In addition, the nodes with large popularity values are likely to have more connections. Specifically, for each node i, we generate its latent space position $Z_i \in \mathbb{R}^1$ using a standard normal distribution. Next, the popularity parameter λ_i ($1 \leqslant i \leqslant N$) is independent and identically drawn from a power-law type distribution $\mathrm{P}(\lambda_i = k) = sk^{-\alpha}$, where s is a normalizing constant and $\alpha = 1.5$. Conditional on $\mathbb{Z} = \{Z_i : 1 \leqslant i \leqslant N\}$, the a_{ij}'s are independently generated according to

$$\mathrm{P}(a_{ij} = 1 \mid Z_i, Z_j, \lambda_j) = \exp\left\{\frac{N^2(Z_i - Z_j)^2}{2\lambda_j^2}\right\}.$$

One typical simulated network structure of this type and its degree distributions are displayed in Figure 14. For this network structure, the in-degree histogram is highly skewed, while the out-degree histogram is approximately symmetric. The numerical results are summarized in Table 6, and they are qualitatively similar to those in Table 2.

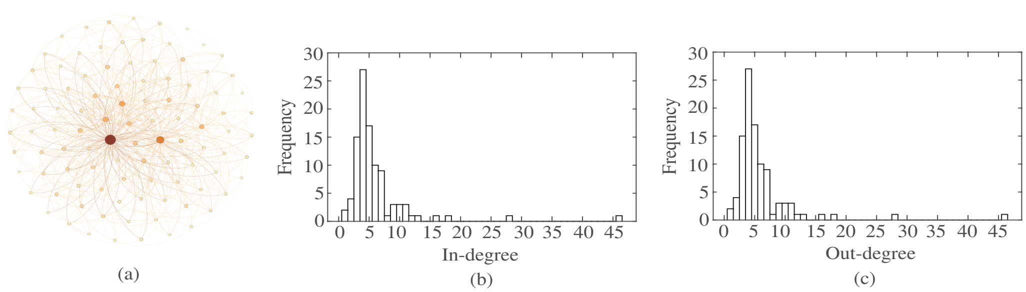
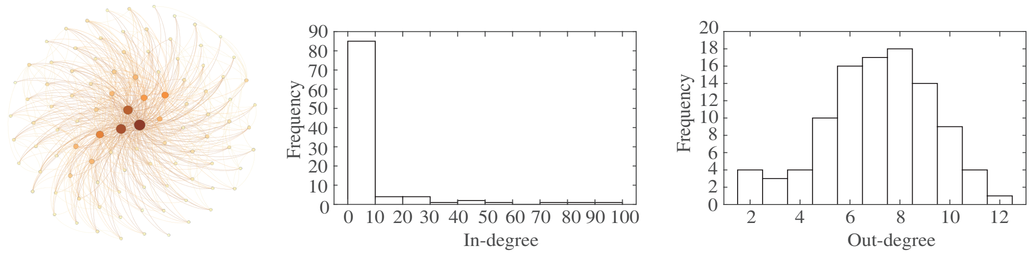


Figure 13 (Color online) One particular random realization of the power-law model. (a) Visualization of the network structure. (b) Histogram of the in-degree distribution. (c) Histogram of the out-degree distribution

Table 5 Simulation results for the power-law model. The RMSE values are reported for every estimator. The ECPs (in %) are given in parentheses

N	$\hat{\rho}$	$\hat{\gamma}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
100	0.2042 (89.3)	0.3164 (87.6)	0.0538 (92.8)	0.0537 (92.8)	0.1811 (90.9)
200	0.1121 (93.4)	0.1723 (91.2)	0.0354 (94.7)	0.0346 (94.7)	0.1039 (93.3)
500	0.0745 (94.0)	0.1100 (93.8)	0.0216 (95.7)	0.0228 (95.7)	0.0628 (94.6)
1,000	0.0543 (94.8)	0.0819 (94.3)	0.0159 (95.3)	0.0162 (95.3)	0.0453 (94.3)
2,000	0.0394 (94.4)	0.0631 (94.4)	0.0118 (93.4)	0.0111 (93.4)	0.0316 (95.1)



(a) (b) (c)

Figure 14 (Color online) One particular random realization of the PSLSM. (a) Visualization of the network structure. (b) Histogram of the in-degree distribution. (c) Histogram of the out-degree distribution

Table 6 Simulation results for the PSLSM. The RMSE values are reported for every estimator. The ECPs (in %) are given in parentheses

N	$\hat{\rho}$	$\hat{\gamma}$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$
100	0.1444 (91.1)	0.2455 (87.9)	0.0515 (94.5)	0.0547 (94.5)	0.1478 (92.1)
200	0.0897 (93.3)	0.1464 (92.7)	0.0360 (93.9)	0.0358 (93.9)	0.1005 (93.0)
500	0.0527 (94.9)	0.0890 (93.9)	0.0236 (93.5)	0.0223 (93.5)	0.0654 (93.5)
1,000	0.0413 (94.4)	0.0656 (94.8)	0.0156 (96.1)	0.0156 (96.1)	0.0462 (94.0)
2,000	0.0276 (94.8)	0.0427 (94.7)	0.0114 (95.4)	0.0113 (95.4)	0.0319 (95.1)

3.4 Hypothesis testing results

We further study the performance of the quasi-likelihood ratio tests by their empirical size and power with a significance level of 0.05. First, we evaluate the empirical size of the quasi-likelihood ratio test under three different cases. Case 1 sets the true value of $\theta = (\rho, \gamma, \beta^{\mathrm{T}}, \sigma^2)^{\mathrm{T}}$ as $\theta_0 = (0, 0, 0, 0, 1)^{\mathrm{T}}$ so that the empirical size of the global significant test can be evaluated. Case 2 sets $\theta_0 = (0, 0, 0.1, 0.1, 1)^{\mathrm{T}}$ so that the empirical size of the shareholder network effect can be evaluated. Case 3 sets $\theta_0 = (0.1, 0.1, 0, 0, 1)^{\mathrm{T}}$ so that the empirical size of the industry effect can be evaluated. Various network sizes N equaling 500, 1,000 and 2,000 are considered. For each simulation example and network size, the experiment is randomly replicated $M = 1{,}000$ times. Let $T_g^{(m)}$ represent the test statistic for global significance obtained in the m-th replication. We independently and identically generate

$$\{\chi^2_{i,n} : i = 1, \ldots, p+3,\ n = 1, \ldots, 10000\}$$

from the chi-squared distribution with degree of freedom 1. The m-th p-value is given by

$$p_m\text{-value} = 10000^{-1} \sum_n I\bigg\{T_g^{(m)} > \sum_{i=1}^{p+3} \lambda_{N,i,g}(\hat{\theta}, \hat{\mu}^{(3)}, \hat{\mu}^{(4)}) \chi^2_{i,n}\bigg\},$$

where $\lambda_{N,i,g}(\hat{\theta}, \hat{\mu}^{(3)}, \hat{\mu}^{(4)})$ is a consistent estimator of $\lambda_{i,g}(\hat{\theta}, \hat{\mu}^{(3)}, \hat{\mu}^{(4)})$ under the null hypothesis stated below Theorem 3.2. The other two testing methods are evaluated similarly, and the simulated results are summarized in Table 7.

Second, we examine the empirical power of the quasi-likelihood ratio test with a significance level of 0.05 under the same setting except for a different $\theta_0 = (0.1, 0.1, 0.05, 0.05, 1)^{\mathrm{T}}$. The simulated results are summarized in Table 8. For the purposes of illustration, we also consider the setting of $\theta_0 = (1.5\varrho, 1.5\varrho, 0.8\varrho, 0.8\varrho, 1)$ with a fixed sample size $N = 500$, where ϱ equaling 0.02, 0.05 and 0.10 measures the signal strength of the parameters. The simulation results are shown in Table 9. From Tables 7–9, we can draw the following conclusions. The first is that the empirical sizes of the three tests are close to a confidence level of 0.05. The second is that the empirical powers of the three tests tend to 100% when the sample size N or the signal strength ϱ obtains larger. These two findings indicate that the three tests perform well when N is large.

Table 7 Empirical size of the quasi-likelihood ratio test

N	Case	ER	DI	SB	PLD	PSLSM
	1	0.042	0.043	0.017	0.039	0.027
500	2	0.063	0.067	0.031	0.050	0.053
	3	0.058	0.078	0.062	0.083	0.070
	1	0.037	0.037	0.023	0.042	0.038
1,000	2	0.055	0.047	0.058	0.054	0.060
	3	0.055	0.061	0.054	0.064	0.061
	1	0.041	0.038	0.033	0.040	0.042
2,000	2	0.054	0.045	0.042	0.063	0.050
	3	0.066	0.060	0.059	0.060	0.054

Table 8 Empirical power of the quasi-likelihood ratio test

N	Case	ER	DI	SB	PLD	PSLSM
	1	0.954	0.959	0.865	0.891	0.967
500	2	0.762	0.769	0.566	0.503	0.777
	3	0.841	0.834	0.807	0.841	0.830
	1	0.999	1.000	0.962	0.998	0.999
1,000	2	0.966	0.970	0.807	0.865	0.944
	3	0.984	0.977	0.961	0.981	0.982
	1	1.000	1.000	0.998	1.000	1.000
2,000	2	1.000	1.000	0.951	0.992	0.999
	3	1.000	1.000	1.000	1.000	1.000

Table 9 Empirical power of the hypothesis testing results

ϱ	Case	ER	DI	SB	PLD	PSLSM
	1	0.158	0.137	0.117	0.132	0.122
0.02	2	0.105	0.091	0.104	0.070	0.067
	3	0.175	0.169	0.171	0.184	0.171
	1	0.776	0.771	0.634	0.662	0.777
0.05	2	0.503	0.490	0.349	0.302	0.512
	3	0.661	0.646	0.631	0.663	0.661
	1	1.000	0.999	0.997	1.000	1.000
0.10	2	0.990	0.988	0.910	0.850	0.987
	3	0.996	0.997	0.994	1.000	0.998

4 Concluding remarks

In this article, we proposed an SARMA model to study inter-stock dependence among multiple stocks in the Chinese stock market. To investigate the effect of inter-stock dependence, a common shareholder network was used to construct an adjacency matrix. The proposed model then made use of the inter-stock dependence, error dependence and exogenous industry information. A thoroughly developed application using the Chinese stock market data was then extensively studied. To solve the parameter estimation problem, a quasi-maximum likelihood estimation method was proposed, and the associated asymptotic properties were established. The performance of the QMLE was proven by extensive simulation studies.

To conclude this article, we discuss several interesting topics for future study. First, for a large-scale dataset, calculating the QMLE is computationally expensive [22, 31]. This is mainly because the determinant and inverse of an ultra-high-dimensional matrix are involved. Some more computationally efficient methods could be considered, such as GMM. Second, many responses in the real world are time series observations. Therefore, time dynamics could be considered and statistically modeled. Third, the SARMA model requires the responses to be continuous. However, discrete responses are frequently encountered in real data analysis. This is also an interesting topic that needs to be further investigated. Lastly, the network connectivity in our case is known and does not need to be estimated. However, a more realistic situation is that the network adjacency matrix is estimated with the non-ignorable statistical error. Thus, exactly how to apply the proposed method is not yet clear at the moment. We consider this an excellent topic for future study.

Acknowledgements The first author was supported by the Major Program of the National Natural Science Foundation of China (Grant No. 11731101) and National Natural Science Foundation of China (Grant No. 11671349). The second author was supported by National Natural Science Foundation of China (Grant No. 72171226), the Beijing Municipal Social Science Foundation (Grant No. 19GLC052) and the National Statistical Science Research Project (Grant No. 2020LZ38). The third author was supported by National Natural Science Foundation of China (Grant Nos. 71532001, 11931014, 12171395 and 71991472) and the Joint Lab of Data Science and Business Intelligence at Southwestern University of Finance and Economics. The fourth author was supported by National Natural Science Foundation of China (Grant No. 11831008) and the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science (Grant No. Klatasds-Moe-Ecnu-Klatasds2101).

References

1 Anselin L, Florax R. New Directions in Spatial Econometrics. London: Springer, 1995
2 Baltagi B, Fingleton B, Pirotte A. A time-space dynamic panel data model with spatial moving average errors. Regional Sci Urban Econom, 2019, 76: 13–31
3 Barabasi A, Albert R. Emergence of scaling in random networks. Science, 1999, 286: 509–512
4 Chan L, Karceski J, Lakonishok J. The risk and return from factors. J Financ Quant Anal, 1998, 33: 159–188
5 Chan W. A comparison of some pattern identification methods for order determination of mixed ARMA models. Statist Probab Lett, 1999, 42: 69–79
6 Chang X, Huang D, Wang H. A popularity scaled latent space model for large-scale directed social network. Statist Sinica, 2019, 29: 1277–1299
7 Clauset A, Shalizi C, Newman M. Power-law distributions in empirical data. SIAM Rev, 2009, 51: 661–703
8 Dogan O. GMM estimation of spatial autoregressive models with moving average disturbances. Regional Sci Urban Econom, 2013, 43: 903–926
9 Erdös P, Rényi A. On the evolution of random graphs. Magyar Tud Akad Mat Kutato Int Közl, 1960, 5: 17–61
10 Fabio M. Learning about the interdependence between the macroeconomy and the stock market. Int Rev Econom Finance, 2017, 49: 223–242
11 Fama E, French K. Common risk factors in the returns on stocks and bonds. J Financ Econom, 1993, 33: 3–56
12 Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Amer Statist Assoc, 2001, 96: 1348–1360
13 Fang N, Jiang W, Luo R. Realized semivariances and the variation of signed jumps in China's stock market. Emerging Market Finance Trade, 2017, 53: 563–586
14 Fingleton B. A generalized method of moments estimator for a spatial model with moving average errors, with application to real estate prices. Empirical Econom, 2008, 34: 35–37
15 Fingleton B, Gallo J, Pirotte A. A multidimensional spatial lag panel data model with spatial moving average nested random effects errors. Empirical Econom, 2018, 55: 113–146
16 Gul F, Kim J, Qiu A. Ownership concentration, foreign shareholding, audit quality, and stock price synchronicity: Evidence from China. J Financ Econom, 2010, 95: 425–442
17 Gupta A, Robinson P. Inference on higher-order spatial autoregressive models with increasingly many parameters. J Econometrics, 2015, 186: 19–31
18 Han B, Yang L. Social networks, information acquisition, and asset prices. Manag Sci, 2013, 59: 1444–1457
19 Holland P, Leinhardt S. An exponential family of probability distributions for directed graphs. J Amer Statist Assoc, 1981, 76: 33–50
20 Kelejian H, Prucha I. On the asymptotic distribution of the Moran I test statistic with applications. J Econometrics, 2001, 104: 219–257
21 Lee L. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models.

Econometrica, 2004, 72: 1899–1925
22 Liu X, Lee L, Bollinger C. An efficient GMM estimator of spatial autoregressive models. J Econometrics, 2010, 159: 303–319
23 Luo R, Liu Y, Lan W. A penalized expected risk criterion for portfolio selection. China Financ Rev Int, 2019, 9: 386–400
24 Ma Y, Pan R, Zou T, et al. A naive least squares method for spatial autoregression with covariates. Statist Sinica, 2020, 30: 653–672
25 Markowitz H. Portfolio selection. J Finance, 1952, 7: 77–91
26 Nowicki K, Snijders T. Estimation and prediction for stochastic block structures. J Amer Statist Assoc, 2001, 96: 1077–1087
27 Ozsoylev H, Walden J. Asset pricing in large information networks. J Econom Theory, 2011, 146: 2252–2280
28 Sharpe W. Capital asset prices: A theory of market equilibrium under conditions of risk. J Finance, 1964, 19: 425–442
29 Wang Y, Wong G. Stochastic block models for directed graphs. J Amer Statist Assoc, 1987, 82: 8–19
30 Yu J, Jong R, Lee L. Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. J Econometrics, 2008, 146: 118–134
31 Zhou J, Chen Y, Wang H. Estimating spatial autocorrelation with sampled network data. J Bus Econom Statist, 2017, 35: 130–138
32 Zhou J, Huang D, Wang H. A note on estimating network dependence in a discrete choice model. Stat Interface, 2018, 11: 433–439
33 Zhou J, Li D, Pan R, et al. Network GARCH model. Statist Sinica, 2020, 30: 1723–1740
34 Zhu X, Chang X, Li R, et al. Portal nodes screening for large scale social networks. J Econometrics, 2019, 209: 145–157
35 Zou T, Lan W, Wang H, et al. Covariance regression analysis. J Amer Statist Assoc, 2017, 112: 266–281
36 Zou T, Luo R, Lan W, et al. Network influence analysis. Statist Sinica, 2021, 31: 1727–1748

Appendix A Four useful lemmas

This appendix introduces four useful lemmas. Before providing the technical lemmas, we first introduce some notations. Let $\|\cdot\|_s$ denote the vector s-norm or the matrix s-norm for $1 \leqslant s \leqslant \infty$. In other words, for any generic vector $x = (x_1, \ldots, x_q)^{\mathrm{T}} \in \mathbb{R}^q$, $\|x\|_s = (\sum_{i=1}^q |x_i|^s)^{1/s}$, and for any generic matrix $G \in \mathbb{R}^{m\times q}$, we have

$$\|G\|_s = \sup\{\|Gx\|_s/\|x\|_s : x \in \mathbb{R}^{q\times 1} \text{ and } x \neq 0\}.$$

In addition, define the element-wise ℓ_∞ norm for any generic matrix G as $|G|_\infty = \|\mathrm{vec}(G)\|_\infty$, where $\mathrm{vec}(G)$ denotes the vectorization for any generic matrix G. We next introduce the following four useful lemmas. Since Lemma A.2 is directly modified from [20], we only present the proofs of the remaining three lemmas.

Lemma A.1. *For any vector $\alpha = (\alpha_1, \ldots, \alpha_p)^{\mathrm{T}} \in \mathbb{R}^p$, we have the matrices $M \in \mathbb{R}^{N\times p}$ and $U \in \mathbb{R}^{N\times N}$. Then for any $s \geqslant 1$, we have $\|UM\alpha\|_s \leqslant N^{1/s}|M|_\infty\|U\|_\infty\|\alpha\|_1$.*

Proof. Write $U = (u_{ij})_{N\times N} \in \mathbb{R}^{N\times N}$ and $M = (m_{ij})_{N\times p} \in \mathbb{R}^{N\times p}$. For any p-dimensional vector $\alpha = (\alpha_1, \ldots, \alpha_p)^{\mathrm{T}}$ and a fixed $1 \leqslant i \leqslant N$, we have

$$\begin{aligned}\left|\sum_{j=1}^N\sum_{k=1}^p u_{ij}m_{jk}\alpha_k\right| &\leqslant \sum_{j=1}^N\sum_{k=1}^p |u_{ij}||m_{jk}||\alpha_k| = \sum_{k=1}^p |\alpha_k| \sum_{j=1}^N |u_{ij}||m_{jk}| \\ &\leqslant |M|_\infty \sum_{k=1}^p |\alpha_k| \max_{1\leqslant i\leqslant N}\sum_{j=1}^N |u_{ij}| = |M|_\infty\|U\|_\infty\|\alpha\|_1.\end{aligned}$$

Then, we have

$$\|UM\alpha\|_s = \left\{\sum_{i=1}^N\left|\sum_{j=1}^N\sum_{k=1}^p u_{ij}m_{jk}x_k\right|^s\right\}^{1/s} \leqslant N^{1/s}|M|_\infty\|U\|_\infty\|\alpha\|_1,$$

which completes the proof. □

Lemma A.2. *Assume ε_i's are independent and identically distributed with a mean of* 0 *and a finite variance of σ^2. Define*

$$Q = \mathcal{E}^{\mathrm{T}} H \mathcal{E} + B^{\mathrm{T}} \mathcal{E} - \sigma^2 \operatorname{tr}(H),$$

where $H = (h_{ij})_{N\times N} \in \mathbb{R}^{N\times N}$ and $B = (b_1, \ldots, b_N)^{\mathrm{T}} \in \mathbb{R}^{N\times 1}$. Suppose that the following assumptions are satisfied: (A) $\sup_{N\geqslant 1} \|H\|_1 = \sup_{N\geqslant 1} \sum_{i=1}^N |h_{ij}| < \infty$, (B) *for some* $\eta_1 > 0$, $\sup_{N\geqslant 1} N^{-1} \|B\|_{2+\eta_1}^{2+\eta_1} < \infty$, *and* (C) *for some* $\eta_2 > 0$, $\mathrm{E}|\varepsilon_i|^{4+\eta_2} < \infty$. *Then we have* $\mathrm{E}(Q) = 0$ *and*

$$\sigma_Q^2 := \sigma^4 \sum_{i=1}^N \sum_{j\neq i} h_{ij}^2 + 2 \sum_{i=1}^N \sum_{j\neq i} h_{ij} b_i \mu^{(3)} + \sigma^2 \sum_{i=1}^N b_i^2 + \{\mu^{(4)} - \sigma^4\} \sum_{i=1}^N h_{ii}^2,$$

where $\mu^{(s)} = \mathrm{E}(\varepsilon_i^s)$ *for* $s = 3, 4$. *Furthermore, suppose* (D) $N^{-1}\sigma_Q^2 \geqslant c$ *for some* $c > 0$. *Then we obtain* $N^{-1/2-\epsilon} Q \to_p 0$ *for any* $\epsilon > 0$ *and* $\sigma_Q^{-1} Q \to_d N(0, 1)$.

Lemma A.3. *Define*

$$\partial\ell(\theta)/\partial\theta = (\partial\ell(\theta)/\partial\rho, \partial\ell(\theta)/\partial\gamma, \partial\ell(\theta)/\partial\beta, \partial\ell(\theta)/\partial\sigma^2)^{\mathrm{T}}.$$

Under Conditions (C1)–(C4), *we have that* $N^{-1/2}\partial\ell(\theta)/\partial\theta \to_d N(0, \Delta + \mathcal{J})$ *as* $N \to \infty$.

Proof. Recall that $K = I - \rho W$, $S = I + \gamma W$ and $G = (I - \rho W)\mathbb{Y} - \mathbb{X}\beta$. We then compute their first-order derivatives as

$$\begin{aligned}
\frac{\partial\ell(\theta)}{\partial\rho} &= -\mathrm{tr}(K^{-1}W) + \frac{1}{\sigma^2}\mathbb{Y}^{\mathrm{T}} W^{\mathrm{T}} (S^{\mathrm{T}})^{-1} S^{-1} G,\\
\frac{\partial\ell(\theta)}{\partial\gamma} &= -\mathrm{tr}(S^{-1}W) + \frac{1}{\sigma^2} G^{\mathrm{T}} (S^{\mathrm{T}})^{-1} W^{\mathrm{T}} (S^{\mathrm{T}})^{-1} S^{-1} G,\\
\frac{\partial\ell(\theta)}{\partial\beta} &= \frac{1}{\sigma^2}\mathbb{X}^{\mathrm{T}} (S^{\mathrm{T}})^{-1} S^{-1} G, \quad \frac{\partial\ell(\theta)}{\partial\sigma^2} = -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} G^{\mathrm{T}} (S^{\mathrm{T}})^{-1} S^{-1} G.
\end{aligned}$$

According to the model (2.1), we have

$$\mathbb{Y} = (I - \rho W)^{-1}\{\mathbb{X}\beta + (I + \gamma W)\mathcal{E}\}.$$

Then after a tedious calculation, we obtain

$$\frac{\partial\ell(\theta)}{\partial\theta} = \begin{pmatrix} \mathcal{E}^{\mathrm{T}} H_1 \mathcal{E} \\ \vdots \\ \mathcal{E}^{\mathrm{T}} H_L \mathcal{E} \end{pmatrix} + \begin{pmatrix} B_1 \\ \vdots \\ B_L \end{pmatrix} \mathcal{E} - \sigma^2 \begin{pmatrix} tr(H_1) \\ \vdots \\ tr(H_L) \end{pmatrix}, \tag{A.1}$$

where $L = 4$, $H_1 = (K^{\mathrm{T}})^{-1} W^{\mathrm{T}}$, $H_2 = (S^{\mathrm{T}})^{-1} W^{\mathrm{T}}$, $H_3 = 1/2\sigma^4 I$, $H_4 = 0_{N\times N}$, $B_1 = \frac{1}{\sigma^2}(\mathbb{X}\beta)^{\mathrm{T}} (K^{\mathrm{T}})^{-1} W^{\mathrm{T}} (S^{\mathrm{T}})^{-1}$, $B_2 = 0$, $B_3 = 0$ and $B_4 = \mathbb{X}^{\mathrm{T}} (S^{\mathrm{T}})^{-1}$. Let

$$Q_l = \mathcal{E}^{\mathrm{T}} H_l \mathcal{E} + B_l^{\mathrm{T}} \mathcal{E} - \sigma^2 \operatorname{tr}(H_l),$$

where $l = 1, \ldots, 4$. In addition, for any generic vector $t = (t_1, \ldots, t_4)^{\mathrm{T}} \neq 0$, let

$$R = t^{\mathrm{T}} Q(t) = \mathcal{E}^{\mathrm{T}} H(t) \mathcal{E} + t^{\mathrm{T}} B(t) \mathcal{E} - \sigma^2 \operatorname{tr}\{H(t)\},$$

where $Q(t) = (Q_1, \ldots, Q_4)^{\mathrm{T}}$, $H(t) = (\sum_{l=1}^4 t_l H_l)_{N\times N}$ and $B(t) = (B_1, \ldots, B_4)^{\mathrm{T}}$.

According to Conditions (C2)–(C3) and Lemma A.1, we have

$$\begin{aligned}
&\|H_1\|_1 \leqslant \|K^{-1}\|_\infty \|W\|_\infty < \infty, \quad \|H_2\|_1 \leqslant \|S^{-1}\|_\infty \|W\|_\infty < \infty,\\
&\|H_3\|_1 = \frac{1}{2\sigma^4}\|I\|_1 = \frac{1}{2\sigma^4} < \infty, \quad \|H_4\|_1 = 0.
\end{aligned}$$

Consequently, we obtain

$$\sup_{N\geqslant 1}\|H(t)\|_1<\infty.$$

As a result, Condition (A) of Lemma A.2 is satisfied. According to Conditions (C2) and (C3) and Lemma A.1 again, for any $\eta_1>0$, we have

$$\begin{aligned}
N^{-1}\|B_1\|_{2+\eta_1}^{2+\eta_1}&=N^{-1}\frac{1}{\sigma^{2(2+\eta_1)}}\{\|(\mathbb{X}\beta)^{\mathrm{T}}(K^{\mathrm{T}})^{-1}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}\|_{2+\eta_1}^{2+\eta_1}\}\\
&\leqslant N^{-1}\frac{1}{\sigma^{2(2+\eta_1)}}\{\|\beta\|_1^{2+\eta_1}|\mathbb{X}|_\infty^{2+\eta_1}\|K^{-1}\|_\infty^{2+\eta_1}\|W\|_\infty^{2+\eta_1}\|S^{-1}\|_\infty^{2+\eta_1}\}<\infty,\\
N^{-1}\|B_2\|_{2+\eta_1}^{2+\eta_1}&=0,\quad N^{-1}\|B_3\|_{2+\eta_1}^{2+\eta_1}=0,\\
N^{-1}\|B_4\|_{2+\eta_1}^{2+\eta_1}&=\|\mathbb{X}^{\mathrm{T}}(S^{\mathrm{T}})^{-1}\|_{2+\eta_1}^{2+\eta_1}\leqslant|\mathbb{X}|_\infty^{2+\eta_1}\|S^{-1}\|_\infty^{2+\eta_1}<\infty.
\end{aligned}$$

This implies that

$$\sup_{N\geqslant 1}\|B(t)\|_{2+\eta_1}^{2+\eta_1}\leqslant 4^{1+\eta_1}\max_{1\leqslant l\leqslant 4}|t_l|^{2+\eta_1}\sup_{N\geqslant 1}N^{-1}\|B_l\|_{2+\eta_1}^{2+\eta_1}<\infty.$$

Hence, Condition (B) of Lemma A.2 is satisfied. Condition (C) of Lemma A.2 is obviously satisfied by Condition (C4). We next evaluate each component of $\mathrm{E}\{\partial\ell(\theta)/\partial\theta\}$ separately as follows:

$$\begin{aligned}
\mathrm{E}\bigg\{\frac{\partial\ell(\theta)}{\partial\rho}\bigg\}&=-\mathrm{tr}(K^{-1}W)+\mathrm{tr}\{(K^{\mathrm{T}})^{-1}W^{\mathrm{T}}\}=0,\\
\mathrm{E}\bigg\{\frac{\partial\ell(\theta)}{\partial\gamma}\bigg\}&=-\mathrm{tr}(S^{-1}W)+\mathrm{tr}\{W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}\}=0,\\
\mathrm{E}\bigg\{\frac{\partial\ell(\theta)}{\partial\beta}\bigg\}&=-\frac{1}{\sigma^2}\mathrm{E}\{\mathbb{X}^{\mathrm{T}}(S^{\mathrm{T}})^{-1}(S^{-1}KK^{-1}(\mathbb{X}\beta+S\mathcal{E})-S^{-1}\mathbb{X}\beta)\}\\
&=-\frac{1}{\sigma^2}\mathrm{E}\{\mathbb{X}^{\mathrm{T}}(S^{\mathrm{T}})^{-1}(S^{-1}\mathbb{X}\beta-S^{-1}\mathbb{X}\beta)\}=0,\\
\mathrm{E}\bigg\{\frac{\partial\ell(\theta)}{\partial\sigma^2}\bigg\}&=\frac{N}{2\sigma^4}-\frac{1}{\sigma^6}\mathrm{E}(\mathcal{E}^{\mathrm{T}}\mathcal{E})=\frac{N}{2\sigma^4}-\frac{1}{\sigma^6}\sigma^2\mathrm{tr}(I)=0.
\end{aligned}$$

Note that $\mathrm{E}(\mathcal{E}\mathcal{E}^{\mathrm{T}})=\sigma^2I$ and $\mathrm{E}(\mathcal{E})=\mathbf{0}$. We then compute $N^{-1}\mathrm{cov}\{\partial\ell(\theta)/\partial\theta\}$ as

$$\begin{aligned}
\frac{1}{N}\mathrm{E}\bigg\{\frac{\partial\ell(\theta)}{\partial\rho}\frac{\partial\ell(\theta)}{\partial\rho}\bigg\}&=\frac{1}{N}\mathrm{tr}(WK^{-1}K^{-1}W)-\frac{2}{N}\mathrm{tr}\{W^{\mathrm{T}}(K^{\mathrm{T}})^{-1}K^{-1}W\}\\
&\quad+\frac{1}{\sigma^2N}\mathrm{tr}\{(\mathbb{X}\beta)^{\mathrm{T}}(K^{\mathrm{T}})^{-1}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}WK^{-1}\mathbb{X}\beta\}\\
&\quad+\frac{2}{\sigma^4N}\mathrm{tr}\{(\mathbb{X}\beta)^{\mathrm{T}}(K^{\mathrm{T}})^{-1}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}WK^{-1}\}\mu^{(3)}\\
&\quad+\frac{1}{\sigma^4N}\mathrm{tr}\{(K^{\mathrm{T}})^{-1}W^{\mathrm{T}}WK^{-1}\}\mu^{(4)}=\Delta_{N,11}+\mathcal{J}_{N,11}\to\Delta_{11}+\mathcal{J}_{11},\\
\frac{1}{N}\mathrm{E}\bigg\{\frac{\partial\ell(\theta)}{\partial\gamma}\frac{\partial\ell(\theta)}{\partial\gamma}\bigg\}&=\frac{1}{\sigma^2N}\mathbb{X}^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}\mathbb{X}=\Delta_{N,33}+\mathcal{J}_{N,33}\to\Delta_{33}+\mathcal{J}_{33},\\
\frac{1}{N}\mathrm{E}\bigg\{\frac{\partial\ell(\theta)}{\partial\sigma^2}\frac{\partial\ell(\theta)}{\partial\sigma^2}\bigg\}&=-\frac{1}{4\sigma^4}+\frac{1}{4\sigma^8N}\mu^{(4)}=\Delta_{N,44}+\mathcal{J}_{N,44}\to\Delta_{44}+\mathcal{J}_{44},\\
\frac{1}{N}\mathrm{E}\bigg\{\frac{\partial\ell(\theta)}{\partial\rho}\frac{\partial\ell(\theta)}{\partial\gamma}\bigg\}&=\frac{1}{N}\mathrm{tr}(WK^{-1}S^{-1}W)-\frac{2}{N}\mathrm{tr}\{W^{\mathrm{T}}(K^{\mathrm{T}})^{-1}S^{-1}W\}\\
&\quad+\frac{1}{\sigma^4N}\mathrm{tr}\{(\mathbb{X}\beta)^{\mathrm{T}}(K^{\mathrm{T}})^{-1}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}\}\mu^{(3)}\\
&\quad+\frac{1}{\sigma^4N}\mathrm{tr}\{(K^{\mathrm{T}})^{-1}W^{\mathrm{T}}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}\}\mu^{(4)}=\Delta_{N,12}+\mathcal{J}_{N,12}\to\Delta_{12}+\mathcal{J}_{12},\\
\frac{1}{N}\mathrm{E}\bigg\{\frac{\partial\ell(\theta)}{\partial\rho}\frac{\partial\ell(\theta)}{\partial\beta^{\mathrm{T}}}\bigg\}&=\frac{1}{\sigma^2N}\mathbb{X}^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}WK^{-1}\mathbb{X}\beta+\frac{1}{\sigma^4N}\mathbb{X}^{\mathrm{T}}(S^{\mathrm{T}})^{-1}WK^{-1}\mu^{(3)}
\end{aligned}$$

$$
\begin{aligned}
&= \Delta_{N,13} + \mathcal{J}_{N,13} \to \Delta_{13} + \mathcal{J}_{13},\\
\frac{1}{N}\mathrm{E}\left\{\frac{\partial \ell(\theta)}{\partial \rho}\frac{\partial \ell(\theta)}{\partial \sigma^2}\right\} &= \frac{1}{2\sigma^4 N}\mathrm{tr}(\mathbb{X}\beta K^{-1}WS^{-1})\mu^{(3)} + \frac{1}{2\sigma^4 N}\mathrm{tr}(K^{-1}W)\mu^{(4)}\\
&= \Delta_{N,14} + \mathcal{J}_{N,14} \to \Delta_{14} + \mathcal{J}_{14},\\
\frac{1}{N}\mathrm{E}\left\{\frac{\partial \ell(\theta)}{\partial \gamma}\frac{\partial \ell(\theta)}{\partial \beta^{\mathrm{T}}}\right\} &= \frac{1}{\sigma^4 N}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}\mathbb{X}\mu^{(3)} = \Delta_{N,23} + \mathcal{J}_{N,23} \to \Delta_{23} + \mathcal{J}_{23},\\
\frac{1}{N}\mathrm{E}\left\{\frac{\partial \ell(\theta)}{\partial \gamma}\frac{\partial \ell(\theta)}{\partial \sigma^2}\right\} &= \frac{1}{2\sigma^6 N}\mathrm{tr}(S^{-1}W)\mu^{(4)} = \Delta_{N,24} + \mathcal{J}_{N,24} \to \Delta_{24} + \mathcal{J}_{24},\\
\frac{1}{N}\mathrm{E}\left\{\frac{\partial \ell(\theta)}{\partial \sigma^2}\frac{\partial \ell(\theta)}{\partial \beta^{\mathrm{T}}}\right\} &= \frac{2}{\sigma^2 N}\mathrm{E}\{\mathbb{X}^{\mathrm{T}}(S^{\mathrm{T}})^{-1}\mathcal{E}\} = 0 = \Delta_{N,34} + \mathcal{J}_{N,34} \to \Delta_{34} + \mathcal{J}_{34},
\end{aligned}
$$

where the convergence is due to Condition (C4). We then have

$$N^{-1}\mathrm{cov}\{\partial \ell(\theta)/\partial\theta\} = \Delta_N + \mathcal{J}_N \to \Delta + \mathcal{J},$$

and Condition (D) of Lemma A.2 is satisfied. In summary, the equation

$$N^{-1/2}R = N^{-1/2}t^{\mathrm{T}}Q(t) \to_d N(0, t^{\mathrm{T}}(\Delta + \mathcal{J})t)$$

can be verified. Together with the Cramér-Wold device, it implies that $N^{-1/2}\partial\ell(\theta)/\partial\theta \to_d N(0, \Delta + \mathcal{J})$, which completes the proof of Lemma A.3. □

Lemma A.4. *Define* $\ell^{(2)}(\theta) = \partial^2\ell(\theta)/\partial\theta\partial\theta^{\mathrm{T}}$*, and according to Conditions* (C1)–(C4)*, we have* $-N^{-1}\ell^{(2)}(\theta) \to_p \Delta$ *as* $N \to \infty$*.*

Proof. We first evaluate the second-order derivatives of $\ell(\theta)$ as follows:

$$
\begin{aligned}
\frac{\partial^2 \ell(\theta)}{\partial \rho^2} &= -\mathrm{tr}(WK^{-1}K^{-1}W) - \frac{1}{\sigma^2}\mathbb{Y}^{\mathrm{T}}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}W\mathbb{Y},\\
\frac{\partial^2 \ell(\theta)}{\partial \gamma^2} &= \mathrm{tr}(WS^{-1}S^{-1}W) - \frac{2}{\sigma^2}G^{\mathrm{T}}(S^{\mathrm{T}})^{-1}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}G\\
&\quad - \frac{1}{\sigma^2}G^{\mathrm{T}}(S^{\mathrm{T}})^{-1}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}WS^{-1}G,\\
\frac{\partial^2 \ell(\theta)}{\partial \beta\partial\beta^{\mathrm{T}}} &= -\frac{1}{\sigma^2}\mathbb{X}(S^{\mathrm{T}})^{-1}S^{-1}\mathbb{X}, \quad \frac{\partial^2 \ell(\theta)}{(\partial \sigma^2)^2} = \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}G^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}G,\\
\frac{\partial^2 \ell(\theta)}{\partial \rho\partial\gamma} &= -\frac{1}{\sigma^2}\mathbb{Y}^{\mathrm{T}}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}G - \frac{1}{\sigma^2}\mathbb{Y}^{\mathrm{T}}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}WS^{-1}G,\\
\frac{\partial^2 \ell(\theta)}{\partial \rho\partial\sigma^2} &= -\frac{1}{\sigma^4}\mathbb{Y}^{\mathrm{T}}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}G, \quad \frac{\partial^2 \ell(\theta)}{\partial \gamma\partial\sigma^2} = -\frac{1}{\sigma^4}G^{\mathrm{T}}(S^{\mathrm{T}})^{-1}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}K^{-1}G,\\
\frac{\partial^2 \ell(\theta)}{\partial \rho\partial\beta^{\mathrm{T}}} &= -\frac{1}{\sigma^2}\mathbb{X}^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}W\mathbb{Y}, \quad \frac{\partial^2 \ell(\theta)}{\partial \gamma\partial\beta^{\mathrm{T}}} = -\frac{2}{\sigma^2}\mathbb{X}^{\mathrm{T}}(S^{\mathrm{T}})^{-1}W^{\mathrm{T}}(S^{\mathrm{T}})^{-1}S^{-1}G,\\
\frac{\partial^2 \ell(\theta)}{\partial \sigma^2\partial\beta^{\mathrm{T}}} &= -\frac{1}{\sigma^4}(S^{-1}\mathbb{X})^{\mathrm{T}}S^{-1}G.
\end{aligned}
$$

Note that $\mathrm{E}(\mathcal{E}) = 0$ and $\mathrm{E}(\mathcal{E}\mathcal{E}^{\mathrm{T}}) = \sigma^2 I$. Then $-N^{-1}\mathrm{E}\{\ell^{(2)}(\theta)\} = \Delta_N$ can be verified. Lastly, we prove $-N^{-1}\ell^{(2)}(\theta) = \Delta_N + o_p(1)$. First, we prove $-N^{-1}\ell^{(2)}(\theta) = -N^{-1}\mathrm{E}\{\ell^{(2)}(\theta)\} + o_p(1)$. After an algebraic calculation, we have

$$
\begin{aligned}
-N^{-1}\frac{\partial^2 \ell(\theta)}{\partial \rho^2} &= -\frac{1}{N}\mathrm{tr}\{W^{\mathrm{T}}(K^{\mathrm{T}})^{-1}K^{-1}W\} + \frac{1}{N\sigma^2}\{\varepsilon^{\mathrm{T}}W^{\mathrm{T}}(K^{\mathrm{T}})^{-1}K^{-1}W\varepsilon\}\\
&\quad + \frac{2}{N\sigma^2}\{\beta^{\mathrm{T}}\mathbb{X}^{\mathrm{T}}W^{\mathrm{T}}(K^{\mathrm{T}})^{-1}(S^{\mathrm{T}})^{-1}K^{-1}W\varepsilon\} - \frac{1}{N}\mathrm{E}\left\{\frac{\partial^2 \ell(\theta)}{\partial \rho^2}\right\}.
\end{aligned}
$$

Let $H = W^{\mathrm{T}}(K^{\mathrm{T}})^{-1}K^{-1}W$ and $B = \beta^{\mathrm{T}}\mathbb{X}^{\mathrm{T}}W^{\mathrm{T}}(K^{\mathrm{T}})^{-1}(S^{\mathrm{T}})^{-1}K^{-1}W$. According to Lemma A.1 and Conditions (C2) and (C3), we have

$$\sup_{N\geqslant 1}\|W^{\mathrm{T}}(K^{\mathrm{T}})^{-1}K^{-1}W\|_1 < \infty$$

and

$$\sup_{N\geqslant 1} N^{-1}\|\mathrm{vec}\, B\|_{2+\eta_1}^{2+\eta_1} < \infty.$$

Thus, Conditions (A) and (B) of Lemma A.2 hold. Accordingly,

$$-N^{-1}\partial^2\ell(\theta)/\partial\rho^2 = -N^{-1}\mathrm{E}\{\partial^2\ell(\theta)/\partial\rho^2\} + o_p(1)$$

is verified by Lemma A.2. Following the same logic, we can show that

$$-N^{-1}\ell^{(2)}(\theta) = -N^{-1}\mathrm{E}\{\ell^{(2)}(\theta)\} + o_p(1).$$

Consequently, the variance of each component of $N^{-1}\ell^{(2)}(\theta)$ is $O(N^{-1})$. Combining this result with Condition (C4), we have $-N^{-1}\ell^{(2)}(\theta) = \Delta_N + o_p(1) \to_p \Delta$. This completes the proof of Lemma A.4. □

Appendix B Proof of Theorem 3.1

Theorem 3.1 is to be proved in the following two steps. In the first step, we show that $\hat\theta$ is $\sqrt{N}$-consistent. In the second step, we verify that $\hat\theta$ is asymptotically normal.

Step 1 (Consistency)**.** Here, we want to prove that $\hat\theta$ is $\sqrt{N}$-consistent, i.e., there is a local optimizer $\hat\theta$ such that $\|\hat\theta - \theta\| = O_p(N^{-1/2})$. Following Fan and Li [12], it suffices to show that for any $\epsilon > 0$, there is a finite constant $C > 0$ such that

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

To this end, we employ Taylor's expansion and obtain

$$R(\theta) = \ell(\theta + N^{-1/2}u) - \ell(\theta) = N^{-1/2}u^{\mathrm{T}}\frac{\partial\ell(\theta)}{\partial\theta} + (2N)^{-1}u^{\mathrm{T}}\frac{\partial^2\ell(\theta)}{\partial\theta\partial\theta^{\mathrm{T}}}u + o_p(1). \tag{B.2}$$

According to Lemma A.3, we know that $N^{-1/2}\partial\ell(\theta)/\partial\theta = O_p(1)$. In addition, according to Lemma A.4, we have $N^{-1}\partial^2\ell(\theta)/\partial\theta\partial\theta^{\mathrm{T}} = -\Delta_N + o_p(1) \to -\Delta$. Because of $\|u\| = C$, the first term in (B.2) is uniformly bounded by $C\|N^{-1/2}\partial\ell(\theta)/\partial\theta\|$, which is linear in C with the coefficient $\|N^{-1/2}\partial\ell(\theta)/\partial\theta\| = O_p(1)$. On the other hand, the second term in (B.2) is uniformly larger than $(2N)^{-1}\lambda_{\min}(\Delta_N)C^2 \to_p (2N)^{-1}\lambda_{\min}(\Delta)C^2$, where $\lambda_{\min}(M)$ refers to the minimal eigenvalue of any generic matrix M. As a result, with the probability tending to one, the second term in (B.2) is uniformly larger than $0.5\lambda_{\min}\Delta C^2$, which is quadratic in C. Therefore, as long as C is sufficiently large, the second term would dominate the first term. This completes the first part of the proof.

Step 2 (Normality)**.** Here, we are going to show that $\hat\theta$ is asymptotically normal. We take a Taylor's expansion of the estimation equation $\partial\ell(\hat\theta)/\partial\theta = 0$ at the true value θ, leading to

$$\partial\ell(\hat\theta)/\partial\theta = \partial\ell(\theta)/\partial\theta + \partial^2\ell(\theta)/\partial\theta\partial\theta^{\mathrm{T}}(\hat\theta - \theta)\{1 + o_p(1)\} = 0.$$

We then have

$$\sqrt{N}(\hat\theta - \theta) = \bigg\{-\frac{1}{N}\frac{\partial^2\ell(\theta)}{\partial\theta\partial\theta^{\mathrm{T}}}\bigg\}^{-1}\frac{1}{\sqrt{N}}\frac{\partial\ell(\theta)}{\partial\theta}\{1 + o_p(1)\}.$$

According to Slutsky's theorem and Lemmas A.3 and A.4, we then have

$$\sqrt{N}(\hat\theta - \theta) = \frac{1}{\sqrt{N}}\bigg\{\frac{d^2\ell(\theta)}{d\theta d\theta^{\mathrm{T}}}\bigg\}^{-1}\frac{d\ell(\theta)}{d\theta} + o_p(1) \to_d N(0, \Delta^{-1} + \Delta^{-1}\mathcal{J}\Delta^{-1}), \tag{B.3}$$

which completes the entire proof of Theorem 3.1.

Appendix C Proof of Theorem 3.2

In the proof of Theorem 3.2, we only present the proof for T_g, since the results for T_s and T_b can be proved in a similar manner. Under the null hypothesis of $H_0 : \widetilde{\theta} = 0$, we have the constrained log-likelihood as $\ell_g(\theta) = -N/2\log(\sigma^2) - (2\sigma^2)^{-1}\mathbb{Y}^{\mathrm{T}}\mathbb{Y}$. We can obtain the constrained QMLE of θ as $\hat{\theta}_g = \arg\max_\theta \ell_g(\theta) = (0, 0, \mathbf{0}^{\mathrm{T}}, \hat{\sigma})^{\mathrm{T}}$ with $\hat{\sigma} = N^{-1}\mathbb{Y}^{\mathrm{T}}\mathbb{Y}$. Then following the same logic in the proof of (B.3), we obtain

$$N^{1/2}(\hat{\theta}_g - \theta) = N^{-1/2}\Delta_g^{-1}(\theta)\frac{\partial \ell(\theta)}{\partial\theta} + o_p(1), \tag{C.1}$$

where Δ_g is the information matrix of $\ell_g(\theta)$ and we have

$$\Delta_g = \begin{pmatrix} 0_{(p+2)\times(p+2)} & 0_{(p+2)\times 1} \\ 0_{1\times(p+2)} & \Delta_{44} \end{pmatrix}$$

with $\Delta_{44} = \sigma^{-4}/2$. Combining this result with Theorem 3.1, we have proven that both $\hat{\theta}$ and $\hat{\theta}_g$ are $\sqrt{N}$-consistent. According to (B.3) and (C.1), we can obtain

$$N^{1/2}(\hat{\theta}_g - \hat{\theta}) = N^{-1/2}\big(\Delta_g^{-1} - \Delta^{-1}\big)\frac{\partial \ell(\theta)}{\partial\theta} + o_p(1) = O_p(1). \tag{C.2}$$

Applying Taylor's expansion, we have

$$T_g = -2\{\ell(\hat{\theta}_g) - \ell(\hat{\theta})\} = (\hat{\theta}_g - \hat{\theta})^{\mathrm{T}}\frac{\partial^2 \ell(\breve{\theta})}{\partial\theta\partial\theta^{\mathrm{T}}}(\hat{\theta}_g - \hat{\theta}),$$

where $\breve{\theta}$ lies between $\hat{\theta}$ and $\hat{\theta}_g$ and it is also $\sqrt{N}$-consistent. According to Conditions (C1)–(C4) and following the same logic in the proof of Lemma A.4, we have $-N^{-1}\{\partial^2\ell(\breve{\theta})\}/(\partial\theta\partial\theta^{\mathrm{T}}) \to_p \Delta$. Combining the above result with (C.2), we have

$$\begin{aligned} T_g &= N^{-1/2}(\hat{\theta}_g - \hat{\theta})^{\mathrm{T}}\Delta N^{-1/2}(\hat{\theta}_g - \hat{\theta}) + o_p(1) \\ &= \left\{N^{-1/2}(\Delta_g^{-1} - \Delta^{-1})\frac{\partial\ell(\theta)}{\partial\theta}\right\}^{\mathrm{T}}\Delta\left\{N^{-1/2}(\Delta_g^{-1} - \Delta^{-1})\frac{\partial\ell(\theta)}{\partial\theta}\right\} + o_p(1) \\ &= \left\{N^{-1/2}(\Delta + \mathcal{J})^{-1/2}\frac{\partial\ell(\theta)}{\partial\theta}\right\}^{\mathrm{T}}(\Delta + \mathcal{J})^{1/2}(\Delta_g^{-1} - \Delta^{-1})\Delta \\ &\quad \times (\Delta_g^{-1} - \Delta^{-1})(\Delta + \mathcal{J})^{1/2}\left\{N^{-1/2}(\Delta + \mathcal{J})^{-1/2}\frac{\partial\ell(\theta)}{\partial\theta}\right\} + o_p(1). \end{aligned}$$

We can verify that $\Delta_g^{-1}\Delta\Delta_g^{-1} = \Delta_g^{-1}$ and $(\Delta_g^{-1} - \Delta^{-1})^{\mathrm{T}}\Delta(\Delta_g^{-1} - \Delta^{-1}) = \Delta_g^{-1} - \Delta$, and we further obtain

$$\begin{aligned} T_g = &\left\{N^{-1/2}(\Delta + \mathcal{J})^{-1/2}\frac{\partial\ell(\theta)}{\partial\theta}\right\}^{\mathrm{T}}(\Delta + \mathcal{J})^{1/2}(\Delta_g^{-1} - \Delta) \\ &\times (\Delta + \mathcal{J})^{1/2}\left\{N^{-1/2}(\Delta + \mathcal{J})^{-1/2}\frac{\partial\ell(\theta)}{\partial\theta}\right\} + o_p(1). \end{aligned}$$

According to Lemma A.3, we have

$$N^{-1/2}(\Delta + \mathcal{J})^{-1/2}\{\partial\ell(\theta)\}/\partial\theta \to_d N(0, I),$$

where the dimension of I is $p+3$. Let $\lambda_{1,g}(\theta), \dots, \lambda_{p+3,g}$ be the eigenvalues of $(\Delta + \mathcal{J})^{1/2}(\Delta_g^{-1} - \Delta)(\Delta + \mathcal{J})^{1/2}$. According to the continuous mapping theorem and Slutsky's theorem, we obtain

$$T_g \to_d \sum_{i=1}^{p+3}\lambda_{i,g}(\theta, \mu^{(3)}, \mu^{(4)})\chi^2_{i,1}.$$

This completes the first part of the proof.

Under the normal assumption of $\mathcal{E}$, the matrix $\mathcal{J}_N = 0$. According to Condition (C4), we have $\mathcal{J} = 0$, which leads to $\Delta + \mathcal{J} = \Delta$. In addition, by using the fact that $(\Delta_g^{-1} - \Delta)^{\mathrm{T}}\Delta(\Delta_g^{-1} - \Delta) = \Delta_g^{-1} - \Delta$, the symmetric matrix $\Delta^{1/2}(\Delta_g^{-1} - \Delta)\Delta^{1/2}$ is idempotent. Let $\Delta = (\Delta_1, \Delta_2, \Delta_3, \Delta_4)$, where Δ_i $(i = 1, 2, 3, 4)$ correspond to the blocks of Δ_g^{-1}. Then we have

$$\mathrm{tr}\{\Delta^{1/2}(\Delta_g^{-1} - \Delta)\Delta^{1/2}\} = \mathrm{tr}\{(\Delta_g^{-1} - \Delta)\Delta\} = \mathrm{tr}(I - \Delta_g^{-1}\Delta) = \mathrm{tr}\left\{\begin{pmatrix} I & 0_{(p+2)\times 1} \\ -\Delta_{44}^{-1}\Delta_3 & 0 \end{pmatrix}\right\} = p + 2.$$

Combining the above results, we have $T_g \to_d \chi^2_{p+2}$, which completes the entire proof.

Appendix D Proof of Theorem 3.3

In this section, we only present the proof of $\mathcal{B}_g$, while the results for $\mathcal{B}_s$ and $\mathcal{B}_b$ can be proved in a similar manner. Let the true value under the local alternative be

$$\theta_{g0} = (\rho_g/\sqrt{N}, \gamma_g/\sqrt{N}, \beta_g^{\mathrm{T}}/\sqrt{N}, \sigma^2)^{\mathrm{T}},$$

the estimator under the null alternative be $\hat{\theta}_g = (0, 0, \mathbf{0}^{\mathrm{T}}, \hat{\sigma})^{\mathrm{T}}$ with $\hat{\sigma} = N^{-1}\mathbb{Y}^{\mathrm{T}}\mathbb{Y}$, and the global quasi-maximum likelihood estimator be $\hat{\theta} = \arg\max_\theta \ell(\theta)$. We take a Taylor's expansion of $\partial\ell(\hat{\theta}_g)/\partial\theta = 0$ at $\theta_{g0} - \theta_{gN}$, where $\theta_{gN} = \theta_g/\sqrt{N} = (\widetilde{\theta}_g^{\mathrm{T}}/\sqrt{N}, 0)^{\mathrm{T}}$ and $\theta_g = (\rho_g, \gamma_g, \beta_g^{\mathrm{T}}, 0)^{\mathrm{T}}$. This leads to

$$\partial\ell(\hat{\theta}_g)/\partial\theta = \frac{\partial\ell(\theta_{g0} - \theta_{gN})}{\partial\theta} + \frac{\partial^2\ell(\breve{\theta}_g)}{\partial\theta\partial\theta^{\mathrm{T}}}(\hat{\theta}_g - (\theta_{g0} - \theta_{gN})),$$

where $\breve{\theta}_g$ lies between $\hat{\theta}_g$ and $\theta_{g0} - \theta_{gN}$. We then have

$$\sqrt{N}(\hat{\theta}_g - \theta_{g0}) = \left\{-\frac{1}{N}\frac{\partial^2\ell(\breve{\theta}_g)}{\partial\theta\partial\theta^{\mathrm{T}}}\right\}^{-1}\frac{1}{\sqrt{N}}\frac{\partial\ell(\theta_{g0} - \theta_{gN})}{\partial\theta} - \sqrt{N}\theta_{gN},$$
$$\sqrt{N}(\hat{\theta}_g - \hat{\theta}) = N^{-1/2}(\Delta_g^{-1} - \Delta^{-1})\frac{\partial\ell(\theta_{g0})}{\partial\theta} - \sqrt{N}\theta_{gN} + o_p(1).$$

By Theorem 3.1 and Lemma A.4, we have

$$\sqrt{N}(\hat{\theta} - \theta_{g0}) \to_d N(0, \Delta^{-1} + \Delta^{-1}\mathcal{J}\Delta^{-1})$$

and

$$-N^{-1}\{\partial^2\ell(\breve{\theta})\}/(\partial\theta\partial\theta^{\mathrm{T}}) \to_p \Delta.$$

Applying Taylor's expansion and $\theta_{gN}^{\mathrm{T}} \to 0$ as $N \to \infty$, we have

$$\begin{aligned} T_g &= -2\{\ell(\hat{\theta}_g) - \ell(\hat{\theta})\} = (\hat{\theta}_g - \hat{\theta})^{\mathrm{T}}\frac{\partial^2\ell(\breve{\theta})}{\partial\theta\partial\theta^{\mathrm{T}}}(\hat{\theta}_g - \hat{\theta}) \\ &= \left\{N^{-1/2}(\Delta + \mathcal{J})^{-1/2}\frac{\partial\ell(\theta)}{\partial\theta}\right\}^{\mathrm{T}}(\Delta + \mathcal{J})^{1/2}(\Delta_g^{-1} - \Delta)(\Delta + \mathcal{J})^{1/2} \\ &\quad \times \left\{N^{-1/2}(\Delta + \mathcal{J})^{-1/2}\frac{\partial\ell(\theta)}{\partial\theta}\right\} + \theta_g^{\mathrm{T}}\Delta\theta_g + o_p(1). \end{aligned}$$

According to Slutsky's theorem, we obtain

$$T_g \to_d \sum_{i=1}^{p+3}\lambda_{i,g}\chi^2_{i,1} + \sqrt{N}\theta_{gN}^{\mathrm{T}}\Delta\sqrt{N}\theta_{gN}$$

under the local alternative setting, where $\lambda_{1,g}, \ldots, \lambda_{p+3,g}$ are the eigenvalues of

$$(\Delta + \mathcal{J})^{1/2}(\Delta_g^{-1} - \Delta)(\Delta + \mathcal{J})^{1/2}.$$

The power function is defined as

$$\begin{aligned}\mathcal{B}_g &= \mathrm{P}(T_g > \chi^2_{\mathrm{weight},g}(1-\alpha) \mid H_1) = \mathrm{P}\bigg(\sum_{i=1}^{p+3}\lambda_{i,g}\chi^2_{i,1} + \theta_g^{\mathrm{T}}\Delta\theta_g > \chi^2_{\mathrm{weight},g}(1-\alpha)\bigg)\\ &= \mathrm{P}\bigg(\bigg|\sum_{i=1}^{p+3}\lambda_{i,g}\chi^2_{i,1}\bigg| > \chi^2_{\mathrm{weight},g}(1-\alpha) - \theta_g^{\mathrm{T}}\Delta\theta_g\bigg)\\ &= 1 - F_g\{\chi^2_{\mathrm{weight},g}(1-\alpha) - \theta_g^{\mathrm{T}}\Delta\theta_g\},\end{aligned}$$

where $\chi^2_{\mathrm{weight},g}(1-\alpha)$ is the $1-\alpha$ theoretical quantile of the weighted chi-square $\sum_{i=1}^{p+3}\lambda_{i,g}\chi^2_{i,1}$, $F_g(\cdot)$ is the cumulative distribution function of the weight chi-square distribution $\sum_{i=1}^{p+3}\lambda_{i,g}\chi^2_{i,1}$ and $\theta_g = (\rho_g, \gamma_g, \beta_g^{\mathrm{T}}, 0)^{\mathrm{T}}$. Then we can prove that

$$\lim_{N\to\infty}\mathcal{B}_g = 1 - F_g\{\chi^2_{\mathrm{weight},g}(1-\alpha) - \theta_g^{\mathrm{T}}\Delta\theta_g\}.$$