Statistica Sinica Preprint No: SS-2022-0035	
Title	Mutual Influence Regression Model
Manuscript ID	SS-2022-0035
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0035
Complete List of Authors	Xinyan Fan, Wei Lan, Tao Zou and Chih-Ling Tsai
Corresponding Authors	Tao Zou
E-mails	tao.zou@anu.edu.au
Notice: Accepted version subject to English editing.	

Statistica Sinica

Mutual Influence Regression Model

Xinyan Fan, Wei Lan*, Tao Zou* and Chih-Ling Tsai

Renmin University of China, Southwestern University of Finance and Economics, The Australian National University and University of California, Davis

Abstract: In this article, we propose the mutual influence regression model (MIR) to establish the relationship between the mutual influence matrix of actors and a set of similarity matrices induced by their associated attributes. This model is able to explain the heterogeneous structure of the mutual influence matrix by extending the commonly used spatial autoregressive model while allowing it to change with time. To facilitate making inferences with MIR, we establish parameter estimation, weight matrices selection and model testing. Specifically, we employ the quasi-maximum likelihood estimation method to estimate unknown regression coefficients, and demonstrate that the resulting estimator is asymptotically normal without imposing the normality assumption and while allowing the number of similarity matrices to diverge. In addition, an extended BIC-type criterion is introduced for selecting relevant matrices from the divergent number

Corresponding author: Tao Zou, The Australian National University, Canberra, ACT 2600, Australia. E-mail: tao.zou@anu.edu.au. Wei Lan, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China. E-mail: lanwei@swufe.edu.cn.

of similarity matrices. To assess the adequacy of the proposed model, we further propose an influence matrix test and develop a novel approach in order to obtain the limiting distribution of the test. The simulation studies support our theoretical findings, and a real example is presented to illustrate the usefulness of the proposed MIR model.

Key words and phrases: Extended Bayesian Information Criterion, Mutual Influence Matrix, Similarity Matrices, Spatial Autoregressive Model

1. Introduction

Due to the possibility of relationships between subjects (such as network connections or spatial interactions), the traditional data assumption of independent and identically distributed observations is no longer valid, and there can be a complex structure of mutual influence between the subjects. Accordingly, understanding mutual influence has become an important topic across various fields and applications such as business, biology, economics, medicine, sociology, political science, psychology, engineering, and science. For example, the study of the mutual influence between actors can help to identify influential users within a network (see Trusov, Bodapati and Bucklin (2010)). In addition, investigating the mutual influence between geographic regions is essential for exploring spillover effects in spatial data (see Golgher and Voss (2016); Zhang and Yu (2018)), and this type of analysis is important for understanding the spread of COVID-19 between different countries and cities (see Han et al. (2021)). Moreover, quantifying mutual influence in mobile social networks is helpful to provide important insights into the design of social platforms and applications (see Peng et al. (2017)). These examples motivate us to introduce the mutual influence regression model so that we are able to effectively and systematically study mutual influence.

Let $Y_{1t}, \cdots, Y_{nt}$ be the responses of n actors observed at time t for $t = 1, \cdots, T$. To characterize the mutual influence among the n actors, the following regression model can be considered for each actor $i = 1, \cdots, n$ at $t = 1, \cdots, T$,

$$Y_{it} = b_{i1t}Y_{1t} + \cdots + b_{i(i-1)t}Y_{(i-1)t} + b_{i(i+1)}Y_{(i+1)t} + \cdots + b_{int}Y_{nt} + \epsilon_{it}, \quad (1.1)$$

where b_{ijt} presents the influence effect of Y_{jt} on Y_{it} and ϵ_{it} is the random noise. Define $Y_t = (Y_{1t}, \cdots, Y_{nt})^\top \in \mathbb{R}^n$, $\epsilon_t = (\epsilon_{1t}, \cdots, \epsilon_{nt})^\top \in \mathbb{R}^n$ and $B_t = (b_{ijt}) \in \mathbb{R}^{n\times n}$ with $b_{iit} = 0$. Then we have the matrix form of (1.1),

$$Y_t = B_tY_t + \epsilon_t, \quad (1.2)$$

where B_t is called the mutual influence matrix and it characterizes the degree of mutual influence among the n actors at time t.

Estimating model (1.2) is a challenging task since it involves a large number of parameters, specifically $n(n-1)$ for each t. The regularization type methods studied by Manresa (2013), de Paula et al. (2019) and Kwok (2020) are not applicable when n is large. To avoid the issue of high dimensionality, one commonly used approach is to employ the spatial autoregressive (SAR) model, which parameterizes the mutual influence matrix B_t by $B_t = \rho W^{(t)}$, where $W^{(t)}$ is the adjacency matrix of a known network or a spatial weight matrix whose elements are a function of geographic or economic distances. In addition, ρ is the single influence parameter that characterizes the influence power among the n actors; see, for example, Lee (2004), Zou et al. (2017) and Huang et al. (2019) for detailed discussions and the references therein. Accordingly, model (1.2) becomes estimable since the number of parameters is greatly reduced from $n(n-1)$ to 1.

Because the SAR model only involves a single influence parameter ρ, it may not fully capture the influential information of B_t. Hence, Lee and Liu (2010), Elhorst, Lacombe and Piras (2012), Lee and Yu (2014), Kwok (2019), and Lam and Souza (2020) considered a higher-order SAR model that includes multiple weight matrices (i.e., $W^{(t)}$s) along with their associated parameters. Gupta and Robinson (2015, 2018) further extended it by allowing the number of weight matrices to diverge. In general, the elements of weight matrix $W^{(t)}$ are functions of the geographic or economic distances among the n actors. For example, a typical choice of distance measure for spatial data is geographic distance (Dou, Parrella and Yao (2016); Zhang and Yu (2018); Gao et al. (2019)). In addition, one natural choice of distance measure for network data is whether there exists a link between the actors through the adjacency matrix (Zhou et al. (2017); Zhu et al. (2017); Huang et al. (2019)). However, the above weight settings cannot be directly applied to the higher-order SAR model for non-geographic or non-network data since these distance measure are not well defined for other types of data. Accordingly, how to parameterize the mutual influence matrix for non-geographic and non-network data is an unsolved problem that needs further investigation. This motivates us to study the following two important and challenging subjects: (i) How to define weight matrices for general non-geographic and non-network data? (ii) How to assess the adequacy of the selected weight matrices?

To resolve challenge (i), we propose using similarity matrices induced from attributes (e.g., gender or income) to be our weight matrices to accommodate non-geographic and non-network data. Specifically, let $\mathbf{Z}^{(t)} = (z_1^{(t)}, \cdots, z_n^{(t)})^\top \in \mathbb{R}^n$ denote the vector of values obtained from the n actors for a given attribute. Then, for any two actors j_1 and j_2, the squared distance between j_1 and j_2 can be defined as the distance between $z_{j_1}^{(t)}$ and $z_{j_2}^{(t)}$, e.g., $(z_{j_1}^{(t)} - z_{j_2}^{(t)})^2$. Following the suggestion of Jenish and Prucha (2012), we consider the similarity matrix as a non-increasing function of the squared distance between actors j_1 and j_2, i.e., $A^{(t)} = (a\{-(z_{j_1}^{(t)} - z_{j_2}^{(t)})^2\})_{n\times n}$ for some bounded and non-decreasing function $a(\cdot)$. Furthermore, we can employ the same procedure to create a set of similarity matrices $A^{(t)}$s deriving from the actors' attributes. In practice, those similarity matrices change along with time t. To this end, we introduce the time heterogeneous matrices, $A^{(t)}$s, which naturally link to the mutual influence matrix B_t. To overcome the aforementioned challenge (ii), we introduce an influence matrix test to examine the adequacy of the selected similarity matrices (i.e., weight matrices) for the high dimensional and time varying mutual influence matrix.

The main contribution of this paper is two-fold. The first is to propose a mutual influence regression (MIR) model that establishes a relationship between the mutual influence matrix and a set of similarity matrices induced by associated attributes of the actors. The emerging model not only broadens the usefulness of the traditional spatial autoregressive model, but also captures the heterogeneous structure of the mutual influence matrix by allowing it to change with time. Accordingly, we study the parameter space of the model and then employ the quasi-maximum likelihood estimation method (see, e.g., Wooldridge (2002)) to estimate unknown regression coefficients. By thoroughly studying the convergence of the Hessian matrix in Frobenius norm, we are able to show that the resulting estimator is asymptotically normal under some mild conditions without imposing the normality assumption while allowing the number of similarity matrices to diverge. Since the number of similarity matrices is diverging, an extended BIC-type criterion motivated from Chen and Chen (2008) is introduced to select relevant matrices. We show that this extended BIC-type criterion is consistent based on a novel result of the exponential tail probability for the general form of quadratic functions.

The second is to introduce an influence matrix test for assessing whether the mutual influence matrix B_t satisfies a linear structure of the time-varying weight matrices. Based on this setting, $\mathrm{cov}(Y_t)$ is a nonlinear function of the time-varying weight matrices. Thus, our test is different from the common hypothesis test for testing whether $\mathrm{cov}(Y_t)$ is a linear structure of the weight matrices (e.g., see Zheng et al. (2019)). Under a nonlinear structure for the mutual influence matrix B_t, however, the quasi-maximum likelihood estimators of regression coefficients can result in a larger variance in the test statistic. As a result, obtaining the asymptotic distribution of the test statistic becomes a challenging task, especially when the number of similarity matrices is diverging. To overcome such difficulties, we develop a novel approach in order to show the asymptotic normality of a summation of the product of quadratic forms with a diverging number of similarity matrices.

The remainder of this article is organized as follows. Section 2 introduces the mutual influence regression model, studies the parameter space, and obtains quasi-maximum likelihood estimators of regression coefficients, which are asymptotically normal. Section 3 presents the extended BIC-type selection criterion as well as its consistency property. In addition, a high dimensional covariance test is given to examine the model adequacy. The theoretical property of this test is provided. Simulation studies and an empirical example are presented in Sections 4 and 5, respectively, while Section 6 concludes the article with a discussion. All theoretical proofs are relegated to the supplementary material.

2. Mutual Influence Regression Model and Estimation

2.1 Model and Notation

We first construct similarity matrices before modeling the mutual influence matrix B_t as a regression function of them. Let $Z_k^{(t)}$ be the k-th $n \times 1$ continuous attribute vector collected at the t-th time for $k = 1, \cdots, d$.

Adapting Jenish and Prucha's (2012) approach in order to incorporate the time effect t, we then obtain heterogeneous similarity matrices: $A_k^{(t)} = A_k^{(t)}(Z_k^{(t)}) = (a\{-(Z_{kj_1}^{(t)} - Z_{kj_2}^{(t)})^2\})_{n\times n}$ for $j_1 = 1, \cdots, n$ and $j_2 = 1, \cdots, n$, where $a(\cdot)$ is a bounded and non-decreasing function and $Z_{kj_1}^{(t)}$ and $Z_{kj_2}^{(t)}$ are the j_1-th and j_2-th elements of $Z_k^{(t)}$, respectively. For continuous attributes, we consider $a(\cdot)$ equal to the exponential function with $a\{-(Z_{kj_1}^{(t)} - Z_{kj_2}^{(t)})^2\} = \exp\{-(Z_{kj_1}^{(t)} - Z_{kj_2}^{(t)})^2\}$ when $|Z_{kj_1}^{(t)} - Z_{kj_2}^{(t)}| < \phi_k^{(t)}$ for some pre-specified positive constant $\phi_k^{(t)}$, and $a\{-(Z_{kj_1}^{(t)} - Z_{kj_2}^{(t)})^2\} = 0$ otherwise. That is, once the distance between any two actors measured by their associated attributes in $Z_k^{(t)}$ exceeds a threshold, the two actors are not mutually influenced. For discrete attributes $Z_k^{(t)}$, we define $a(Z_{kj_1}^{(t)}, Z_{kj_2}^{(t)}) = 1$ if $Z_{kj_1}^{(t)}$ and $Z_{kj_2}^{(t)}$ belong to the same class, and $a(Z_{kj_1}^{(t)}, Z_{kj_2}^{(t)}) = 0$ otherwise. In this case, $A_k^{(t)}$ can be regarded as the adjacency matrix of the network induced by attributes $Z_k^{(t)}$.

To establish the relationship between the mutual influence matrix and a set of similarity matrices, motivated from Anderson (1973), Qu, Lindsay and Li (2000) and Zheng et al. (2019), we parameterize the mutual influence matrix B_t as a function of attributes $Z_k^{(t)}$s ($k = 1, \cdots, d$) given below.

$$B_t(\lambda) \triangleq B_t(Z_1^{(t)}, \cdots, Z_d^{(t)}, \lambda) = \lambda_1 W_1^{(t)} + \cdots + \lambda_d W_d^{(t)}, \qquad (2.1)$$

where $w(Z_{kj_1}^{(t)}, Z_{kj_2}^{(t)}) = a(Z_{kj_1}^{(t)}, Z_{kj_2}^{(t)}) / \sum_{j_2} a(Z_{kj_1}^{(t)}, Z_{kj_2}^{(t)})$ and $W_k^{(t)} = (w(Z_{kj_1}^{(t)}, Z_{kj_2}^{(t)}))_{n\times n}$

is the row-normalized version of $A_k^{(t)}$. We name $W_k^{(t)}$ as the weight matrix for $k = 1, \cdots, d$, which is also called the similarity matrix in the rest of the article. The reason for adopting the row-normalization method is primarily its wide applicability (see, e.g., Lee (2004)). In practice, there are several alternative normalization methods that can be considered, such as the column normalization and the normalization based on the maximum absolute row (or column) sum norm; see Kelejian and Prucha (2010) for detailed discussions.

Substituting (2.1) into (1.2), we introduce the following mutual influence regression (MIR) model,

$$Y_t = B_t(Z_1^{(t)}, \cdots, Z_d^{(t)}, \lambda)Y_t + \epsilon_t = \big(\lambda_1 W_1^{(t)} + \cdots + \lambda_d W_d^{(t)}\big)Y_t + \epsilon_t, \quad (2.2)$$

where $\lambda_1, \cdots, \lambda_d$ are unknown regression coefficients. This model is able to explain the structure of the mutual influence matrix B_t at each time t via a set of similarity matrices $W_k^{(t)}$, induced by the covariates $Z_k^{(t)}$ and their associated influence parameter λ_k. For the sake of simplicity, we refer to the above model as MIR in the rest of the paper. To ease notation, we use B_t rather than $B_t(\lambda)$ in the rest of article. Define $\Delta_t(\lambda) = I_n - B_t = I_n - \big(\lambda_1 W_1^{(t)} + \cdots + \lambda_d W_d^{(t)}\big)$, where I_n is the identity matrix of dimension n. Then, model (2.2) leads to $\Delta_t(\lambda)Y_t = \epsilon_t$. To assure (2.2) identifiable, we require that $\Delta_t(\lambda)$ is invertible.

It is worth noting that, for $d = 1$ and $W_1^{(t)} = W$ constructed by network or spatial data, MIR is the classical spatial autoregressive model of LeSage and Pace (2009). Furthermore, by model (2.1), we have $b_{j_1j_2t} = \lambda_1 w(Z_{1j_1}^{(t)}, Z_{1j_2}^{(t)}) + \cdots + \lambda_d w(Z_{dj_1}^{(t)}, Z_{dj_2}^{(t)})$. Accordingly, the influence effect of node j_2 on j_1, $b_{j_1j_2t}$, is the linear combination of similarity matrices at time t. Specifically, for $k = 1, \cdots, d$, the similarity matrix $w(Z_{kj_1}^{(t)}, Z_{kj_2}^{(t)})$ measures the distance between nodes j_1 and j_2, and its effect is determined by the influence parameter λ_k. Suppose $\lambda_k > 0$. Based on the MIR model (2.2), for any two actors j_1 and j_2, the smaller the distance between $Z_{kj_1}^{(t)}$ and $Z_{kj_2}^{(t)}$, the larger the influence effect between Y_{j_1t} and Y_{j_2t}. Therefore, the covariate $Z_k^{(t)}$ yields a positive effect on the mutual influence between responses of the n actors. In sum, models (2.1) and (2.2) link the mutual influence matrix with a large number of exogenous attributes to responses, which can lead to insightful findings and provide practical interpretations.

Remark 1: It is of interest to note that our concept is similar to the covariance tapering of Furrer et al. (2006). For any given $t = 1, \cdots, T$, we follow Furrer et al. (2006) in assuming that Y_{it}, the response of node i, can be affected by the responses of nearby nodes. However, our method differs in the following two aspects. First, for the geographic data considered in Furrer et al. (2006), the distance between nodes is well defined. However, for general non-geographic and non-network data, the "distance" measure has not been clearly defined. Motivated by the concept of the near-epoch dependent (NED) process of Jenish and Prucha (2012), we define the similarity matrices that are induced by the distances between the attributes of different actors. Second, the goals of these two methods are different. The goal of our paper is to establish the relationship between the mutual influence matrix of actors and a set of similarity matrices induced by their associated attributes, whereas Furrer et al. (2006) focused on the interpolation of large spatial datasets.

2.2 Parameter Estimation

In this paper, we assume that ϵ_ts are iid random variables with mean 0 and covariance matrix $\sigma^2 I_n$ for $t = 1, \cdots, T$, where σ^2 is a scaled parameter. By (2.2), we have $Y_t = \Delta_t^{-1}(\lambda)\epsilon_t$. Then $E(Y_t) = 0$ and $\text{Var}(Y_t) \triangleq \Sigma_t = \sigma^2 \Delta_t^{-1}(\lambda)\{\Delta_t^\top(\lambda)\}^{-1}$, and we obtain the quasi-loglikelihood function following Lee (2004),

$$\begin{aligned}\ell(\theta) = & -\frac{nT}{2}\log(2\pi) - \frac{nT}{2}\log(\sigma^2) + \sum_{t=1}^{T}\log|\det(\Delta_t(\lambda))| \\ & -\frac{1}{2\sigma^2}\sum_{t=1}^{T} Y_t^\top \Delta_t^\top(\lambda)\Delta_t(\lambda)Y_t,\end{aligned} \tag{2.3}$$

where $\theta = (\lambda^\top, \sigma^2)^\top$.

We next employ the concentrated quasi-likelihood approach to estimate θ. Specifically, given λ, one can estimate σ^2 by

$$\widehat{\sigma}^2(\lambda) = (nT)^{-1} \sum_t Y_t^\top \Delta_t^\top(\lambda) \Delta_t(\lambda) Y_t.$$

Plugging this into (2.3), the resulting quasi-concentrated log-likelihood function is

$$\ell_c(\lambda) = -\frac{nT}{2}\log(2\pi) - \frac{nT}{2} - \frac{nT}{2}\log\left\{\widehat{\sigma}^2(\lambda)\right\} + \sum_{t=1}^{T} \log|\det(\Delta_t(\lambda))|. \quad (2.4)$$

Accordingly, we obtain the quasi-maximum likelihood estimator of λ, which is $\widehat{\lambda} = \text{argmax}_{\lambda\in\Lambda}\ell_c(\lambda)$ and Λ is the parameter space. To make $\widehat{\lambda}$ estimable, it is necessary to specify the parameter space Λ. Based on model (2.2) and the definition of $\Delta_t(\lambda)$, one should naturally require that, for any $\lambda \in \Lambda$, $\Delta_t(\lambda)$ is invertible. It is worth noting that a sufficient condition for the invertibility of $\Delta_t(\lambda)$ is $\|\sum_{k=1}^d \lambda_k W_k^{(t)}\| < 1$, where $\|\cdot\|$ denotes the L_2 (i.e., spectral) norm. Using the fact that $W_k^{(t)}$ is row-normalized, we have that $\|\sum_{k=1}^d \lambda_k W_k^{(t)}\| \le \max_k \|W_k^{(t)}\| \sum_{k=1}^d |\lambda_k| \le \sum_{k=1}^d |\lambda_k|$. Accordingly, a sufficient condition for the invertibility of $\Delta_t(\lambda)$ is $\sum_{k=1}^d |\lambda_k| < 1$. This leads us to define the parameter space of λ as follows:

$$\Lambda = \left\{\lambda : \sum_{k=1}^{d} |\lambda_k| < 1 - \varsigma\right\},$$

where ς is some sufficiently small positive number. The reason for introducing ς is to ensure that $\sum_{k=1}^d |\lambda_k|$ is away from 1. In practice, we can set

ς to be a small positive number such as 0.01. This specification does not affect parameter estimation as long as $\sum_{k=1}^{d} |\lambda_k|$ is smaller than 1.

Using the assumption of $\sigma^2 > 0$, the parameter space of θ is

$$\Theta = \big\{\theta = (\lambda^\top, \sigma^2)^\top : \lambda \in \Lambda \text{ and } \sigma^2 > 0\big\}.$$

In addition, σ^2 can be estimated by $\widehat{\sigma}^2 = \widehat{\sigma}^2(\widehat{\lambda})$, which leads to the quasi-maximum likelihood estimator (QMLE), i.e., $\widehat{\theta} = (\widehat{\lambda}^\top, \widehat{\sigma}^2)^\top$.

Denote by $\theta_0 = (\lambda_0^\top, \sigma_0^2)^\top$ the unknown true parameter vector, where $\lambda_0 = (\lambda_{01}, \cdots, \lambda_{0d})^\top \in \Lambda$ and $\sigma_0^2 > 0$. By Lemma 3 and Condition (C4) in Section S1 of the supplementary material, the second order derivative matrix of $\ell(\theta)$ is negative definite for sufficiently large nT in a small neighborhood of θ_0. Accordingly, the parameter estimator $\widehat{\theta}$ exists and lies in Θ. To avoid the problem of local optima in computing QMLE, we recommend using a random initialization method (see, e.g., Wang et al. (2022)). Specifically, we generate many randomized initial values and find the solution which yields the maximum value of the objective function. Our simulation results in Section 5 indicate that this algorithm works satisfactorily in various settings. The asymptotic property of $\widehat{\theta}$ is given in the following theorem.

Theorem 1. *Under Conditions (C1)–(C5) in Section S1 of the supplementary material, as* $nT \to \infty$, $(nT/d)^{1/2} D\mathcal{I}(\theta_0)(\widehat{\theta} - \theta_0)$ *is asymptoti-*

cally normal with mean 0 and covariance matrix $G(\theta_0)$, where D is an arbitrary $M \times (d+1)$ matrix with $M < \infty$ satisfying $\|D\| < \infty$ and $d^{-1}D\mathcal{J}(\theta_0)D^\top \to G(\theta_0)$, and $\mathcal{I}(\theta_0)$ and $\mathcal{J}(\theta_0)$ are defined in Condition (C4).

Note that $nT \to \infty$ in the above theorem means that either n or T go to infinity. To make this theorem practically useful, one needs to estimate $\mathcal{I}(\theta_0)$ and $\mathcal{J}(\theta_0)$ consistently. For $k = 1, \cdots, d+1$ and $l = 1, \cdots, d+1$, define $\mathcal{I}_{nT}(\theta_0) = -(nT)^{-1}E\{\frac{\partial^2 \ell(\theta_0)}{\partial\theta\partial\theta^\top}\} \triangleq (\mathcal{I}_{nT,kl}) \in \mathbb{R}^{(d+1)\times(d+1)}$ and $\mathcal{J}_{nT}(\theta_0) = (nT)^{-1}\mathrm{Var}(\frac{\partial \ell(\theta_0)}{\partial\theta}) \triangleq (\mathcal{J}_{nT,kl}) \in \mathbb{R}^{(d+1)\times(d+1)}$. By Condition (C4), it suffices to show that the plug-in estimators $\mathcal{I}_{nT}(\widehat{\theta})$ and $\mathcal{J}_{nT}(\widehat{\theta})$ are consistent of $\mathcal{I}(\theta_0)$ and $\mathcal{J}(\theta_0)$, respectively.

After simple calculation, we have that, for any $k = 1, \cdots, d$ and $l = 1, \cdots, d$,

$$\mathcal{I}_{nT,k(d+1)} \triangleq -(nT)^{-1}E\Big\{\frac{\partial^2 \ell(\theta_0)}{\partial\lambda_k\partial\sigma^2}\Big\} = \frac{1}{nT\sigma^2}\sum_{t=1}^{T} tr(W_k^{(t)}\Delta_t^{-1}(\lambda_0)) = \frac{1}{nT\sigma^2}tr(U_k),$$

$$\mathcal{I}_{nT,kl} \triangleq -(nT)^{-1}E\Big\{\frac{\partial^2 \ell(\theta_0)}{\partial\lambda_k\partial\lambda_l}\Big\} = (nT)^{-1}\sum_{t=1}^{T} tr\{\Delta_t^{-1\top}(\lambda_0)W_k^{(t)\top}W_l^{(t)}\Delta_t^{-1}(\lambda_0)\}$$
$$+(nT)^{-1}\sum_{t=1}^{T} tr\{W_k^{(t)}\Delta_t^{-1}(\lambda_0)W_l^{(t)}\Delta_t^{-1}(\lambda_0)\} = \frac{2}{nT}tr(U_kU_l),$$

where $U_k = \mathrm{diag}\big\{s(W_k^{(1)}\Delta_t^{-1}(\lambda_0)), \cdots, s(W_k^{(T)}\Delta_T^{-1}(\lambda_0))\big\} \in \mathbb{R}^{(nT)\times(nT)}$ and

$s(A) = (A + A^\top)/2$ for any arbitrary matrix A. In addition,

$$\mathcal{I}_{nT,(d+1)(d+1)} \triangleq -(nT)^{-1} E\Big\{\frac{\partial^2 \ell(\theta_0)}{\partial^2 \sigma^2}\Big\} = \frac{1}{2\sigma_0^4}.$$

Using the result $\widehat{\theta} \to_p \theta_0$ in Theorem 1, we have $\mathcal{I}_{nT}(\widehat{\theta}) \to_p \mathcal{I}_{nT}(\theta_0)$. This, together with Condition (C4), implies $\mathcal{I}_{nT}(\widehat{\theta}) \to_p \mathcal{I}(\theta_0)$.

After algebraic calculation, we next obtain that, for any $k = 1, \cdots, d$ and $l = 1, \cdots, d$,

$$\mathcal{J}_{nT,k(d+1)} \triangleq (nT)^{-1}\text{cov}\Big\{\frac{\partial \ell(\theta_0)}{\partial \lambda_k}, \frac{\partial \ell(\theta_0)}{\partial \sigma^2}\Big\} = \frac{1}{2nT\sigma_0^2}\big\{(\mu^{(4)} - 1)tr(U_k)\big\} \quad \text{and}$$

$$\mathcal{J}_{nT,kl} \triangleq (nT)^{-1}\text{cov}\Big\{\frac{\partial \ell(\theta_0)}{\partial \lambda_k}, \frac{\partial \ell(\theta_0)}{\partial \lambda_l}\Big\} = \frac{2}{nT}tr(U_k U_l) + \frac{\mu^{(4)} - 3}{nT}tr(U_k \otimes U_l),$$

where $\mu^{(4)} = E(\epsilon_{it}^4)/\sigma_0^4$ can be estimated by $\widehat{\mu}^{(4)} = (nT)^{-1}\sum_{i=1}^n \sum_{t=1}^T \widehat{\epsilon}_{it}^4/\widehat{\sigma}^4$ with $\widehat{\epsilon}_t = \Delta_t^{-1}(\widehat{\lambda})Y_t$ and $\widehat{\epsilon}_t = (\widehat{\epsilon}_{1t}, \cdots, \widehat{\epsilon}_{nt})^\top$. Furthermore,

$$\mathcal{J}_{nT,(d+1)(d+1)} \triangleq (nT)^{-1}\text{Var}\Big\{\frac{\partial \ell(\theta_0)}{\partial \sigma^2}\Big\} = \frac{1}{4\sigma_0^4}\big\{2 + (\mu^{(4)} - 3)\big\}.$$

As a result, $\mathcal{J}(\theta_0)$ can be consistently estimated by $\mathcal{J}_{nT}(\widehat{\theta})$. In sum, one can practically apply Theorem 1 by replacing $\mathcal{I}(\theta_0)$ and $\mathcal{J}(\theta_0)$ with their corresponding estimators $\mathcal{I}_{nT}(\widehat{\theta})$ and $\mathcal{J}_{nT}(\widehat{\theta})$, respectively.

According to Theorem 1, we are able to assess the significance of λ_{0k}, which allows us to determine the influential similarity matrices, $W_k^{(t)}$, induced by their associated covariates $Z_k^{(t)}$ for $k = 1, \cdots, d$. In addition, based on the estimated $\widehat{\lambda}$, the mutual influence matrix B_t can be estimated by $\widehat{B}_t = \widehat{\lambda}_1 W_1^{(t)} + \cdots + \widehat{\lambda}_d W_d^{(t)}$, whose asymptotic property is given below.

Theorem 2. *Under Conditions (C1)–(C5) in Section S1 of the supplementary material, as* $nT \to \infty$, $\sup_{t\leq T} \|\widehat{B}_t - B_t\| = O_p\{d(nT)^{-1/2}\}$.

The above theorem indicates that the estimated mutual influence matrix $\widehat{B}_t$ is consistent uniformly for any t under the L_2 norm, as either n or T goes to infinity and $d = o\{(nT)^{1/4}\}$ is from Condition (C5). Hence, $\widehat{B}_t$ can be a consistent estimator of B_t even for finite T. After estimating the mutual influence matrix, we next study the selection of similarity matrices and test the fitness of B_t.

3. Similarity Matrix Selection and Influence Matrix Test

3.1 Selection Consistency

In MIR, the number of similarity matrices is diverging, which motivates us to consider the similarity matrix selection. Note that assessing the significance of λ_{0k} separately for $k = 1, \cdots, d$ via Theorem 1 can result in multiple testing problems (see, e.g., Storey et al. 2004 and Fan et al. 2012). In addition, the traditional Bayesian information criterion (BIC) becomes overly liberal when d is diverging as demonstrated by Chen and Chen (2008). Hence, we modify the extended Bayesian information criterion (EBIC) to select similarity matrices. To this end, we define the true model $\mathcal{S}_T = \{k : \lambda_{0k} \neq 0\}$, which consists of all relevant $W_k^{(t)}$s. In addition, let $\mathcal{S}_F = \{1, \cdots, d\}$ denote the full model and $\mathcal{S}$ represent an arbitrary candidate model such that $\mathcal{S} \subset \mathcal{S}_F$. Moreover, let $\widehat{\theta}_{\mathcal{S}} = (\widehat{\theta}_{k,\mathcal{S}} : k \in \mathcal{S})$ be the maximum likelihood estimator of $\theta_{0\mathcal{S}} = (\theta_{0k} : k \in \mathcal{S}) \in \mathbb{R}^{|\mathcal{S}|}$. In practice, the true model $\mathcal{S}_T$ is unknown. Motivated from Chen and Chen (2008), we propose the information criterion given below to select similarity matrices,

$$\text{EBIC}_{\gamma}(\mathcal{S}) = -2\ell(\widehat{\theta}_{\mathcal{S}}) + |\mathcal{S}| \log(nT) + \gamma|\mathcal{S}| \log(d)$$

for some $\gamma > 0$. Based on this criterion, one can select the optimal model, which is $\widehat{\mathcal{S}} = \text{argmin}_{\mathcal{S}} \text{EBIC}_{\gamma}(\mathcal{S})$. It is worth noting that the third term involved in $\text{EBIC}_{\gamma}(\mathcal{S})$ (i.e., $\gamma|\mathcal{S}| \log(d)$) presents the effect of assigning different prior probabilities to candidate models with different number of weight matrices, and the tuning parameter γ characterizes this strength; we refer to Chen and Chen (2008) for more detailed discussions.

Define as $\mathbb{A}_0 = \{\mathcal{S} : \mathcal{S}_T \subset \mathcal{S}, |\mathcal{S}| \leq q\}$ and $\mathbb{A}_1 = \{\mathcal{S} : \mathcal{S}_T \not\subset \mathcal{S}, |\mathcal{S}| \leq q\}$ the sets of the overfitted and underfitted models, respectively, where the size of any candidate model is no larger than the positive constant q defined in Condition (C7) in Section S1 of the supplementary material. Then, we obtain the theoretical properties of EBIC_{γ} given below.

Theorem 3. *Under Conditions (C1)–(C7) in Section S1 of the supplemen-*

tary material, as $nT \to \infty$*, we have*

$$P\Big\{\min_{\mathcal{S}\in\mathbb{A}_1} \mathrm{EBIC}_\gamma(\mathcal{S}) \le \mathrm{EBIC}_\gamma(\mathcal{S}_T)\Big\} \to 0$$

for any $\gamma > 0$ *and*

$$P\Big\{\min_{\mathcal{S}\in\mathbb{A}_0,\mathcal{S}\neq\mathcal{S}_T} \mathrm{EBIC}_\gamma(\mathcal{S}) \le \mathrm{EBIC}_\gamma(\mathcal{S}_T)\Big\} \to 0$$

for $\gamma > q^2C_w^2/\tau_2 c_{\min,3}\sigma_0^4 - 4$*, where* C_w*,* $c_{\min,3}$ *and* τ_2 *are finite positive constants which are defined in Conditions (C3), (C7) and Lemma 3 (ii), respectively, in Section S1 of the supplementary material.*

The above theorem holds as long as either n or T go to infinity. Note that the assumption $\min_{k\in\mathcal{S}_T}|\lambda_{0k}|\{nT/\log(nT)\}^{1/2} \to \infty$ given in Condition (C6) is modified from Chen and Chen (2008). This assumption is essential for showing the selection consistency of EBIC. Specifically, we demonstrate that $\widehat{\lambda}_k$ for $k \notin \mathcal{S}_T$ converges to 0 of order $(nT)^{-1/2}$. Under some mild conditions, we can further show that $\max_{k\notin\mathcal{S}_T}|\widehat{\lambda}_k| = O_p(\sqrt{\log(d)/nT}) = O_p(\sqrt{\log(nT)/nT})$. Thus, Condition (C6) indicates that $\min_{k\in\mathcal{S}_T}|\lambda_{0k}|$ is larger than $\max_{k\notin\mathcal{S}_T}|\widehat{\lambda}_k|$ asymptotically even with the diverging number of similarity matrices. Our simulation results indicate that $\gamma = 2$ performs satisfactorily under various settings. It is worth noting that we employ the popularly used backward elimination method to implement EBIC (see, e.g., Zhang and Wang (2011) and Schelldorfer et al. (2014)). This approach reduces the computational complexity from 2^d to $O(d^2)$. Thus, EBIC is computable when d is large.

3.2 Influence Matrix Test

To examine the adequacy of model (2.1) for modeling the mutual influence matrix B_t as a linear combination of weight matrices $W_k^{(t)}$ $(k = 1, \cdots, d)$, we consider the following hypotheses,

$$H_0 : B_t = \lambda_{01} W_1^{(t)} + \cdots + \lambda_{0d} W_d^{(t)} \text{ for all } t = 1, \cdots, T, \text{ vs}$$
$$H_1 : B_t \neq \lambda_{01} W_1^{(t)} + \cdots + \lambda_{0d} W_d^{(t)} \text{ for some } t = 1, \cdots, T. \quad (3.1)$$

Note that, under H_0, we have $\Sigma_t = \sigma_0^2 (I_n - B_t)^{-1} (I_n - B_t^\top)^{-1}$, which is a nonlinear function of the weight matrices $W_k^{(t)}$s. This is different from the covariance structure considered in Qu, Lindsay and Li (2000) and Zheng et al. (2019) which assumes that Σ_t is a linear function of the weight matrices.

To test (3.1), it is natural to compare the estimates of B_t calculated under the null and alternative hypotheses, respectively. Then reject the null hypothesis of (3.1) if their difference is relatively large. However, the computation of B_t under the alternative hypothesis is infeasible since it involves $n(n-1)T$ unknown parameters. Hence, we propose to test (3.1) by comparing the covariance matrix of Y_t under the null and alternative hypotheses, respectively. Under H_0, we have $\text{cov}(Y_t) = \Sigma_t = \sigma_0^2 (I_n - B_t)^{-1} (I_n - B_t^\top)^{-1}$.

Based on Theorem 2, B_t can be consistently estimated by $\widehat{B}_t = B_t(\widehat{\lambda})$. Accordingly, one can approximate $\text{cov}(Y_t)$ by $\widehat{\Sigma}_t = \widehat{\sigma}^2(I_n - \widehat{B}_t)^{-1}(I_n - \widehat{B}_t^\top)^{-1}$, where $\widehat{\sigma}^2 = (nT)^{-1}\sum_t Y_t^\top \Delta_t^\top(\widehat{\lambda})\Delta_t(\widehat{\lambda})Y_t$. On the other hand, $\text{cov}(Y_t)$ can be approximated by its sample version under the alternative, and we expect that $E(Y_tY_t^\top) \approx \widehat{\Sigma}_t$ under the null hypothesis, which motivates us to employ the quadratic loss function $tr(Y_tY_t^\top\widehat{\Sigma}_t^{-1} - I_n)^2$ to measure the difference between $Y_tY_t^\top$ and $\widehat{\Sigma}_t$. It is expected that, under H_0, the difference should be small across $t = 1, \cdots, T$. Hence, we propose the following test statistic,

$$T_{ql} = (nT)^{-1}\sum_{t=1}^{T} tr(Y_tY_t^\top\widehat{\Sigma}_t^{-1} - I_n)^2,$$

to assess the adequacy of (2.1).

To show the asymptotic distribution of T_{ql}, let $\mu_{ql} = n + \mu^{(4)} - 2$ and

$$\begin{aligned}\sigma_{ql}^2 = \;& (4\mu^{(4)} - 4)n/T + 4n^{-2}T^{-4}\sigma_0^4 \sum_{t_1\neq t_2\neq t_3}\sum_{k_1,l_1}\sum_{k_2,l_2}[\mathcal{I}_{k_1l_1}^{-1}(\theta_0)\mathcal{I}_{k_2l_2}^{-1}(\theta_0) \\ & \times\{tr(U_{t_1k_1}U_{t_1k_2}) + (\mu^{(4)} - 3)tr(U_{t_1k_1}\otimes U_{t_1k_2})\}tr(V_{t_2l_1})tr(V_{t_3l_2})] \\ & +(8\mu^{(4)} - 8)n^{-1}T^{-3}\sigma_0^4\sum_{t_1\neq t_2}\sum_{k,l}\mathcal{I}_{kl}^{-1}(\theta_0)tr(U_{t_1k})tr(V_{t_2l}), \qquad (3.2)\end{aligned}$$

where $\mathcal{I}_{kl}^{-1}(\theta_0)$ is the kl-th element of $\mathcal{I}^{-1}(\theta_0)$, $U_{tk} = s\{W_k^{(t)}\Delta_t^{-1}(\lambda_0)\}$, $V_{tk} = \{\Delta_t^{-1}(\lambda_0)\}^\top\widetilde{\Lambda}_{tk}\Delta_t^{-1}(\lambda_0)$ and $\widetilde{\Lambda}_{tk}$ is the matrix form of $\partial\text{vec}\{\Sigma_t^{-1}(\theta_0)\}/\partial\theta_k$ for $t_1, t_2, t_3, t = 1, \cdots, T$, $k_1, k_2, k = 1, \cdots, d$, and $l_1, l_2, l = 1, \cdots, d$. Then, the next theorem presents the asymptotic property of T_{ql}.

Theorem 4. *Under the null hypothesis of H_0, Conditions (C1)–(C5) in Section S1 of the supplementary material and assuming that $n/T \to c$ and $\sigma_{ql}^2 > c_\sigma$ for some finite positive constants c and c_σ, we have*

$$(T_{ql} - \mu_{ql})/\sigma_{ql} \to_d N(0,1)$$

as $nT \to \infty$.

Unlike Theorems 1–3, the above result requires that both n and T tend to infinity with $n/T \to c$ for some finite positive constant c. This condition is reasonable since we need the replications of similarity matrices to test the adequacy of MIR. Note that this condition is commonly used for testing high dimensional covariance structures (see, e.g., Ledoit and Wolf (2002) and Zheng et al. (2019)). The above theorem indicates that the asymptotic variance of T_{ql} is σ_{ql}^2, which is given in (3.2) and it includes three components. The first component $(4\mu^{(4)} - 4)c$ is the leading term of variance of $(nT)^{-1}\sum_{t=1}^T tr(Y_t Y_t^\top \Sigma_t^{-1} - I_n)^2$ obtained by assuming that λ_0 is known, while the last two components are of orders $O(d^2)$ and $O(d)$, respectively, and cannot be ignored. These two non-negligible components are mainly induced by the estimator $\widehat{\lambda}$, which makes the proof of Theorem 4 more complicated. Thus, we develop Lemma 4 in Section S1 of the supplementary material to resolve this challenging task.

To make the above theorem practically useful, one needs to estimate the two unknown terms μ_{ql} and σ_{ql}. Note that $\mu^{(4)}$ in μ_{ql} can be consistently estimated by $\widehat{\mu}^{(4)}$, which is defined in the explanation of Theorem 1. As a result, $\widehat{\mu}_{ql} = n + \widehat{\mu}^{(4)} - 2$ is a consistent estimator of μ_{ql}. It is also worth noting that U_{tk}, V_{tk} and $\mathcal{I}_{kl}^{-1}(\theta_0)$ can be consistently estimated by $\widehat{U}_{tk} = s(W_k^{(t)}\Delta_t^{-1}(\widehat{\lambda}))$, $\widehat{V}_{tk} = \{\Delta_t^{-1}(\widehat{\lambda})\}^\top \widehat{\Lambda}_{tk}\Delta_t^{-1}(\widehat{\lambda})$ and $\mathcal{I}_{kl}^{-1}(\widehat{\theta})$, respectively, for $t = 1, \cdots, T$ and $k, l = 1, \cdots, d$, where $\widehat{\Lambda}_{tk}$ is the matrix form of $\partial \text{vec}\{\Sigma_t^{-1}(\widehat{\theta})\}/\partial\theta_k$ and $s(A) = (A + A^\top)/2$ for any arbitrary matrix A defined in Section 2.2. Accordingly, $\widehat{\sigma}_{ql}$, obtained by replacing unknown parameters with their corresponding estimators, is a consistent estimator of σ_{ql}. Consequently, for a given significance level α, we are able to reject the null hypothesis of H_0 if $|T_{ql} - \widehat{\mu}_{ql}| > \widehat{\sigma}_{ql} z_{1-\alpha/2}$, where z_α stands for the α-th quantile of the standard normal distribution.

4. Simulation Studies

To demonstrate the finite sample performance of our proposed MIR model, we conduct the following simulation studies. The similarity matrices $A_k^{(t)} = (a(Z_{kj_1}^{(t)}, Z_{kj_2}^{(t)})) \in \mathbb{R}^{n\times n}$ with zero diagonal elements and $a(Z_{kj_1}^{(t)}, Z_{kj_2}^{(t)}) = \exp\{-(Z_{kj_1}^{(t)} - Z_{kj_2}^{(t)})^2\}$ if $|Z_{kj_1}^{(t)} - Z_{kj_2}^{(t)}| < \phi_k^{(t)}$, $a(Z_{kj_1}^{(t)}, Z_{kj_2}^{(t)}) = 0$ otherwise, where j_1 and j_2 range from 1 to n and $Z_k^{(t)} = (Z_{k1}^{(t)}, \cdots, Z_{kn}^{(t)})^\top$ are iid according to a multivariate normal distribution with mean 0 and covariance matrix I_n for $k = 1, \cdots, d$ and $t = 1, \cdots, T$, and $\phi_k^{(t)}$ is selected to control the density of $A_k^{(t)}$ (i.e., the proportion of nonzero elements) defined as $10/n$ for any k and t (see, e.g., Zou et al. (2017)). Accordingly, we obtain $W_k^{(t)} = (w(Z_{kj_1}^{(t)}, Z_{kj_2}^{(t)}))_{n\times n}$ with $w(Z_{kj_1}^{(t)}, Z_{kj_2}^{(t)}) = a(Z_{kj_1}^{(t)}, Z_{kj_2}^{(t)})/\sum_{j_2} a(Z_{kj_1}^{(t)}, Z_{kj_2}^{(t)})$. The random errors ϵ_{it} are iid and simulated from three distributions: (i) the standard normal distribution $N(0, 1)$; (ii) the standardized exponential distribution; (iii) the mixture distribution $0.9N(0, 5/9) + 0.1N(0, 5)$. The last two distributions allow us to examine the robustness of parameter estimates to other distributions. Finally, the response vectors Y_t are generated by $Y_t = (I_n - \lambda_1 W_1^{(t)} - \cdots - \lambda_d W_d^{(t)})^{-1}\epsilon_t$ for $t = 1, \cdots, T$. Note that the random error ϵ_t is independent of $Z_k^{(t)}$ for any $k = 1, \cdots, d$ and $t = 1, \cdots, T$.

For each of the random error distributions, we consider three different numbers of observations $T = 25, 50$ and 100, three different numbers of actors $n = 25, 50$ and 100, and all of the results are generated with 500 realizations. Since the results for all three error distributions are qualitatively similar, we only present the results for the standard normal distribution, and the results for the mixture normal and the standardized exponential distributions are relegated to the supplementary material.

To assess the performance of parameter estimators, we consider three different numbers of covariates $d = 2, 6$ and 12, where $d = 2$ is borrowed from Zou et al. (2017)), $d = 6$ is used in our real data analysis, and $d = 12$ is an exploration of larger similarity matrices. Since the simulation results for $d = 12$ are qualitatively similar to those for $d = 2$ and 6, we report them in the supplementary material. The regression coefficients are $\lambda_k = 0.1$ for $k = 1, \cdots, d$. In addition, let $\widehat{\lambda}^{(m)} = (\widehat{\lambda}_1^{(m)}, \cdots, \widehat{\lambda}_d^{(m)})^\top \in \mathbb{R}^d$ be the parameter estimate in the m-th realization obtained via the proposed QMLE. For each $k = 1, \cdots, d$, we evaluate the average bias of $\widehat{\lambda}_k^{(m)}$ by BIAS=$500^{-1} \sum_m (\widehat{\lambda}_k^{(m)} - \lambda_k)$. Using the results of Theorem 1, we compute the standard error of $\widehat{\lambda}_k^{(m)}$ via its asymptotic distribution, and denote it $\mathrm{SE}^{(m)}$. Then, the average of the estimated standard errors is SE=$500^{-1} \sum_m \mathrm{SE}^{(m)}$. To assess the validity of the estimated standard errors, we also calculate the true standard error via the 500 realizations and denote it $\mathrm{SE}^* = 500^{-1} \sum_m (\widehat{\lambda}_k^{(m)} - \bar{\lambda}_k)^2$, where $\bar{\lambda}_k = 500^{-1} \sum_m \widehat{\lambda}_k^{(m)}$.

Table 1 presents the results of BIAS, SE and SE* over 500 realizations for $k = 1, \cdots, d$ and $d = 2$ and 6. It indicates that the biases of the parameter estimates are close to 0 for any n and T, and they become smaller as either n or T gets larger. In addition, the variation of the parameter estimate, SD, shows similar findings to those of BIAS. Moreover, the difference between SD and SD* is quite small when either n or T is large. In sum,

Table 1 demonstrates that the asymptotic results obtained in Theorem 1 are reliable and satisfactory.

We next assess the performance of the proposed EBIC criterion by considering three sizes of the full model, $d = 6, 8$, and 12, while the size of the true model is $|\mathcal{S}_T| = 3$. We set $\lambda_k = 0.2$ for any $k \in \mathcal{S}_T$ and $\lambda_k = 0$ otherwise. To implement the EBIC criterion, we set $\gamma = 2$ in this simulation study. Four performance measures are used: (i) the average size (AS) of the selected model $|\widehat{\mathcal{S}}|$; (ii) the average percentage of the correct fit (CT), $I(\widehat{\mathcal{S}} = \mathcal{S}_T)$; (iii) the average true positive rate (TPR), $|\widehat{\mathcal{S}} \cap \mathcal{S}_T|/|\mathcal{S}_T|$; and (iv) the average false positive rate (FPR), $|\widehat{\mathcal{S}} \cap \mathcal{S}_T^c|/|\mathcal{S}_T^c|$. Since the results for all three values of d exhibit a quantitatively similar pattern, we only present the results for $d = 8$.

Table 2 shows that the average percentage of correct fit, CT, increases toward to 100% when either n or T gets large. It is worth noting that the CTs are larger than 70% even when both n and T are small, i.e., $n = 25$ and $T = 25$. Furthermore, the average true positive rate, TPR, is 100%, which indicates that EBIC is unlikely to select an underfitted model even when both n and T are small. In contrast, the average false positive rate, FPR, decreases toward 0 when either n or T becomes large. Moreover, the average size (AS) of the selected model, $|\widehat{\mathcal{S}}|$, approaches the true model

Table 1: The bias and standard error of the parameter estimates when the true parameters are $\lambda_k = 0.1$ for $k = 1, \cdots, d$, and the random errors follow a normal distribution. BIAS: the average bias; SE: the average of the estimated standard errors via Theorem 1; SE*: the standard error of parameter estimates calculated from 500 realizations.

			$d=2$		$d=6$					
n	T		λ_1	λ_2	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
25	25	BIAS	-0.009	0.006	-0.001	-0.004	0.001	0.000	-0.002	-0.006
		SE	0.054	0.054	0.055	0.055	0.055	0.055	0.055	0.055
		SE*	0.058	0.055	0.056	0.055	0.054	0.056	0.052	0.052
25	50	BIAS	-0.002	-0.004	0.001	-0.006	0.001	-0.002	0.001	-0.001
		SE	0.038	0.038	0.039	0.039	0.039	0.039	0.039	0.039
		SE*	0.039	0.041	0.039	0.042	0.042	0.039	0.038	0.039
25	100	BIAS	-0.001	-0.002	0.002	-0.002	-0.000	-0.002	-0.000	0.001
		SE	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.027
		SE*	0.027	0.029	0.028	0.027	0.024	0.026	0.027	0.027
50	25	BIAS	-0.002	-0.003	-0.003	-0.003	-0.001	-0.000	0.001	-0.000
		SE	0.038	0.038	0.037	0.037	0.037	0.037	0.037	0.037
		SE*	0.038	0.037	0.035	0.034	0.038	0.037	0.033	0.036
50	50	BIAS	-0.001	0.000	0.000	-0.000	-0.001	0.001	0.001	-0.005
		SE	0.027	0.027	0.026	0.026	0.026	0.026	0.026	0.026
		SE*	0.025	0.028	0.026	0.028	0.028	0.027	0.027	0.028
50	100	BIAS	-0.001	-0.002	-0.001	-0.001	-0.000	-0.000	0.001	0.001
		SE	0.019	0.019	0.018	0.018	0.018	0.018	0.018	0.018
		SE*	0.019	0.019	0.018	0.020	0.019	0.018	0.017	0.019
100	25	BIAS	0.000	-0.001	-0.001	-0.000	-0.003	0.000	0.001	-0.002
		SE	0.026	0.027	0.026	0.026	0.026	0.026	0.026	0.026
		SE*	0.026	0.028	0.026	0.026	0.028	0.026	0.028	0.026
100	50	BIAS	-0.001	0.001	0.000	-0.001	0.000	0.001	0.001	0.000
		SE	0.019	0.019	0.018	0.018	0.018	0.018	0.018	0.018
		SE*	0.019	0.018	0.017	0.016	0.018	0.019	0.016	0.018
100	100	BIAS	-0.001	-0.001	-0.000	-0.001	0.001	0.002	-0.000	0.001
		SE	0.013	0.013	0.013	0.013	0.013	0.013	0.013	0.013
		SE*	0.013	0.013	0.012	0.013	0.014	0.012	0.013	0.013

Table 2: Model selection via EBIC when $d = 8$ and the random errors are normally distributed. AS: the average size of the selected model; CT: the average percentage of the correct fit; TPR: the average true positive rate; FPR: the average false positive rate.

n	T	AS	CT	TPR	FPR
25	25	3.3	72.6	91.8	9.8
	50	3.2	77.1	95.7	8.5
	100	3.1	81.2	100.0	5.9
50	25	3.2	78.2	94.0	7.9
	50	3.1	80.8	97.2	5.8
	100	3.1	84.7	100.0	5.1
100	25	3.1	82.3	100.0	6.7
	50	3.1	83.8	100.0	5.1
	100	3.0	87.7	100.0	4.2

size. The above results indicate that EBIC performs satisfactorily in finite samples.

Lastly, we examine the performance of the proposed goodness of fit test. We consider a generative model $B_t = \lambda_1 W_1^{(t)} + \cdots + \lambda_d W_d^{(t)} + \kappa E E^\top$, where $E \in \mathbb{R}^n$ is a random normal vector of dimension n with each elements that are iid simulated from a standard normal distribution. The parameter κ is a measure of departure from the null model of H_0. Specifically, $\kappa = 0$ corresponds to the null model, while $\kappa > 0$ represents alternative models. Accordingly, the results for $\kappa = 0$ represent empirical sizes, while the results for $\kappa > 0$ denote empirical powers.

Table 3 indicates that the empirical sizes are slightly conservative when both n and T are small. However, they approach the significance level of 5% when either n or T becomes large. Furthermore, the empirical powers increase as either n or T gets larger. Moreover, they become stronger when κ increases; in particular the empirical power approaches 1 when either n or T equals 100 and $\kappa = 0.2$. The above findings are robust to non-normal error distributions; see Tables S.4 and S.7 in the supplementary material. Consequently, our proposed goodness of fit test not only controls the size well, but is also consistent. It is worth noting that the above estimation, selection and test findings are also robust to non-normal error distributions;

see Tables S.2 to S.7 in the supplementary material.

Table 3: The empirical sizes and powers of the goodness of fit test. The $\kappa = 0$ corresponds to the null model and $\kappa > 0$ represents alternative models. The random errors are normally distributed, and the full model sizes are $d = 2$ and 6.

			d=2			d=6	
n	T	κ=0	κ=0.1	κ=0.2	κ=0	κ=0.1	κ=0.2
25	25	0.030	0.296	0.664	0.024	0.242	0.584
	50	0.034	0.528	0.838	0.030	0.424	0.748
	100	0.042	0.660	0.910	0.042	0.560	0.822
50	25	0.028	0.434	0.772	0.022	0.342	0.654
	50	0.037	0.582	0.878	0.036	0.476	0.786
	100	0.044	0.706	0.974	0.048	0.654	0.954
100	25	0.034	0.510	0.976	0.030	0.452	0.964
	50	0.040	0.738	1.000	0.034	0.588	0.996
	100	0.048	0.910	1.000	0.046	0.830	1.000

5. Real Data Analysis

5.1 Background and Data

To demonstrate the practical usage of our proposed MIR model, we present an empirical example for exploring the mechanism of spillover effects in Chinese mutual funds. It is known that the income and profit of a mutual fund is largely compensated from the management fees, which are charged as a fixed proportion of the total net assets under management. As a result, the variation in cash flow across time is one of the most influential indices closely monitored by fund managers. Thus, exploring the mechanism of cash flow is extremely essential (see e.g., Spitz (1970); Nanda, Wang and Zheng (2004); Brown and Wu (2016)). However, past literatures mainly focus on addressing the characteristics of the mutual funds that affect their cash flow from a cross-sectional prospective (see, e.g., Brown and Wu (2016)). In this study, we employ our proposed MIR model to identify the mutual fund characteristics that can yield mutual influence on fund cash flows (i.e., a spillover effect) from a network perspective.

To proceed with our study, we collect quarterly data from 2010-2017 on actively managed open ended mutual funds through the WIND financial database, which is one of the most authoritative databases regarding the

Chinese financial market. After removing funds with missing observations or existing for less than one year, there are $n = 90$ mutual funds in this empirical study with $T = 32$. The response variable, the cash flow rate of fund i at time t, can be calculated as follows (Nanda, Wang and Zheng (2004)):

$$C_{it} = \frac{TA_{it} - TA_{i,t-1}(1 + r_{it})}{TA_{it}},$$

where TA_{it} and r_{it} are the total net assets and the return of fund i at time t, respectively.

We next generate the similarity matrices to explore the mechanism of spillover effects among mutual funds. To this end, we consider the following five covariates in the spirit of the pioneering work of Spitz (1970). (i). Size: the logarithm of the total net asset of fund i at time $t-1$; (ii) Age: the logarithm of the age of fund i at time $t-1$; (iii) Return; the return of fund i at time $t-1$; (iv) Alpha: the risk-adjusted return of fund i at time $t-1$ measured by the intercept of Carhart (1997) four factor model; (v) Volatility: the standard deviation of the weekly return of fund i and time $t-1$. We next generate the similarity matrices. For the Size covariate, we standardize the data to have zero mean and unit variance, and denote it SIZE_{it} for $i = 1, \cdots, n$ and $t = 1, \cdots, T$. Then, the similarity matrix induced by Size is $A_1^{(t)} = (a(Z_{1j_1}^{(t)}, Z_{1j_2}^{(t)}))$ with zero diagonal elements and $a(Z_{1j_1}^{(t)}, Z_{1j_2}^{(t)}) =$

$\exp\{-(Z_{1j_1}^{(t)} - Z_{1j_2}^{(t)})^2\}$ when $|Z_{1j_1}^{(t)} - Z_{1j_2}^{(t)}| < \phi_1^{(t)}$ for a pre-specified finite positive constant $\phi_1^{(t)}$, and $a(Z_{1j_1}^{(t)}, Z_{1j_2}^{(t)}) = 0$ otherwise. As given in simulation studies, $\phi_1^{(t)}$ is selected so that the proportion of nonzero elements of $A_1^{(t)}$ is $10/n$. Subsequently, we obtain $W_1^{(t)} = (w(Z_{1j_1}^{(t)}, Z_{1j_2}^{(t)}))_{n\times n}$ and $w(Z_{1j_1}^{(t)}, Z_{1j_2}^{(t)}) = a(Z_{1j_1}^{(t)}, Z_{1j_2}^{(t)})/\sum_{j_2} a(Z_{1j_1}^{(t)}, Z_{1j_2}^{(t)})$, which is the row-normalized version of $A_1^{(t)}$. Analogously, we can construct the similarity matrices $W_2^{(t)}, \cdots, W_5^{(t)}$ associated with the remaining four covariates, respectively.

5.2 Empirical Results

We first employ the adequacy test to assess whether the five covariates are sufficient to explain the mutual influence matrix. The resulting p-value for testing the null hypothesis of H_0 in (3.1) is 0.660, which is not significant under the significance level of 5%. This indicates that one or more of the five covariates in the MIR model provide a good fit to the data.

We next employ the proposed QMLE method to estimate the model. Table 5 presents the parameter estimates, standard errors, and their associated p-values. It indicates that the covariates Return, Age and Volatility are significant and positive. It is worth noting that these three covariates are all related to the funds' performance and operating capacity. Hence, we conclude that the funds' cash flows are influenced by other funds with similar performance and operating capacity. Furthermore, the estimate of Size is positive and significant, which implies that the funds' cash flows are influenced by other funds of similar size. In other words, investors tend to invest in larger mutual funds. Moreover, the estimate of Alpha is positive but not significant. Hence, investors pay more attention to raw returns than risk-adjusted returns in judging a fund's performance. This can be due to the fact that raw returns are easier to observe.

Table 4: The QMLE parameter estimates and associated standard errors and p-values for the five covariates.

	Estimate	Standard-Error	p-Value
Alpha	0.005	0.027	0.853
Return	0.569	0.019	0.000
Size	0.330	0.014	0.000
Age	0.036	0.018	0.046
Volatility	0.209	0.020	0.000

Subsequently, we employ EBIC to determine the most relevant covariates that are related to the cash flow with $\gamma = 2$ as in the simulation studies. The resulting model consists of the covariates Return and Size. This implies that fund managers tend to learn relevant information from other funds with a large size and good performance. This finding is consistent with existing studies (see, e.g., Brown, Harlow and Starks (1996)). To check the robustness of our results against the selection of $\phi_k^{(t)}$, we also consider $\phi_k^{(t)}$ so that the proportion of nonzero elements of the weight matrices are $5/n$ and $20/n$. The results yield similar findings to that of $10/n$. Moreover, we consider the two alternative non-decreasing functions of $a(\cdot)$, i.e., $a(x) = 1/(1 + x^2)$ and $a(x) = 1/(1 + x^2)^2$. The estimation results (not reported here) are almost identical to those in Table 4. Hence, our results are not affected by these two alternatives. In sum, the MIR model can provide valuable insight for understanding the mechanism of mutual influence among mutual funds.

6. Conclusion

In this article, we propose the mutual influence regression (MIR) model to explore the mechanism of mutual influence by establishing a relationship between the mutual influence matrix and a set of similarity matrices induced by their associated attributes among the actors. In addition, we allow the number of similarity matrices to diverge. The theoretical properties of the MIR model's estimations, selections, and assessments are established. The Monte Carlo studies support the theoretical findings, and an empirical example illustrates the practical application.

To broaden the usefulness of MIR, we identify six possible avenues for future research. The first avenue is to allow the regression coefficients to change with t that increases model flexibility. The second avenue is to generalize the model by accommodating discrete responses. The third avenue is to extend the linear regression structure of MIR to the nonparametric or semiparametric setting by changing $\lambda_k W_k^{(t)}$ to $g(\lambda_k, W_k^{(t)})$ for some unknown smooth function $g(\cdot)$. The forth avenue is to develop a fast algorithm with theoretical justification that can implement MIR when n or d is large, such as the one-step estimate proposed by Gupta (2021). The fifth avenue is to develop a criterion to obtain the optimal γ for EBIC. The last avenue is to introduce a method for choosing the thresholds or cut-off points of the weight matrices. We believe that these efforts would further increase the application of the MIR model.

Supplementary Material

The Supplementary Material contains the conditions and proofs of the theorems and additional simulation settings and results.

Acknowledgements

Xinyan Fan's research was supported by the National Natural Science Foundation of China (NSFC, 12201626), and the Public Computing Cloud, Renmin University of China. Wei Lan's research was supported by the National Natural Science Foundation of China (NSFC, 71991472, 12171395, 11931014), National Key R&D Program of China (2022YFA1003702 & 2022YFA1003700), the Joint Lab of Data Science and Business Intelligence at Southwestern University of Finance and Economics, and the Fundamental Research Funds for the Central Universities (JBK1806002). Tao Zou's research was supported by ANU College of Business and Economics Early Career Researcher Grant, the RSFAS Cross-Disciplinary Grant, and the assistance of computational resources provided by the Australian Government through the National Computational Infrastructure (NCI) under the ANU Merit Allocation Scheme (ANUMAS).

References

Anderson, T, W. (1973). Asymptotically efficient estimation of covariance matrices with linear structure. *Annals of Statistics* **1**, 135–141.

Brown, C., Harlow, V. and Starks, T. (1996). Of tournaments and temptations: An analysis of managerial incentives in the mutual fund industry. *Journal of Finance* **51**, 85–110.

Brown, D. P. and Wu, Y. (2016). Mutual fund flows and cross-fund learning within families. *Journal of Finance* **71**, 383-424.

Carhart, M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance* **52**, 57–82.

Chen, J. and Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95**, 759–771.

Dou, B., Parrella, M. and Yao, Q. (2016). Generalized Yule-Walker estimation for spatio-temporal models with unknown diagonal coefficients. *Journal of Econometrics* **94**, 369–382.

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

Furrer, R., Genton, M. and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics* **15**, 502–523.

Gao, Z., Ma, Y., Wang, H. and Yao, Q. (2019). Banded spatio-temporal autoregressions. *Journal of Econometrics* **208**, 211–230.

Golgher, A. and Voss, P. (2016). How to interpret the coefficients of spatial models: spillovers, direct and indirect effects. *Spatial Demography* **4**, 175–205.

Gupta, A. (2021). Efficient closed-form estimation of large spatial autoregressions. *Journal of Econometrics*, In Press.

Gupta, A. and Robinson, P. (2015). Inference on higher-order spatial autoregressive models with increasingly many parameters. *Journal of Econometrics* **186**, 19–31.

Gupta, A. and Robinson, P. (2018). Pseudo maximum likelihood estimation of spatial autoregressive models with increasing dimension. *Journal of Econometrics* **202**, 92–107.

Han, X., Hu, Y., Fan, L., Huang, Y., Xu, M. and Gao, S. (2021). Quantifying COVID-19 importation risk in a dynamic network of domestic cities and international countries. *Proceeding of the National Academy of Sciences* **118**, e2100201118.

Huang, D., Lan, W., Zhang, H. and Wang, H. (2019). Least squares estimation for social autocorrelation in large-scale networks. *Electronic Journal of Statistics* **13**, 1135–1165.

Jenish, N. and Prucha, I. R. (2012). On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics* **170**, 178–190.

Kelejian, H. and Prucha, I. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* **157**, 53–67.

Kwok, H. H. (2019). Identification and estimation of linear social interaction models. *Journal of Econometrics* **210**, 434-458.

Kwok, H. H. (2020). Identification methods for social interactions models with unknown networks. *In The Econometrics of Networks.* Emerald Publishing Limited.

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

Ledoit, O. and Wolf, M. (2002). Some hypotheses tests for the covariance matrix when the dimension is large compare to the sample size. *Annals of Statistics* **30**, 1081–1102.

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

Lee, L. F. and Liu, X. (2010). Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory* **26**, 187-230.

Lee, L. F. and Yu, J. (2014). Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics* **180**, 174-197.

LeSage, J. and Pace, R. K. (2009). *Introduction to Spatial Econometrics*, New York: Chapman & Hall.

Manresa, E. (2013). Estimating the structure of social interactions using panel data. *Unpublished Manuscript*, CEMFI, Madrid.

Nanda, N., Wang, J. and Zheng, L. (2004). Family values and the star phenomenon: Strategies of mutual fund families. *Review of Financial Studies* **17**, 667–698.

Peng, S., Yang, A., Cao, L., Yu, S. and Xie, D. (2017). Social influence modeling using information theory in mobile social networks. *Information Sciences* **379**, 146–159.

de Paula, Á., Rasul, I. and Souza, P. (2019). Identifying network ties from panel data: Theory and an application to tax competition. arXiv preprint arXiv:1910.07452.

Qu, A., Lindsay, G. and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87**, 823–836.

Schelldorfer, J., Meier, L. and Bühlmann, P. (2014). Glmmlasso: an algorithm for high-dimensional generalized linear mixed models using ℓ_1-penalization, *Journal of Computational and Graphical Statistics* **23**, 460-477.

Spitz, E. (1970). Mutual fund performance and cash inflows. *Applied Economics* **2**, 141–145.

Trusov, M., Bodapati, A. and Bucklin, R. (2010). Determining influential users in internet social networks. *Journal of Marketing Research* **47**, 643–558.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, Mass.

Wang, D., Zheng, Y., Lian, H. and Li, G. (2022). High-dimensional vector autoregressive time series modeling via tensor decomposition. *Journal of the American Statistical Association* **117**, 1338–1356.

Zhang, Q. and Wang, H. (2011). On BIC's selection consistency for discriminant analysis, *Statistica Sinica* **21,** 731–740.

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

Zheng, S., Chen, Z., Cui, H. and Li, R. (2019). Hypothesis testing on linear structures of high dimensional covariance matrix. *Annals of Statistics* **47**, 3300–3334.

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. and Wang, H. (2017). Network vector autoregression. *Annals of Statistics* **45**, 1096–1123.

Zou, T., Lan, W., Wang, H. and Tsai, C. L. (2017). Covariance regression analysis. *Journal of the American Statistical Association* **112**, 266–281.

Xinyan Fan

Renmin University of China, Beijing 100086, China.

E-mail: 1031820039@qq.com

Wei Lan

Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China.

E-mail: lanwei@swufe.edu.cn

Tao Zou

The Australian National University, Canberra, ACT 2600, Australia.

E-mail: tao.zou@anu.edu.au

Chih-Ling Tsai

University of California, Davis, CA 95616, USA.

E-mail: cltucd@gmail.com