Statistica Sinica Preprint No: SS-2021-0377	
Title	Spatial Autoregressive Models with Generalized Spatial Disturbances
Manuscript ID	SS-2021-0377
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202021.0377
Complete List of Authors	Kuangnan Fang, Wei Lan, Dan Pu and Qingzhao Zhang
Corresponding Authors	Wei Lan
E-mails	lanwei@swufe.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

Spatial Autoregressive Models with Generalized Spatial Disturbances

Kuangnan Fang[1], Wei Lan[2], Dan Pu[1] and Qingzhao Zhang[1]

[1] *Xiamen University and*

[2] *Southwestern University of Finance and Economics*

Abstract: We propose a spatial autoregressive model with generalized disturbances to simultaneously model the spatial effects between the response variables and those between the disturbance terms. By directly modeling the covariance matrix of the disturbance terms as a polynomial function of a row-normalized adjacency matrix with a prespecified upper order that may tend to infinity, our model includes the traditional spatial autoregressive model with moving average disturbances and that with autoregressive disturbances as special cases. We propose a quasi-maximum likelihood estimator (QMLE) for estimating the model, and use an approximate maximum likelihood estimator (AMLE), which is feasible for large-scale networks, to alleviate the computational cost. We establish the asymptotic properties of both estimators (i.e., QMLE and AMLE), without imposing any distribution assumptions. Because the number of matrix predictors diverges, we propose a type of extended Bayesian information criterion method for model selection, and demonstrate its selection consistency. The results of our simulation studies and an analysis of the spatial effects in mutual fund cash

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

inflows demonstrate the usefulness of the proposed model.

Key words and phrases: Approximate maximum likelihood estimator, Extended Bayesian information criterion, Generalized disturbances, Quasi-maximum likelihood estimator, Spatial autoregressive model.

1. INTRODUCTION

Network data are becoming increasingly available, owing to the rapid development of online social networks (e.g., Facebook and Weibo), and interest in the analysis of such data has increased accordingly; for a good summary, see Knoke and Yang (2008), Kolaczyk (2009), and Newman (2010). Because the nodes within a network can be connected, the traditional independent and identically distributed (i.i.d) assumption for sampled data is no longer valid, and can result in invalid statistical inferences (see, e.g., Anselin, 1988). As a result, finding ways of exploring the dependence structure of network data has attracted much interest (see, e.g., Hoff et al., 2002; Chang et al., 2019).

The spatial autoregressive model (SAR) is a popular method for modeling the dependence structure of different nodes within a network (see, e.g., Cliff and Ord, 1973; Anselin, 1988). Given a network of n nodes, indexed by $1 \leq i \leq n$, we define the adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$, where

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

$a_{ij} = 1$ if nodes i and j are connected, and $a_{ij} = 0$ otherwise. We define $a_{ii} = 0$, for any i. To model the dependence structure of the responses $Y = (Y_1, \cdots, Y_n)^\top$ of the n nodes, the SAR with covariates is given by

$$Y = \rho WY + X\boldsymbol{\alpha} + \varepsilon,$$

where $W = (w_{ij}) \in \mathbb{R}^{n\times n}$, with $w_{ij} = a_{ij}/\sum_j a_{ij}$, is a row-normalized adjacency matrix, $X = (X_1, \cdots, X_n)^\top \in \mathbb{R}^{n\times p}$ represents the covariates collected from the n nodes, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_p)^\top \in \mathbb{R}^p$ is the unknown regression coefficient, and $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)^\top \in \mathbb{R}^n$ is a random noise term. The SAR model characterizes the sophisticated network dependency in a parsimonious manner using the autoregressive parameter ρ. To account for more complicated situations, several extensions of the SAR model have been developed, and its usefulness is widely recognized in research domains such as real estate (e.g., Osland, 2010), criminology (e.g., Kakamu et al., 2008), economics and finance (e.g., Arnold et al., 2013), and sociology (e.g., Lin, 2010; Hsieh and Lee, 2016).

However, despite many recent advances and successful applications of the SAR model, to the best of our knowledge, most models assume that the disturbance term ε is independent across different nodes, which is often overly restrictive in practice. For example, Behrens et al. (2012) found that in addition to the spatial interdependence between trade flows, cross-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

sectional correlation exists among disturbance terms, which might be caused by measurement errors. Catania and Billé (2017) find that unobservable factors result in the disturbance terms of financial returns within different sectors being spatially correlated; for additional applications, see Baltagi and Bresson (2011), Fingleton and Szumilo (2019), and Fingleton (2020). Thus, it is essential that we model the spatial effects in the dependent variable and those in the disturbance term simultaneously. To do so, we consider the following spatial autoregressive model with a spatial autoregressive process in the disturbance term (SARAR(1,1)) (see, e.g., Kelejian and Prucha, 1998; Lee, 2003; Kelejian and Prucha, 2010; Liu et al., 2010):

$$Y = \rho WY + X\boldsymbol{\alpha} + \varepsilon, \varepsilon = \lambda W\varepsilon + u,$$

where λ is the spatial autoregressive parameter of the error term and $u = (u_1, \cdots, u_n)^\top \in \mathbb{R}^n$, the elements of which are i.i.d with mean zero and variance σ^2. If $|\lambda| < 1$, the covariance matrix of the original disturbance ε is $E(\varepsilon\varepsilon^\top) = \sigma^2(I_n - \lambda W)^{-1}(I_n - \lambda W^\top)^{-1} = \sigma^2(\sum_{k=0}^{\infty} \lambda^k W^k)(\sum_{i=0}^{\infty} \lambda^k W^k)^\top$, where $I_n \in \mathbb{R}^{n\times n}$ is the identity matrix of dimension n, and W^k denotes the kth power of W. Therefore, the SARAR(1,1) model links the covariance matrix to a row-normalized adjacency matrix of infinite order. Furthermore, $\varepsilon = (I_n - \lambda W)^{-1}u = (\sum_{k=0}^{\infty} \lambda^k W^k)u$ indicates that the SARAR(1,1) model allows a nodal-specific shock to transmit to all other nodes through higher-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

order neighbors. However, in some cases, shocks are bounded to a small local neighborhood, and are not transmitted to other nodes. Hence, we consider the following alternative model, called the spatial autoregressive model with moving average disturbances of order (1,1), that is, SARMA(1,1)(see, e.g., Anselin and Florax, 1995; Fingleton, 2008; Dogan and Taspinar, 2013; Dogan, 2015):

$$Y = \rho WY + X\boldsymbol{\alpha} + \varepsilon, \varepsilon = u + \lambda Wu.$$

In this model, the corresponding covariance matrix of the original disturbance is $E(\varepsilon\varepsilon^\top) = \sigma^2(I_n - \lambda W)(I_n - \lambda W)^\top = \sigma^2\{I_n - \lambda(W + W^\top) + \lambda^2 WW^\top\}$, which is limited to the first and second order of the row-normalized adjacency matrix. In addition to the above two types of models, we examine high-order and infinite-order spatial lags in the disturbance terms (see, e.g., Lee and Liu, 2010; Badinger and Egger, 2011; Gupta and Robinson, 2018). Although the SARAR and SARMA models are commonly used, which model should be used in empirical studies is unclear. Some non-nested tests have been developed to distinguish spatial models (see, e.g., Delgado and Robinson, 2015). In contrast, we propose a general model that includes the traditional SARMA and SARAR models as special cases.

We call the proposed model the spatial autoregressive model with generalized spatial disturbances (SARg), and model the covariance matrix of

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

the disturbances as a polynomial function of the row-normalized adjacency matrix. That is, we directly link the covariance matrix Σ of the disturbances with the matrix predictors W^k $(k = 1, \cdots, d)$ using unknown parameters. By allowing the number of matrix predictors to increase slowly with the network size, this model includes the SARAR and SARMA models as special cases. Specifically, we set $d = \infty$ for SARAR(1,1) and $d = 2$ for SARMA(1,1). Accordingly, in contrast to SARAR and SARMA, the proposed model is sufficiently flexible to capture complex dependence structures.

First, to estimate the model, we employ the popular quasi-maximum likelihood estimator (QMLE; see, e.g., Smirnov and Anselin, 2001; Lee, 2004). Although the consistency and the asymptotic normality of the QMLE with fixed-dimensional parameters are well established, whether they hold with diverging parameters is doubtful. In this case, traditional pointwise convergence is no longer valid, and we need to consider the consistency of the vector and the matrix norms.

Second, when d is large, it is essential to identify which matrix predictors (i.e., W^k, for $k = 1, \cdots, d$) are relevant. However, when d is diverging, traditional model selection criteria, such as the Bayesian information criterion (BIC), tend to choose many spurious predictors and lack selection

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

consistency (see, e.g., Chen and Chen, 2008; Wang et al., 2009). To overcome this difficulty, the extended Bayesian information criterion (EBIC) is considered for linear regression models (Chen and Chen, 2008). In this study, we investigate the properties of an EBIC-type criterion for the proposed SARg model and establish its selection consistency. However, this process is technically challenging, because it involves a large deviation result on the first-order derivative of the log-likelihood function, which is a quadratic form of independent random variables. Accordingly, we need to carefully study the tail probability of the quadratic form of these variables, which presents a challenge when proving the consistency of the EBIC.

Third, despite its theoretical attractiveness, the QMLE experiences bottlenecks when considering large-scale networks (see, e.g., Barry and Pace, 1999; Smirnov and Anselin, 2001; Ma et al., 2020). The associated computational cost is very high when the network size n is large, because the computational complexity of calculating the determinant and the inverse is, in general, $O(n^3)$ (Trefethen and Bau, 1997; Barry and Pace, 1999). In reality, many social networks are enormous. For example, Facebook (www.facebook.com) has more than 1.79 billion daily active users. Thus, an alternative computationally feasible estimator is needed to alleviate the computational burden of the QMLE.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

This study contributes to the literature as follows. First, we show that under certain mild conditions, the QMLE is consistent and asymptotically normal when the number of parameters tends to infinity as the sample size increases. Second, we derive an upper bound for the tail probability of the quadratic form of independent random variables, and thus a large deviation result for the first-order derivatives. This approach is particularly useful for establishing the selection consistency of the EBIC for the proposed SARg model. Third, we propose an approximate maximum likelihood estimator (AMLE) for the SARg model. The basic idea is to approximate the determinant in the quasi log-likelihood function using a truncated-matrix Taylor expansion. Compared with the QMLE, the AMLE is computationally feasible and attractive for large-scale network analysis. We also investigate the theoretical properties of the AMLE.

The rest of this paper is organized as follows. Section 2 introduces the SARg model and the properties of its estimators. Additionally, an EBIC is proposed for selecting the best model. Section 3 and 4 present numerical studies and a real-data example, respectively. Finally, Section 5 concludes the paper.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

2. MODELS AND METHODOLOGY

2.1 Model and Notation

In addition to the spatial effect in the dependent variable, we consider a spatial spillover in the disturbance term. If two nodes are connected, their corresponding disturbances are more likely to be correlated. Therefore, the covariance matrix of the disturbance term may be affected by the network structure. Therefore, we model the covariance matrix as a polynomial function of the row-normalized adjacency matrix, with a prespecified upper order, and propose the following model:

$$\begin{aligned} Y &= \rho WY + X\boldsymbol{\alpha} + \varepsilon, \\ E(\varepsilon) = 0, \text{ and } \operatorname{cov}(\varepsilon) &= \Sigma(\boldsymbol{\beta}) = \beta_0 I_n + \beta_1 \widetilde{W} + \cdots + \beta_d \widetilde{W}^d. \end{aligned} \tag{2.1}$$

Here, $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_d)^\top \in \mathbb{R}^{d+1}$ is the coefficient vector used to characterize the dependence structure of the disturbance term ε, where β_k measures the contribution of $\widetilde{W}^k$, for $k = 1, \cdots, d$, and d is the upper order of the polynomial function. For completeness, define $\widetilde{W}^0 = I_n$. Here, to guarantee the symmetry of the covariance matrix, we use $\widetilde{W} = \frac{1}{2}(W + W^\top)$, instead of W. Notably, in model (2.1), β_k has practical meaning, because $\widetilde{W}^k = (\widetilde{w}_{ij}^{(k)}) \in \mathbb{R}^{n\times n}$ can capture the nodal relationship in terms of the k-step path. For example, if $w_{ij}^{(1)} > 0$, there exists a direct con-

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

nection from node i to node j, or from node j to node i. When $k = 2$, $w_{ij}^{(2)} = \sum_{l=1}^{n} w_{il}^{(1)} w_{lj}^{(1)}$ and $w_{ij}^{(2)} > 0$ imply that there exists a node l such that $w_{il}^{(1)} > 0$ and $w_{lj}^{(1)} > 0$. As a result, there exists a two-step path connecting node i and node j through intermediary node l. Accordingly, $w_{ij}^{(k)} > 0$ indicates that node i and node j are connected via a k-path. In summary, the terms $\widetilde{W}^k$, for $k = 1, \cdots, d$, contain useful information on network dependence, and thus may affect the covariance structure of the disturbance term. To assess the effect of $\widetilde{W}^k$ on Σ, we assign an appropriate weight β_k to each $\widetilde{W}^k$, and use β_k to measure the effect of the k-path nodal relationship on the covariance structure of ε. In equation (2.1), d can be regarded as a positive integer representing the maximal path length. As the network size $n \to \infty$, we allow d to increase slowly. For simplicity, we assume p is fixed, which is commonly considered in the extant literature (see, e.g., Lee, 2004; Kelejian and Prucha, 2010; Lee and Liu, 2010; Ma et al., 2020).

Remark 1. Two types of models, SARAR and SARMA, have been proposed to model the spatial effect in the disturbance term. The SARAR model assumes that the disturbances follow a spatial autoregressive process. This assumption links the covariance matrix of disturbances to the row-normalized adjacency matrix of infinite order. Alternatively, a spatial

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

moving average process can be imposed on the disturbances of the SAR model, in which the spatial correlation between disturbances is limited to a finite order of the row-normalized adjacency matrix. By linking the covariance matrix to the network adjacency matrix using matrix polynomial regression, our proposed model is sufficiently flexible to capture different types of spatial dependence. When the upper order d is set appropriately, the SARAR and SARMA models are special cases of our model framework.

Remark 2. In model (2.1), the upper order d needs to be specified a priori. In practice, we can set d to be a quite large integer. A similar approach has been used to specify the lag order in the vector autoregressive model and vector autoregressive moving average model (see, e.g., Hsu et al., 2008; Wilms et al., 2021). Moreover, as illustrated in our simulation results, our estimation results are robust to different choices of d.

2.2 Parameter Space

In this section, before estimating the parameters, we discuss the parameter space of the proposed model. Define the parameter vector $\boldsymbol{\theta} = (\rho, \boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top \in \mathbb{R}^{d+p+2}$ and the true parameter vector $\boldsymbol{\theta}_0 = (\rho_0, \boldsymbol{\beta}_0^\top, \boldsymbol{\alpha}_0^\top)^\top \in \mathbb{R}^{d+p+2}$. From model (2.1), we have $Y = (I_n - \rho W)^{-1}(X\boldsymbol{\alpha} + \varepsilon)$. Because W is a row-normalized adjacency matrix, the largest eigenvalue of W is one (Banerjee

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

et al., 2004). A sufficient condition to ensure the invertibility of $I_n - \rho W$ is $|\rho| < 1$. Accordingly, we assume $|\rho| < 1$ throughout this paper. In addition, the covariance matrix $\Sigma = cov(\varepsilon)$ is required to be positive definite. Following Fan and Lv (2008), we require that there exist positive constants $\tau_{\min}$ and $\tau_{\max}$ that satisfy

$$0 < \tau_{\min} < \lambda_{\min}(\Sigma) < \lambda_{\max}(\Sigma) < \tau_{\max} < \infty,$$

where $\lambda_{\min}(\Sigma)$ and $\lambda_{\max}(\Sigma)$ are the minimal and maximal eigenvalues, respectively, of the covariance matrix Σ. Using a spectral decomposition, we decompose $\widetilde{W}$ as $\widetilde{W} = UDU^\top$, where U is an orthonormal matrix, $D = \text{diag}(\lambda_1, \cdots, \lambda_n)$, and λ_j is the jth largest eigenvalue of $\widetilde{W}$. Therefore, we have

$$U^\top \Sigma U = \text{diag}(\sum_{i=0}^{d} \beta_i \lambda_j^i). \tag{2.2}$$

Define $a_{\max} = \max\{|\lambda_1|, |\lambda_n|\}$. As a result, we construct the parameter space as follows:

$$\Theta = \Big\{ \boldsymbol{\theta} = (\rho, \boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top : |\rho| < 1, |\alpha_i| < \infty \text{ for } i = 1, \cdots, p,$$
$$\tau_{\min} < \sum_{i=0}^{d} \beta_i \lambda^i < \tau_{\max} \text{ for any } \lambda \in [-a_{\max}, a_{\max}] \Big\}.$$

If $\boldsymbol{\theta} \in \Theta$, all the eigenvalues of the covariance matrix Σ are positive and bounded. Clearly, Θ is a nonempty set, because it contains $\big\{(\rho, \boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top)^\top : \rho = 0, \boldsymbol{\alpha} = 0, \beta_0 > \tau_{\min}, \beta_k = 0, \text{ for any } k = 1, \cdots, d\big\}$ as a nontrivial

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

subspace. Moreover, the parameter space Θ is an open set. For detailed results and a proof, see Proposition 1 in the Supplementary Material, S.2.

2.3 QMLE

Here, we use the QMLE method to estimate the unknown parameters in model (2.1). Because $E(Y) = (I_n - \rho W)^{-1} X\boldsymbol{\alpha}$ and $\mathrm{cov}(Y) = (I_n - \rho W)^{-1}\Sigma(I_n - \rho W^\top)^{-1}$, the quasi-log-likelihood function, ignoring the constant, is

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{2}\log|\Sigma| + \frac{1}{2}\log|I_n - \rho W|^2 - \frac{1}{2}(Y - \rho WY - X\boldsymbol{\alpha})^\top \Sigma^{-1}(Y - \rho WY - X\boldsymbol{\alpha}).$$

Consequently, the QMLE of $\boldsymbol{\theta}_0$ can be obtained as $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}\in\Theta} \mathcal{L}(\boldsymbol{\theta})$. This problem can be solved using the Newton–Raphson method with the Armijo line search. To avoid the problem of local optima, we recommend using a random initialization method in practice (see, e.g., Wang et al., 2021). Specifically, we can generate many randomized initial values, and then use the estimation that yields the maximum value of the objective function. Our simulation studies (unreported) demonstrate that this method works satisfactorily and provides stable estimation results for different sets of randomized initial values.

To investigate the theoretical properties of the QMLE, we first introduce some notation that will be used to develop the theoretical distributions of $\hat{\boldsymbol{\theta}}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

For any generic vector $x = (x_1, \cdots, x_n)^\top \in \mathbb{R}^n$, the vector q-norm is defined as $\|x\|_q = (\sum_{i=1}^n |x_i|^q)^{1/q}$. For any generic matrix $M = (m_{ij}) \in \mathbb{R}^{n\times n}$, define $\|M\|_q$ as the matrix q-norm, for $q = 1, 2, \infty$, and the Frobenius norm $\|M\|_F = \sum_i \sum_j m_{ij}^2$. Moreover, define $|M|_\infty = \|\operatorname{vec}(M)\|_\infty$. Denote $\mathcal{I}_n(\boldsymbol{\theta}_0)$ as the Fisher information matrix of the quasi-log-likelihood, and $\mathcal{I}_n(\boldsymbol{\theta}_0) + \mathcal{J}_n(\boldsymbol{\theta}_0)$ as the variance of the score function, explicit forms of which are shown in the Supplementary Material, S.1. We next state several conditions.

(C1) As $n \to \infty$, we assume $d = O(n^\kappa)$, for some $0 < \kappa < 1/4$.

(C2) Define $Z = \{\Sigma(\boldsymbol{\beta}_0)\}^{-1/2}\varepsilon = (z_1, \cdots, z_n)^\top$. Assume $z_1, \cdots, z_n$ are independent subGuassian random variables satisfying $E(z_i) = 0$, $E(z_i^2) = 1$, $E(z_i^k) = \mu_k$ for $k = 3, 4$, and $E\exp(z_i^2/t^2) \le 2$ for any $t \ge K$, where μ_k $(k = 3, 4)$ and $K > 0$ are finite constants.

(C3) (i) Define τ as the number of distinct eigenvalues of $\widetilde{W}$ that satisfies $d \le \tau \le n$ as $n \to \infty$; (ii) Assume $\sup_{n\ge1} \|W\|_1 < \infty$, $\sup_{n\ge1} \|W\|_\infty < \infty$, and $\sup_{n\ge1, 1\le k\le d} \|\widetilde{W}^k\|_1 < \infty$.

(C4) Assume there exists a large enough open subset $\widetilde{\Theta} \subset \Theta$ that contains the true parameter $\boldsymbol{\theta}_0$, such that $\sup_{n\ge1} \|(I_n - \rho W)^{-1}\|_1 < \infty$, $\sup_{n\ge1} \|(I_n - \rho W)^{-1}\|_\infty < \infty$, $\sup_{n\ge1} \|\big(\Sigma(\boldsymbol{\beta})\big)^{1/2}\|_1 < \infty$, and

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

$\sup_{n\geq 1} \|\big(\Sigma(\boldsymbol{\beta})\big)^{-1/2}\|_1 < \infty$, for any $\boldsymbol{\theta} \in \widetilde{\Theta}$.

(C5) The auxiliary information X satisfies $\sup_{n>1} |X|_\infty < \infty$.

(C6) Assume $\|\mathcal{I}_n(\boldsymbol{\theta}_0) - \mathcal{I}(\boldsymbol{\theta}_0)\|_F = o(1)$ and $\|\mathcal{J}_n(\boldsymbol{\theta}_0) - \mathcal{J}(\boldsymbol{\theta}_0)\|_F = o(1)$. We further assume $c_1 < \lambda_{\min}\big(\mathcal{I}(\boldsymbol{\theta}_0)\big) < \lambda_{\max}\big(\mathcal{I}(\boldsymbol{\theta}_0)\big) = O(d)$ and $c_2 < \lambda_{\min}\big(\mathcal{I}(\boldsymbol{\theta}_0) + \mathcal{J}(\boldsymbol{\theta}_0)\big) < \lambda_{\max}\big(\mathcal{I}(\boldsymbol{\theta}_0) + \mathcal{J}(\boldsymbol{\theta}_0)\big) = O(d)$, for some finite positive constants c_1 and c_2.

Condition (C1) allows d to diverge as the network size $n \to \infty$. Condition (C2) is a moment condition that is much weaker than commonly used distribution assumptions; see, for example, Zhou et al. (2017). We assume that the noise terms are subGuassian random variables in order to bound the tail probability of the first derivatives of the log-likelihood function. This is an essential condition to establish the selection consistency of the EBIC. Condition (C3)(i) assumes $\widetilde{W}$ has divergent distinct eigenvalues, which ensures the identifiability of the parameters $\boldsymbol{\beta}$. If the number of distinct eigenvalues of $\widetilde{W}$ is smaller than d, some parameters β_k in model (2.1) will be unidentifiable. Conditions (C3)(ii) and (C4) are standard regular conditions that limit the spatial correlation to a manageable degree. Similar conditions appear in Lee (2004) and Kelejian and Prucha (2010). Condition (C5) assumes that the regressors are bounded, as is common in

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

the literature; see, for example, Lee (2004). Condition (C6) is a law of large numbers-type condition that guarantees the convergence of the Fisher information matrix and the variance of the score function. Similar conditions can be found in Lee (2007) and Zou et al. (2021). Based on the above conditions, the theoretical distribution of $\hat{\boldsymbol{\theta}}$ is given in the following theorem.

Theorem 1. *Let* $\hat{\boldsymbol{\theta}} = (\hat{\rho}, \hat{\boldsymbol{\beta}}^\top, \hat{\boldsymbol{\alpha}}^\top)^\top$ *be the QMLE of* $\boldsymbol{\theta}_0 = (\rho_0, \boldsymbol{\beta}_0^\top, \boldsymbol{\alpha}_0^\top)^\top$. *Under Conditions (C1)–(C6), we have*

(a) Consistency:

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 = O_p(\sqrt{d/n}).$$

(b) Asymptotic normality:

$$\sqrt{n/d}\boldsymbol{t}^\top(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, s^2),$$

for any generic vector $\boldsymbol{t} = (t_1, \cdots, t_{d+p+2})^\top \neq 0 \in \mathbb{R}^{d+p+2}$ *satisfying* $\|\boldsymbol{t}\|_2 < C$, $s^2 = d^{-1}\boldsymbol{t}^\top\mathcal{I}^{-1}(\boldsymbol{\theta}_0)\big(\mathcal{I}(\boldsymbol{\theta}_0)+\mathcal{J}(\boldsymbol{\theta}_0)\big)\mathcal{I}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{t}$, *and* $d^{-1}\boldsymbol{t}^\top\big(\mathcal{I}(\boldsymbol{\theta}_0)+\mathcal{J}(\boldsymbol{\theta}_0)\big)\boldsymbol{t}$ *is a finite positive constant.*

The proof of Theorem 1 is given in the Supplementary Material. According to Theorem 1, the QMLE is a $\sqrt{n/d}$-consistent estimator. The asymptotic distribution of $\hat{\boldsymbol{\theta}}$ remains valid when d diverges. In addition, when d is finite, the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ simplifies to

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N\big(0, \mathcal{I}^{-1}(\boldsymbol{\theta}_0)\big(\mathcal{I}(\boldsymbol{\theta}_0) + \mathcal{J}(\boldsymbol{\theta}_0)\big)\mathcal{I}^{-1}(\boldsymbol{\theta}_0)\big).$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

In practice, to make valid statistical inferences, we need to estimate $\mathcal{I}(\boldsymbol{\theta}_0)$ and $\mathcal{J}(\boldsymbol{\theta}_0)$ consistently, which we can do using the sample-based counterparts. Specifically, we can estimate $\mathcal{I}(\boldsymbol{\theta}_0)$ and $\mathcal{J}(\boldsymbol{\theta}_0)$ by $\mathcal{I}_n(\hat{\boldsymbol{\theta}})$ and $\mathcal{J}_n(\hat{\boldsymbol{\theta}})$, respectively, where $\hat{\mu}_k = n^{-1}\sum_{i=1}^n \hat{z}_i^k$ $(k = 3, 4)$ and $\hat{Z} = (\hat{z}_1, \cdots, \hat{z}_n) = \{\Sigma(\hat{\boldsymbol{\beta}})\}^{-1/2}(Y - \hat{\rho}WY - X\hat{\boldsymbol{\alpha}}^\top)$. Notably, when no spatial effect exists in the disturbance term, that is, $d = 0$, the asymptotic distribution of the QMLE is the same as that studied by Lee (2004) for the pure SAR model.

2.4 Model Selection

When d is large, it is essential to identify which $\widetilde{W}^k$ terms are relevant. Therefore, we propose an EBIC-type model selection method, motivated by Chen and Chen (2008), for diverging d. Specifically, we define the true model as $\mathcal{S}_0 = \{k \geq 0 : \beta_{0k} \neq 0\}$, which collects the index of relevant terms $\widetilde{W}^k$. We assume that the number of relevant terms is finite, that is, $|\mathcal{S}_0| < \infty$. The full model is defined as $\mathcal{S}_F = \{0, 1, \cdots, d\}$. Let $\mathcal{S} \subset \mathcal{S}_F$ be a candidate model. The corresponding coefficient vector is defined as $\boldsymbol{\theta}(\mathcal{S}) = (\rho, \boldsymbol{\beta}(\mathcal{S})^\top, \boldsymbol{\alpha}^\top)^\top$, $\boldsymbol{\beta}(\mathcal{S}) = (\beta_i, i \in \mathcal{S}) \in \mathbb{R}^{|\mathcal{S}|}$, and $|\mathcal{S}|$ represents the size of model $\mathcal{S}$. Then, we propose the following EBIC:

$$\text{EBIC}_\gamma(\mathcal{S}) = -2\mathcal{L}\big\{\hat{\boldsymbol{\theta}}(\mathcal{S})\big\} + |\mathcal{S}|\log(n) + 2\gamma|\mathcal{S}|\log(d),$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

for some $\gamma \geq 0$, where $\hat{\boldsymbol{\theta}}(\mathcal{S})$ is the QMLE for model $\mathcal{S}$ and γ is a scale parameter. Then, the best model selected by the EBIC is $\mathcal{S}_{\text{EBIC}} = \arg\min_{\mathcal{S}\subset\mathcal{S}_F} \text{EBIC}_\gamma(\mathcal{S})$.

Similarly to Chen and Chen (2008), we define the collection of underfitted models as $\mathscr{A}_1 = \{\mathcal{S} : \mathcal{S}_0 \not\subset \mathcal{S}, |\mathcal{S}| \leq cd_0\}$, and the collection of overfitted models as $\mathscr{A}_2 = \{\mathcal{S} : \mathcal{S}_0 \subset \mathcal{S}, |\mathcal{S}| \leq cd_0\}$, where $d_0 = |\mathcal{S}_0|$, and $c > 1$ is a fixed constant defined in Condition (C8). The following conditions are needed before establishing the properties of the EBIC.

(C7) $\min_{i\in\mathcal{S}_0} |\beta_{i0}| \geq Cn^{-1/4}$, for some constant $C > 0$.

(C8) Define $\ddot{\mathcal{L}}(\boldsymbol{\theta}) = \partial^2\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$. For any given $\delta > 0$, there exist constants $\xi > 0$ and $c > 1$ such that, when n is sufficiently large,

$$(1-\delta)\mathcal{I}_n\{\boldsymbol{\theta}_0(\mathcal{S}\cup\mathcal{S}_0)\} \leq -n^{-1}E\ddot{\mathcal{L}}\{\boldsymbol{\theta}(\mathcal{S}\cup\mathcal{S}_0)\} \leq (1+\delta)\mathcal{I}_n\{\boldsymbol{\theta}_0(\mathcal{S}\cup\mathcal{S}_0)\},$$

for all $\mathcal{S}$ such that $|\mathcal{S}| \leq cd_0$ and $\|\boldsymbol{\theta}(\mathcal{S}\cup\mathcal{S}_0) - \boldsymbol{\theta}_0(\mathcal{S}\cup\mathcal{S}_0)\|_2 \leq \xi$.

Condition (C7) places a requirement on the order of the nonzero coefficients, and is popular in classic linear regression models. Condition (C8) extends (C6) to a small neighborhood of the true parameter $\boldsymbol{\theta}_0(\mathcal{S}\cup\mathcal{S}_0)$. Similar conditions can be found in Chen and Chen (2012) and Chen and Luo (2013). Based on these conditions, the selection consistency of the EBIC can be obtained as follows.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

Theorem 2. *Under Conditions (C1)–(C8), as $n \to \infty$, we have*

$$P\Big\{ \min_{\mathcal{S} \in \mathscr{A}_1} EBIC_\gamma(\mathcal{S}) \leq EBIC_\gamma(\mathcal{S}_0) \Big\} \to 0,$$

for any $\gamma \geq 0$ satisfying $\gamma = o(n^{1/2}/\log d)$;

$$P\Big\{ \min_{\mathcal{S} \in \mathscr{A}_2} EBIC_\gamma(\mathcal{S}) \leq EBIC_\gamma(\mathcal{S}_0) \Big\} \to 0,$$

for any $\gamma \geq 0$ satisfying $\gamma > \frac{1}{\min\{c_5^*, c_6^*\}} - \frac{5}{2}$, where c_5^* and c_6^* are defined in the Supplementary Material.

The above theorem implies that the EBIC can determine the true model consistently, as long as $n \to \infty$. In practice, we recommend using $\gamma = 0.5$, which performs well in our simulations. To implement the EBIC, we apply the forward-backward selection procedure (see, e.g., Zhang, 2009; Ma et al., 2021), which reduces the computational complexity from $O(2^d)$ to $O(d^2)$. Thus, the EBIC is applicable even for divergent d.

2.5 AMLE

Despite the theoretical attractiveness of the QMLE, it is computationally infeasible when the network size n is large, because each iteration calculates the log-determinant and the inverse of the matrix $I_n - \rho W$.

To avoid having to calculate $\log|I_n - \rho W|$, the following well-known formula of matrix powering expansion is commonly used (see, e.g., Martin,

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

1993; Barry and Pace, 1999; Boutsidis et al., 2017). That is,

$$\log|I_n - \rho W| = -\sum_{k=1}^{\infty} \frac{\mathrm{tr}(\rho^k W^k)}{k}.$$

Notably, this expansion holds even when W is asymmetric; therefore, we can approximate $\log|I_n - \rho W|$ using a truncated-matrix Taylor expansion. The approximate quasi-log-likelihood function is

$$\mathcal{L}_a(\boldsymbol{\theta}) = -\frac{1}{2}\log|\Sigma| - \sum_{k=1}^{m} \frac{\mathrm{tr}(\rho^k W^k)}{k} - \frac{1}{2}(Y - \rho WY - X\boldsymbol{\alpha})^\top \Sigma^{-1}(Y - \rho WY - X\boldsymbol{\alpha}),$$

where the constant independent of $\boldsymbol{\theta}$ is ignored. Maximizing $\mathcal{L}_a(\boldsymbol{\theta})$ leads to the AMLE $\hat{\boldsymbol{\theta}}_{\mathrm{a}} = \arg\max_{\boldsymbol{\theta}\in\Theta} \mathcal{L}_a(\boldsymbol{\theta})$, and we can employ the Newton–Raphson method to find the AMLE. Although computing the trace for the large-scale matrix is still computationally expensive, this procedure needs to be conducted only once, in advance. Furthermore, the log-determinant and the inverse of Σ can be calculated easily after conducting the spectral decomposition of the matrix $\widetilde{W}$. Compared with the QMLE, the AMLE is computationally efficient, according to our numerical studies. We next establish the theoretical properties for the AMLE.

Theorem 3. *Let* $\hat{\boldsymbol{\theta}}_a = (\hat{\rho}_a, \hat{\boldsymbol{\beta}}_a^\top, \hat{\boldsymbol{\alpha}}_a^\top)^\top$ *be the AMLE of* $\boldsymbol{\theta}_0 = (\rho_0, \boldsymbol{\beta}_0^\top, \boldsymbol{\alpha}_0^\top)^\top$. *Under Conditions (C1)–(C6), if* $m\rho_0^m = o(n^{-1/2})$, *we have*

(a) Consistency:

$$\|\hat{\boldsymbol{\theta}}_a - \boldsymbol{\theta}_0\|_2 = O_p(\sqrt{d/n}).$$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

(b) Asymptotic normality:

$$\sqrt{n/d}\boldsymbol{t}^\top(\hat{\boldsymbol{\theta}}_a - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, s^2),$$

for any generic vector $\boldsymbol{t} = (t_1, \cdots, t_{d+p+2})^\top \neq 0 \in \mathbb{R}^{d+p+2}$ *satisfying* $\|\boldsymbol{t}\|_2 < C$, $s^2 = d^{-1}\boldsymbol{t}^\top\mathcal{I}^{-1}(\boldsymbol{\theta}_0)\big(\mathcal{I}(\boldsymbol{\theta}_0)+\mathcal{J}(\boldsymbol{\theta}_0)\big)\mathcal{I}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{t}$, *and* $d^{-1}\boldsymbol{t}^\top\big(\mathcal{I}(\boldsymbol{\theta}_0)+\mathcal{J}(\boldsymbol{\theta}_0)\big)\boldsymbol{t}$ *is a finite positive constant.*

According to Theorem 3, the asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_{\mathrm{a}}$ is the same as that of $\hat{\boldsymbol{\theta}}$, mainly because of the limiting condition $m\rho_0^m = o(n^{-1/2})$. Intuitively, this technical condition holds when either m goes to infinity or ρ_0 converges to zero. In practice, when ρ_0 is fixed, m is a tuning parameter that can be selected using the BIC. Specifically, the optimal m is selected using $m_{BIC} = \arg\min_{1\leq m\leq m_{\max}} -2\mathcal{L}_a\big\{\hat{\boldsymbol{\theta}}_a(m)\big\} + m\log(n)$, where $m_{\max}$ is a pre-specified large integer, and $\hat{\boldsymbol{\theta}}_a(m)$ is the corresponding AMLE for truncated order m.

3. SIMULATION STUDIES

3.1 Simulation Settings

To evaluate the finite-sample performance of the proposed model, we conduct Monte Carlo simulations in various settings. We consider three types of network models, each generating its own mechanisms for the adjacency

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

matrix A.

Example 1. We first consider a simple ER model (Erdös and Rényi, 1959); specifically, the diagonal elements $a_{ii}(i = 1, \cdots, n)$ of the adjacency matrix A are set to zero. The off-diagonal elements $a_{ij}(i \neq j)$ are independent and identically generated from Bernoulli distributions with probability $n^{-0.8}$.

Example 2. To reflect the clustering property observed in many real networks, we follow Nowicki and Snijders (2001) and generate a stochastic block model, as follows. We assume there exist five blocks, and that each node is randomly appointed to one of the five blocks. If node i and node j $(i \neq j)$ are from the same block, $P(a_{ij} = 1) = 10n^{-1}$; otherwise, $P(a_{ij} = 1) = n^{-1}$. Similarly, we set $a_{ii} = 0$, for all $i = 1, \cdots, n$.

Example 3. In many social networks, there exist some nodes with extremely large in-degree values. This motivates us to simulate a power-law-type network structure to mimic such a highly skewed in-degree distribution. Thus, we first generate the in-degree of each node using a discrete power-law distribution (see, e.g., Clauset et al., 2009) with probability mass function $ck^{-\alpha}$, where c is a constant, k is the in-degree of the node, and $\alpha = 2.5$. For each node i, with $i = 1, \cdots, n$, we randomly sample $\min(k, n)$ nodes, without replacement, and define this set as S_i. If $j \in S_i$

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

$(j = 1, \cdots, n)$, then $a_{ij} = 1$; otherwise, $a_{ij} = 0$. Finally, we force $a_{ii} = 0$ for each node i $(i = 1, \cdots, n)$.

Given the adjacency matrix A, W can be computed by row-normalizing the matrix A. Set the auxiliary information matrix $X = (X_1, \cdots, X_n)^\top \in \mathbb{R}^{n\times 2}$. For $X_i = (x_{i1}, x_{i2})^\top \in \mathbb{R}^2$, let $x_{i1} = 1$ and x_{i2} be independent and identically generated from a standard normal distribution $N(0, 1)$. Then, the response vector Y is generated by $Y = (I_n - \rho W)^{-1}(X\boldsymbol{\alpha} + \varepsilon)$. Here, we set $\rho = 0.2$, $\boldsymbol{\alpha} = (3, 6)^\top$, $\varepsilon = \Sigma^{1/2}Z$, with $\Sigma = 0.1I_n + 0.3\widetilde{W} + 0.7\widetilde{W}^2 + 1.5\widetilde{W}^3 + 2\widetilde{W}^4$, and $\widetilde{W} = (W + W^\top)/2$. Each element of $Z = (z_1, \cdots, z_n)^\top$ is independent and identically generated from either a standard normal distribution or a mixture distribution $0.9N(0, 5/9) + 0.1N(0, 5)$. To study the consistency of the proposed EBIC, we consider different full model sizes with $d = 6, 9, 12$ and the size of the true model $d_0 = 4$. We set $\gamma = 0.5$ to select the best model, because other settings yield similar conclusions. Three network sizes (i.e., $n = 500, 1000, 1500$) are considered, and each setting is replicated randomly 500 times.

3.2 Simulation Results

Define $\boldsymbol{\theta_0} = \left(\theta_{0,1}, \cdots, \theta_{0,d+p+2}\right)^\top \in \mathbb{R}^{d+p+2}$ as the true parameter, and $\hat{\boldsymbol{\theta}}^{(k)} = \left(\hat{\theta}_1^{(k)}, \cdots, \hat{\theta}_{d+p+2}^{(k)}\right)^\top \in \mathbb{R}^{d+p+2}$ as the estimator of $\boldsymbol{\theta_0}$ at the kth

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

simulation replication. For parameter $\theta_{0,j}$, define the bias as $\text{BIAS}_j = 500^{-1}\sum_k \left(\hat{\theta}_j^{(k)} - \theta_{0,j}\right)$ and the standard deviation as $\text{SD}_j = \left\{500^{-1}\sum_k \left(\hat{\theta}_j^{(k)} - \bar{\theta}_j\right)^2\right\}^{1/2}$, with $\bar{\theta}_j = 500^{-1}\sum_k \hat{\theta}_j^{(k)}$. In each iteration, we also calculate the standard error estimate $\text{SE}_j^{(k)}$, based on Theorem 1 for the QMLE and Theorem 3 for the AMLE. Then, the average of the estimated standard error is $\text{SE}_j = 500^{-1}\sum_k \text{SE}_j^{(k)}$. The average CPU time (in seconds) consumed by both estimators (i.e., QMLE and AMLE) is recorded using a PC with 3.40 GHz and 288 GB RAM.

To assess the performance of the proposed EBIC, we calculate several evaluation indices. Define $\mathcal{S}_T$ and $\hat{\mathcal{S}}$ as the index sets of the true model and the selected model, respectively. The index set $\mathcal{S}_T^c$ represents the complementary set of the index set $\mathcal{S}_T$. Then, we can calculate the average size of the selected model $|\hat{\mathcal{S}}|$, average percentage of the correctly fitted model (CF) $I(\hat{\mathcal{S}} = \mathcal{S}_T)$, average true positive rate (TPR) $|\hat{\mathcal{S}} \cap \mathcal{S}_T|/|\mathcal{S}_T|$, and average false positive rate (FPR) $|\hat{\mathcal{S}} \cap \mathcal{S}_T^c|/|\mathcal{S}_T^c|$. The simulation results are presented in Tables 1–2. To save space, the results for Example 1 when the noise term follows the mixture distribution and the results for Examples 2 and 3 are relegated to the Supplementary Material, S.5.1, because they yield similar conclusions.

According to Table 1, the BIAS values are small and the standard

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

deviations (SD) are close to the standard error estimate (SE) for the QMLE. Furthermore, the SD and SE values approach zero as the network size n increases. These results are consistent with the asymptotic theory given in Theorem 1. The AMLE shows the same pattern as that of the QMLE. Specifically, the SD values of the QMLE and AMLE are similar. In addition, the SD values of the AMLE are robust to different choices of truncated order m. However, the AMLE with the optimal m selected using the BIC outperforms the other methods: it has the smallest SD values, and its SD values and SE values are almost equal. In addition, the AMLE is computationally more efficient than the QMLE.

Table 2 shows the performance of the EBIC. From Table 2, we can draw the following conclusions. First, for any fixed d, the performance of the EBIC improves as the network size n increases. The percentage of correctly fitted models increases with the dimension n. Second, the EBIC performs well when the network size n is large. Specifically, when $n = 1,500$, the average size of the selected model approaches that of the true model, the average positive selection rate increases to one, and the average false discovery rate decreases to zero. These findings are consistent with our theoretical results, and indicate that the EBIC can consistently select the true model.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

4. REAL-DATA ANALYSIS

To demonstrate the practical usefulness of the proposed method, we study the spillover effect of mutual funds, which is crucial for both fund managers and general investors (see, e.g., Spitz, 1970; Nanda et al., 2004). By applying the proposed model, we simultaneously consider the spillover effects of the response variable and the disturbance term from a network perspective.

We collect data on actively managed open-ended mutual funds in the first quarter of 2021 from the WIND financial database, an authoritative database for the Chinese financial market. After removing funds with missing observations, 261 funds remain. The response variable, cash inflow rate Y_i, is defined as $Y_i = \big(TA_i^{new} - TA_i^{old}(1 + r_i^{new})\big)/TA_i^{new}$, where TA_i^{new} and TA_i^{old} are the total net assets of fund i at the end of the first quarter of 2021 and at the end of 2020, respectively, and r_i^{new} is the fund return during the first quarter of 2021 (Nanda et al., 2004). To avoid the effect of outliers caused by cash flow, we remove the 1% of funds with the highest cash inflow rate, leaving 258 funds in our study.

Motivated by Spitz (1970) and Sawicki and Finn (2002), we next introduce the intercept term and four related exogenous covariates to explore their effect on cash flow: X_{i1} is the logarithm of the total net assets of fund i at the end of 2020; X_{i2} is the logarithm of the fund's age; X_{i3} records the

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

fund's raw return at the end of 2020; and X_{i4} is the risk-adjusted return of fund i, which is measured by the intercept of Carhart's (1997) four-factor model. Histograms of the cash inflow rate and four exogenous covariates are depicted in Figure S.1 in the Supplementary Material, S.6. Before starting our analysis, we standardize all the exogenous covariates.

To explore the network effect of mutual funds on cash flow, we construct the adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{n\times n}$, as follows. As described in Pareek (2012), we define $a_{ij} = a_{ji} = 1$ if two funds i and j allocate at least 5% of their portfolios to the same stock; otherwise, $a_{ij} = a_{ji} = 0$. Additionally, we require the diagonal elements of the adjacency matrix A to be equal to zero, for completeness. Next, we study the spillover effect of mutual funds based on the cash flow and the observed network structure.

We first use the proposed EBIC method to select the best candidate model, given the size of the full model $d = 6$ and the tuning parameter $\gamma = 0.5$. This leads to $\mathcal{S}_{EBIC} = \{0, 1\}$; thus, the covariance matrix is related to matrices I_n and W. Consequently, we obtain the QMLE for the parameters ρ, $\boldsymbol{\beta}$, and $\boldsymbol{\alpha}$ and the corresponding log-likelihood values. The final results are given in Table 3. For comparison purposes, we present the QMLE estimates of the SARAR(1,1) model and the SARMA(1,1) model. In addition, we present the QMLE of the pure SAR model, which ignores

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

the spatial effect of the disturbance term. We draw the following conclusions from the estimation results. First, compared with other models, the proposed model fits the data best, because it yields the largest log-likelihood value (see, e.g., Keller and Shiue, 2007). Second, when considering the spatial effect of the disturbances, $\hat{\rho}$ is positive and significant at the 5% significance level, which suggests a positive and significant spillover effect between cash flows. Third, in addition to the spatial effect of mutual funds on cash flow, a local spatial effect exists in the disturbance term, because we observe a significant negative spatial dependence for directly connected neighbors at the 5% level of significance. Fourth, the estimates for the four exogenous covariates are similar to those of the SARAR(1,1) model, SARMA(1,1) model, and pure SAR model. The age, raw return, and risk-adjusted return of funds have a significant effect on cash flow, after adjusting for the spatial effect between cash flows. Furthermore, to determine whether the proposed model fits the data adequately well, we present a QQ-plot for the estimated error term $\hat{Z} = \hat{\Sigma}(\hat{\boldsymbol{\beta}})^{-1/2}(Y - \hat{\rho}WY - X\hat{\boldsymbol{\alpha}})$, where $\hat{\rho}$, $\hat{\boldsymbol{\alpha}}$, and $\hat{\boldsymbol{\beta}}$ are estimated values. If the spatial dependence between the response variables and between the disturbance terms is modeled well, we expect the error term Z to be standard normal. As shown in Figure 1, each element of the error term $\hat{Z}$ is roughly normal, implying that the

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

proposed model fits the data well. Accordingly, it is necessary to consider the spatial correlations between disturbances.

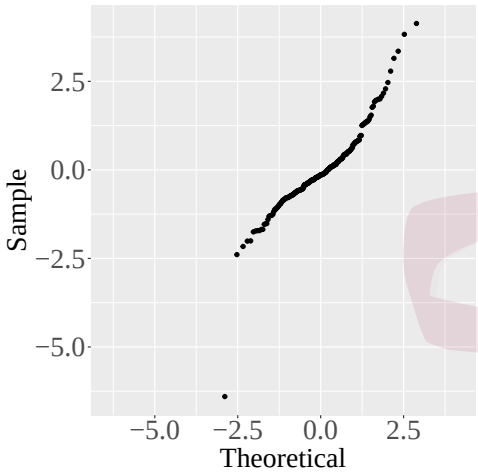

Figure 1: QQ-plot for the error term $\hat{Z}$.

5. CONCLUSION

In this paper, we have proposed a novel method called the spatial autoregressive model with generalized spatial disturbances to simultaneously model the spatial effects between the response variables and between the disturbance terms. We use the QMLE to estimate the model, and establish the asymptotic properties of the QMLE without imposing any distribution assumptions. Because the QMLE is computationally infeasible for large-scale network data, we propose an AMLE and establish its asymptotic

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

properties. Our numerical studies show that the AMLE is more computationally efficient than the QMLE, and is applicable to large-scale network data sets. Because the number of matrix predictors is diverging, we propose using an EBIC-type method to select the best model, and demonstrate the consistency of this criterion. We demonstrate the performance of the proposed method using simulation studies and a real-data example.

Several interesting topics are available for future research. First, it is reasonable to allow the coefficients to change over time. Therefore, we can extend our model to a semiparametric or nonparametric model, and use the information of neighboring time points. Second, in our model, the adjacency matrix is observed and fixed; it would be of interest to allow the adjacency matrix to be random or partially observed. Third, our proposed model requires that the dependent variable be continuous. Thus, modeling the spatial effects between discrete dependent variables is also an interesting topic. Fourth, similarly to Wang et al. (2007), it would be meaningful to develop a criterion for choosing the optimal γ for the EBIC. Lastly, we would like to develop a test to assess the adequacy of the proposed model using high-dimensional structured covariance matrix tests (see, e.g., Zhong et al., 2017). However, this is challenging because of the complex dependence in the dependent variable and in the disturbance term. We believe these

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

efforts would extend the usefulness of our proposed method.

Supplementary Material

This online Supplementary Material comprises six parts. Section S.1 provides the explicit forms of $\mathcal{I}_n(\boldsymbol{\theta}_0)$ and $\mathcal{J}_n(\boldsymbol{\theta}_0)$. Section S.2 presents the results of Proposition 1, as well as its proof. Sections S.3–S.4 present six useful lemmas and the proof of Theorems 1–3, respectively. Additional simulation results and a descriptive analysis for real data are presented in Sections S.5–S.6, respectively.

Acknowledgements

The authors sincerely thank the Editor, Associate Editor, and two reviewers for their thoughtful and constructive suggestions. Authors are listed in alphabetical order. This research was supported by the National Natural Science Foundation of China (72071169, 71991472, 12171395,11931014, 11971404, 71988101), National Social Science Foundation of China (21&ZD146), National Bureau of Statistics of China (2022LZ34), the Fundamental Research Funds for the Central Universities (JBK1806002), and the Joint Lab of Data Science and Business Intelligence at Southwestern University of Finance and Economics.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Boston: Kluwer Academic Publishers.

Anselin, L. and R. Florax (1995). *New Directions in Spatial Econometrics*. London: Springer.

Arnold, M., S. Stahlberg, and D. Wied (2013). Modeling different kinds of spatial dependence in stock returns. *Empirical Economics 44*(2), 761–774.

Badinger, H. and P. Egger (2011). Estimation of higher-order spatial autoregressive cross-section models with heteroscedastic disturbances. *Papers in Regional Science 90*(1), 213–235.

Baltagi, B. H. and G. Bresson (2011). Maximum likelihood estimation and lagrange multiplier tests for panel seemingly unrelated regressions with spatial lag and spatial errors: An application to hedonic housing prices in paris. *Journal of Urban Economics 69*(1), 24–42.

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton, FL : Chapman & Hall/ CRC Press.

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

Behrens, K., C. Ertur, and W. Koch (2012). 'Dual' gravity: Using spatial econometrics to control for multilateral resistance. *Journal of Applied Econometrics 27*(5), 773–794.

Boutsidis, C., P. Drineas, P. Kambadur, E.-M. Kontopoulou, and A. Zouzias (2017). A randomized algorithm for approximating the log determinant of a symmetric positive definite

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

matrix. *Linear Algebra and its Applications 533*, 95–117.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance 52*(1), 57–82.

Catania, L. and A. G. Billé (2017). Dynamic spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Applied Econometrics 32*(6), 1178–1196.

Chang, X., D. Huang, and H. Wang (2019). A popularity scaled latent space model for large-scale directed social network. *Statistica Sinica 29*(3), 1277–1299.

Chen, J. and Z. Chen (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika 95*(3), 759–771.

Chen, J. and Z. Chen (2012). Extended BIC for small-n-large-P sparse GLM. *Statistica Sinica 22*(2), 555–574.

Chen, Z. and S. Luo (2013). Selection consistency of EBIC for GLIM with non-canonical links and diverging number of parameters. *Statistics and Its Interface 6*(2), 275–284.

Clauset, A., C. R. Shalizi, and M. E. Newman (2009). Power-law distributions in empirical data. *SIAM review 51*(4), 661–703.

Cliff, A. D. and J. K. Ord (1973). *Spatial Autocorrelation*. London: Pion.

Delgado, M. A. and P. M. Robinson (2015). Non-nested testing of spatial correlation. *Journal of Econometrics 187*(1), 385–401.

Dogan, O. (2015). Heteroskedasticity of unknown form in spatial autoregressive models with a

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

moving average disturbance term. *Econometrics 3*(1), 101–127.

Dogan, O. and S. Taspinar (2013). GMM estimation of spatial autoregressive models with moving average disturbances. *Regional Science and Urban Economics 43*(6), 903–926.

Erdös, P. and A. Rényi (1959). On random graphs I. *Publ. Math. Debrecen 6*(18), 290–297.

Fan, J. and J. Lv (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70*(5), 849–911.

Fingleton, B. (2008). A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors. *Spatial Economic Analysis 3*(1), 27–44.

Fingleton, B. (2020). Exploring Brexit with dynamic spatial panel models: some possible outcomes for employment across the EU regions. *The Annals of Regional Science 64*(2), 455–491.

Fingleton, B. and N. Szumilo (2019). Simulating the impact of transport infrastructure investment on wages: a dynamic spatial panel model approach. *Regional Science and Urban Economics 75*, 148–164.

Gupta, A. and P. M. Robinson (2018). Pseudo maximum likelihood estimation of spatial autoregressive models with increasing dimension. *Journal of Econometrics 202*(1), 92–107.

Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association 97*(460), 1090–1098.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

REFERENCES

Hsieh, C. S. and L. F. Lee (2016). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics 31*(2), 301–319.

Hsu, N.-J., H.-L. Hung, and Y.-M. Chang (2008). Subset selection for vector autoregressive processes using lasso. *Computational Statistics & Data Analysis 52*(7), 3645–3657.

Kakamu, K., W. Polasek, and H. Wago (2008). Spatial interaction of crime incidents in Japan. *Mathematics and Computers in Simulation 78*(2), 276–282.

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

Kelejian, H. H. and I. R. Prucha (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics 157*(1), 53–67.

Keller, W. and C. H. Shiue (2007). The origin of spatial interaction. *Journal of Econometrics 140*(1), 304–332.

Knoke, D. and S. Yang (2008). *Social Network Analysis (2nd edition)*. New York: Sage.

Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models*. New York: Springer.

Lee, L. F. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews 22*(4), 307–335.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

REFERENCES

Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica 72*(6), 1899–1925.

Lee, L. F. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics 137*(2), 489–514.

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

Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics 28*(4), 825–860.

Liu, X., L. F. Lee, and C. R. Bollinger (2010). An efficient GMM estimator of spatial autoregressive models. *Journal of Econometrics 159*(2), 303–319.

Ma, Y., S. Guo, and H. Wang (2021). Sparse spatio-temporal autoregressions by profiling and bagging. *Journal of Econometrics*, In press.

Ma, Y., W. Lan, F. Zhou, and H. Wang (2020). Approximate least squares estimation for spatial autoregressive models with covariates. *Computational Statistics & Data Analysis 143*, 106833.

Ma, Y., R. Pan, T. Zou, and H. Wang (2020). A naive least squares method for spatial autoregression with covariates. *Statistica Sinica 30*(2), 653–672.

Martin, R. J. (1993). Approximations to the determinant term in gaussian maximum likelihood estimation of some spatial models. *Communications in Statistics - Theory and*

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

Methods 22(1), 189–205.

Nanda, V., Z. J. Wang, and L. Zheng (2004). Family values and the star phenomenon: Strategies of mutual fund families. *The Review of Financial Studies 17*(3), 667–698.

Newman, M. (2010). *Networks: An Introduction*. Oxford: Oxford University Press.

Nowicki, K. and T. A. B. Snijders (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American statistical association 96*(455), 1077–1087.

Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research 32*(3), 289–320.

Pareek, A. (2012). Information networks: Implications for mutual fund trading behavior and stock returns. In *AFA 2010 Atlanta Meetings Paper*.

Sawicki, J. and F. Finn (2002). Smart money and small funds. *Journal of Business Finance & Accounting 29*(5-6), 825–846.

Smirnov, O. and L. Anselin (2001). Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach. *Computational Statistics & Data Analysis 35*(3), 301 – 319.

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

Trefethen, L. N. and D. Bau (1997). *Numerical Linear Algebra*, Volume 50. Philadelphia: SIAM.

Wang, D., Y. Zheng, H. Lian, and G. Li (2021). High-dimensional vector autoregressive time

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

series modeling via tensor decomposition. *Journal of the American Statistical Association*, In press.

Wang, H., B. Li, and C. Leng (2009). Shrinkage tuning parameter selection with a diverging number of parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71*(3), 671–683.

Wang, H., R. Li, and C.-L. Tsai (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika 94*(3), 553–568.

Wilms, I., S. Basu, J. Bien, and D. S. Matteson (2021). Sparse identification and estimation of large-scale vector autoregressive moving averages. *Journal of the American Statistical Association*, In press.

Zhang, T. (2009). Adaptive forward-backward greedy algorithm for sparse learning with linear models. *Advances in neural information processing systems*, 1921–1928.

Zhong, P.-S., W. Lan, P. X. Song, and C.-L. Tsai (2017). Tests for covariance structures with high-dimensional repeated measurements. *The Annals of Statistics 45*(3), 1185–1213.

Zhou, J., Y. Tu, Y. Chen, and H. Wang (2017). Estimating spatial autocorrelation with sampled network data. *Journal of Business & Economic Statistics 35*(1), 130–138.

Zou, T., R. Luo, W. Lan, and C.-L. Tsai (2021). Network influence analysis. *Statistica Sinica*, In press.

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

Table 1: Detailed simulation results for Example 1 when the true parameters are $(\rho, \boldsymbol{\beta}^\top, \boldsymbol{\alpha}^\top) = (0.2, 0.1, 0.3, 0.7, 1.5, 2, 3, 6)$ and z_i $(i = 1, \cdots, n)$ follows a normal distribution. BIAS: the average bias; SD: the standard deviation computed from 500 replications; SE: the average of the estimated standard error. We also report the computational time (Time) in seconds.

Method	n	Measure	ρ	β_0	β_1	β_2	β_3	β_4	α_1	α_2	Time
	500	BIAS	0.000	-0.001	-0.002	0.002	0.010	-0.004	-0.001	0.000	29.782
		SD	0.005	0.012	0.080	0.205	0.459	0.671	0.053	0.017	
		SE	0.005	0.011	0.075	0.201	0.426	0.590	0.050	0.016	
QMLE	1000	BIAS	0.000	0.000	0.002	-0.001	-0.024	-0.029	0.002	0.000	247.419
		SD	0.004	0.008	0.053	0.149	0.310	0.440	0.039	0.011	
		SE	0.004	0.008	0.054	0.147	0.323	0.466	0.038	0.011	
	1500	BIAS	0.000	0.000	-0.002	-0.009	0.002	0.020	-0.003	0.000	1010.097
		SD	0.003	0.007	0.049	0.118	0.313	0.465	0.033	0.009	
		SE	0.003	0.006	0.044	0.122	0.277	0.409	0.032	0.009	
	500	BIAS	0.000	-0.001	-0.001	0.003	0.010	-0.006	-0.002	0.000	3.619
		SD	0.005	0.012	0.082	0.203	0.462	0.667	0.052	0.016	
		SE	0.005	0.011	0.075	0.201	0.426	0.590	0.050	0.016	
AMLE	1000	BIAS	0.000	0.000	0.002	-0.001	-0.024	-0.029	0.002	0.000	16.269
$(m = 2)$		SD	0.004	0.008	0.053	0.149	0.310	0.440	0.039	0.011	
		SE	0.004	0.008	0.054	0.147	0.323	0.466	0.038	0.011	
	1500	BIAS	0.000	0.000	-0.002	-0.009	0.002	0.020	-0.003	0.000	39.087
		SD	0.003	0.007	0.049	0.119	0.311	0.465	0.033	0.009	
		SE	0.003	0.006	0.044	0.122	0.277	0.409	0.032	0.009	
	500	BIAS	0.000	-0.001	-0.003	0.001	0.014	0.001	-0.002	0.000	3.537
		SD	0.005	0.012	0.080	0.204	0.451	0.655	0.052	0.016	
		SE	0.005	0.011	0.075	0.201	0.427	0.591	0.050	0.016	
AMLE	1000	BIAS	0.000	0.000	0.002	-0.001	-0.024	-0.029	0.002	0.000	16.437
$(m = 3)$		SD	0.004	0.008	0.053	0.149	0.310	0.440	0.039	0.011	
		SE	0.004	0.008	0.054	0.147	0.323	0.466	0.038	0.011	
	1500	BIAS	0.000	0.000	-0.002	-0.008	-0.001	0.014	-0.003	0.000	48.840
		SD	0.003	0.007	0.051	0.119	0.325	0.483	0.033	0.009	
		SE	0.003	0.006	0.044	0.122	0.276	0.408	0.032	0.009	
	500	BIAS	0.000	-0.001	-0.004	0.000	0.021	0.011	-0.002	0.000	9.775
		SD	0.005	0.011	0.078	0.203	0.435	0.631	0.051	0.016	
		SE	0.005	0.011	0.075	0.201	0.428	0.592	0.050	0.016	
AMLE	1000	BIAS	0.000	0.000	0.002	-0.001	-0.024	-0.029	0.002	0.000	43.578
(m_{BIC})		SD	0.004	0.008	0.053	0.149	0.310	0.440	0.039	0.011	
		SE	0.004	0.008	0.054	0.147	0.323	0.466	0.038	0.011	
	1500	BIAS	0.000	0.000	-0.004	-0.011	0.009	0.031	-0.003	0.000	129.494
		SD	0.003	0.007	0.044	0.115	0.285	0.423	0.033	0.009	
		SE	0.003	0.006	0.044	0.122	0.278	0.410	0.032	0.009	

Statistica Sinica: Preprint
doi:10.5705/ss.202021.0377

Table 2: Detailed model selection results of the EBIC for Example 3, with $d_0 = 4$, $\gamma = 0.5$, and z_i $(i = 1, \cdots, n)$ following a normal distribution. $|\hat{\mathcal{S}}|$: the average size of the selected model; CF: the average percentage of correctly fitted models; TPR: the average true positive rate; FPR: the average false positive rate.

n	$d = 6$				$d = 9$				$d = 12$			
	$\lvert\hat{\mathcal{S}}\rvert$	CF	TPR	FPR	$\lvert\hat{\mathcal{S}}\rvert$	CF	TPR	FPR	$\lvert\hat{\mathcal{S}}\rvert$	CF	TPR	FPR
500	3.410	0.410	0.770	0.165	3.230	0.240	0.685	0.098	3.240	0.210	0.678	0.066
1000	3.940	0.810	0.930	0.110	4.000	0.670	0.908	0.074	4.000	0.600	0.885	0.058
1500	3.990	0.950	0.985	0.025	4.090	0.780	0.940	0.066	4.190	0.640	0.935	0.056

Table 3: Detailed estimation results for the mutual fund data set.

Variables	SARg			pure SAR			SARAR(1,1)			SARMA(1,1)		
	Coef	Std	p-value	Coef	Std	p-value	Coef	Std	p-value	Coef	Std	p-value
ρ	0.192	0.020	0.000	0.045	0.240	0.859	0.337	0.351	0.337	0.050	0.088	0.566
λ	-	-	-	-	-	-	-0.555	0.539	0.303	-0.261	0.277	0.346
I_n	0.019	0.004	0.000	0.020	0.004	0.000	-	-	-	-	-	-
W	-0.018	0.004	0.000	-	-	-	-	-	-	-	-	-
Intercept	-0.072	0.002	0.000	-0.081	0.019	0.000	-0.060	0.027	0.027	-0.081	0.007	0.000
log(Size)	0.005	0.009	0.615	0.001	0.010	0.944	0.002	0.010	0.830	0.001	0.010	0.885
log(Age)	0.029	0.009	0.002	0.032	0.010	0.001	0.033	0.010	0.001	0.033	0.010	0.001
Return	0.032	0.009	0.000	0.036	0.010	0.000	0.036	0.010	0.000	0.036	0.009	0.000
Alpha	0.013	0.008	0.099	0.017	0.009	0.068	0.016	0.009	0.063	0.016	0.009	0.068
Log-likelihood	141.318			136.904			137.598			137.307		