

Network autoregression model with grouped factor structures*

ZHANG Zhiyuan, ZHU Xuening[✉]

School of Data Science, Fudan University, Shanghai 200433, China

Abstract: Network autoregression and factor model are effective methods for modeling network time series data. In this study, we propose a network autoregression model with a factor structure that incorporates a latent group structure to address nodal heterogeneity within the network. An iterative algorithm is employed to minimize a least-squares objective function, allowing for simultaneous estimation of both the parameters and the group structure. To determine the unknown number of groups and factors, a PIC criterion is introduced. Additionally, statistical inference of the estimated parameters is presented. To assess the validity of the proposed estimation and inference procedures, we conduct extensive numerical studies. We also demonstrate the utility of our model using a stock dataset obtained from the Chinese A-Share stock market.

Key words: network autoregression; factor structure; heterogeneity; latent group structure; network time series

CLC number: F22 **Document code:** A **Article ID:** 2097 – 0137 (2023) 05 – 0024 – 14

1 Introduction

Modeling network time series data is a crucial undertaking in numerous fields, as the popularity of network data continues to grow. Network data modeling has a broad range of applications, including social network analysis(Lewis et al. , 2008), financial risk management(Matthews, 2013), and spatial-temporal data modeling, among others.

In order to model network time series data, Zhu et al. (2017) proposed a network autoregression model to capture the dynamic relationships between network nodes. While useful, this model is limited by its assumption of homogeneous regression coefficients across all nodes in the network. To allow for greater heterogeneity, a possible extension is to incorporate a group structure among the regression coefficients. This approach is related to recent literature on group panel data models. For example, Lin et al. (2012) studied linear panel data models with parameter heterogeneity and employed group-specific slope coefficients to capture the group effect. They used a threshold-based method for parameter estimation and group membership assignment. Liu et al. (2020) extended this approach to both linear and nonlinear panel data models with group structure and proposed an iterative estimation algorithm that ensures consistency, provided that the number of groups is properly specified. For network data, Zhu et al. (2022) proposed a group network autoregression (GNAR) model that characterizes the network interaction effect.

* **Received:** 2023 – 04 – 04

Accepted: 2023 – 04 – 11

Published online: 2023 – 08 – 30

Supported by National Natural Science Foundation of China(72222009, 71991472)

✉ **Corresponding author:** ZHU Xuening(xueningzhu@fudan.edu.cn)

ZHANG Zhiyuan(zhiyuanzhang20@fudan.edu.cn)

The GNAR model, while able to flexibly model time series data collected from a large number of network nodes, is still limited in its ability to fully capture the complex dynamic patterns exhibited by all network nodes. One potential solution to this issue is to incorporate the factor model structure into the GNAR model. The factor model has been extensively studied in the literature and has a wide range of applications. For example, Fama et al. (1993) proposed the well-known five-factor model, which includes bond-market and stock-market factors to explain returns on stocks and bonds. Fan et al. (2011) used approximate factor models for high-dimensional covariance estimation, assuming that the error terms adopt a sparse covariance matrix, and used an adaptive thresholding technique for estimation. Ando et al. (2017) proposed a grouped factor model and applied it to cluster high-dimensional time series of financial data to capture the co-movement of stocks. Other related works include Hou et al. (2015), Bai(2003), and Stock et al. (2011). Combining the GNAR model with the grouped factor model is a natural way to enhance the interpretability power of complex time series collected from network nodes.

In this work, we propose a network autoregression model with a factor structure, which introduces a latent group structure to account for nodal heterogeneity and leverages the factor structure to model complex dynamic patterns. To estimate the parameters and group structure simultaneously, we utilize an iterative algorithm. To identify the appropriate group and factor numbers, we introduce a PIC criterion for model selection. Furthermore, we provide a statistical inference procedure for the estimated parameters, and validate the estimation and inference procedures through extensive numerical studies. Finally, we demonstrate the practical usefulness of the proposed model through an empirical analysis using data collected from the Chinese A-Share market.

The rest of this paper is organized as follows. Section 2 introduces the model and notations, the algorithm used to estimate the model, the procedures used for model selection, and the statistical inference. Section 3 presents the results of simulation studies and an empirical data analysis. Finally, Section 4 offers concluding remarks.

2 Network autoregression with factor structure

2.1 Model and notations

Consider a network with N nodes indexed as $i = 1, \dots, N$. For the i th node, let $Y_{it} \in \mathbb{R}$ be the dynamic response collected from the i th node at the t th time point for $1 \leq t \leq T$. For instance, Y_{it} can be the weekly return from the i th stock at the t th week. To characterize the network relationship among the network nodes, we employ an adjacency matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{N \times N}$. Specifically, $a_{ij} = 1$ implies that the i th node follows the j th node, otherwise $a_{ij} = 0$. Following the convention we let $a_{ii} = 0$. Let $\mathbf{W} = (w_{ij}) = (n_i^{-1} a_{ij}) \in \mathbb{R}^{N \times N}$ be the row-normalized adjacency matrix, where $n_i = \sum_j a_{ij}$ is the out-degree of the i th node. In addition, associated with the i th node at time t , we can collect a set of covariates $\mathbf{x}_{it} \in \mathbb{R}^p$.

To characterize the nodes' heterogeneity patterns, we introduce a group structure among the nodes. Specifically, let $g_i \in [G]$ be the group membership of the i th node. To model the dynamic pattern of Y_{it} , we consider the following autoregression model by embedding the observed network structure information,

$$Y_{it} = \sum_j \beta_{g_i, g_j} w_{ij} Y_{j(t-1)} + \nu_{g_i} Y_{i(t-1)} + \mathbf{x}_{it}^T \boldsymbol{\gamma}_{g_i} + \varepsilon_{it}, \quad (1)$$

where ε_{it} is a noise term. Here β_{g_i, g_j} captures the network effect between node i and j , ν_{g_i} is the momentum effect of the node, and $\boldsymbol{\gamma}_{g_i}$ is the regression coefficient for the exogenous covariates. Although the model (1) characterizes the network dependence among the nodes, it does not account for the common driven factors for nodes within the same group. To address this issue, we following Ando et al. (2017) and introduce a grouped factor model structure for the noise term, which is given by:

$$\varepsilon_{it} = \boldsymbol{\lambda}_{g_i, i}^T \mathbf{f}_{g_i, t} + e_{it}. \quad (2)$$

Here $\mathbf{f}_{g_i, t} \in \mathbb{R}^s$ is a group specific factor with factor loading $\boldsymbol{\lambda}_{g_i, i} \in \mathbb{R}^s$. Lastly, e_{it} is an independent and identically

distributed noise term with mean zero and variance σ^2 .

Notations Through out this paper, we define $[n] = \{1, \dots, n\}$ for any integer $n \geq 1$. In addition, we denote \mathbf{I}_n as identity matrix of dimension $n \times n$ for any positive integer $n \geq 1$. Besides, for an arbitrary vector $\mathbf{v} = (v_1, \dots, v_n)^\top \in \mathbb{R}^n$, we denote its L_2 -norm as $\|\mathbf{v}\| = \left(\sum_{i=1}^n v_i^2\right)^{1/2}$.

2.2 Model estimation

We first discuss the estimation when the number of groups G and the number of group specific factors r_g are given. Let $\boldsymbol{\beta} = (\boldsymbol{\beta}_{g'g'} : g, g' \in [G]) \in \mathbb{R}^{G \times G}$, $\boldsymbol{\nu} = (\nu_g : g \in [G])^\top$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_G)^\top$, and we collect the parameters in $\Theta = \{\boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\gamma}\}$. In addition, let $\boldsymbol{\Lambda}_g = (\boldsymbol{\lambda}_{g,i} : g_i = g)^\top \in \mathbb{R}^{N_g \times r_g}$ and $\mathbf{F}_g = (\mathbf{f}_{g,1}, \dots, \mathbf{f}_{g,T})^\top \in \mathbb{R}^{T \times r_g}$, where N_g is the number of nodes in the g th group. Lastly, let $\mathcal{G} = (g_i : i \in [N])^\top$ collects the group memberships of all subjects. To estimate the parameters, we minimize the following objective function,

$$\mathcal{L}(\Theta, \boldsymbol{\Lambda}_1, \mathbf{F}_1, \dots, \boldsymbol{\Lambda}_G, \mathbf{F}_G, \mathcal{G}) = \sum_{i,t} \left(Y_{it} - \sum_j \boldsymbol{\beta}_{g_i g_j} w_{ij} Y_{j(t-1)} - \nu_{g_i} Y_{i(t-1)} - \mathbf{x}_{it}^\top \boldsymbol{\gamma}_{g_i} - \boldsymbol{\lambda}_{g_i, i}^\top \mathbf{f}_{g_i, t} \right)^2.$$

We optimize the above objective function using an iterative algorithm. First, given $\{\Theta, \mathcal{G}\}$, we define $\mathbf{Z}_g = (\mathbf{z}_{g,1}, \dots, \mathbf{z}_{g, N_g})^\top \in \mathbb{R}^{N_g \times T}$, where $\mathbf{z}_{g,i} = (\varepsilon_{it} : g_i = g, t \in [T])^\top$. Following Bai et al. (2002), we can estimate \mathbf{F}_g by using the eigenvalue decomposition of $\mathbf{Z}_g^\top \mathbf{Z}_g$ under the identification condition $\mathbf{F}_g^\top \mathbf{F}_g / T = \mathbf{I}_{r_g}$. Specifically, we should have $\hat{\mathbf{F}}_g = \sqrt{T} \tilde{\mathbf{F}}_g$, where $\tilde{\mathbf{F}}_g$ collects the eigenvectors of $\mathbf{Z}_g^\top \mathbf{Z}_g$ corresponding to the largest r_g eigenvalues. Given $\hat{\mathbf{F}}_g$, the estimate of factor loading can be obtained as $\hat{\boldsymbol{\Lambda}}_g = \mathbf{Z}_g \hat{\mathbf{F}}_g / T$.

Next, given $\{\boldsymbol{\Lambda}_1, \mathbf{F}_1, \dots, \boldsymbol{\Lambda}_G, \mathbf{F}_G, \mathcal{G}\}$, we can update Θ via OLS method. We define $\Theta_g = (\boldsymbol{\beta}_g, \boldsymbol{\nu}_g, \boldsymbol{\gamma}_g^\top)^\top \in \mathbb{R}^{G+p+1}$, where $\boldsymbol{\beta}_g$ is the g th row vector of $\boldsymbol{\beta}$. In addition, we let $\mathcal{X}_{it} = \left(\sum_j w_{ij} I(g_j = 1) Y_{j(t-1)}, \dots, \sum_j w_{ij} I(g_j = G) Y_{j(t-1)}, Y_{i(t-1)}, \mathbf{x}_{it}^\top \right)^\top \in \mathbb{R}^{G+p+1}$, $\tilde{Y}_{it} = Y_{it} - \boldsymbol{\lambda}_{g_i, i}^\top \mathbf{f}_{g_i, t}$. Then we have

$$\hat{\Theta}_g = \left(\sum_{i \in \mathcal{C}_g} \sum_t \mathcal{X}_{it} \mathcal{X}_{it}^\top \right)^{-1} \left(\sum_{i \in \mathcal{C}_g} \sum_t \mathcal{X}_{it} \tilde{Y}_{it} \right), \quad (2)$$

where $\mathcal{C}_g = \{i : g_i = g\}$.

Lastly, given $\{\Theta, \boldsymbol{\Lambda}_1, \mathbf{F}_1, \dots, \boldsymbol{\Lambda}_G, \mathbf{F}_G\}$, we update the group memberships \mathcal{G} by minimizing the sum of squared residuals among all G possible groups for each node i sequentially from 1 to N . Specifically, with the information of all group memberships except for node i denoted by $\mathcal{G}_{(-i)} = (g_j : j \neq i)$, we assign node i to the group that has the smallest sum of squared residuals, which takes the form

$$\hat{g}_i = \operatorname{argmin}_g \sum_t \left(Y_{it} - \sum_j \boldsymbol{\beta}_{g g_j} w_{ij} Y_{j(t-1)} - \nu_g Y_{i(t-1)} - \mathbf{x}_{it}^\top \boldsymbol{\gamma}_g - \boldsymbol{\lambda}_{g, i}^\top \mathbf{f}_{g, t} \right)^2 + \sum_{j: g_j = 1, t} \left(Y_{jt} - \sum_{k: k \neq i} \boldsymbol{\beta}_{g_j g_k} w_{jk} Y_{k(t-1)} - \boldsymbol{\beta}_{g_j g} w_{ji} Y_{i(t-1)} - \nu_{g_j} Y_{j(t-1)} - \mathbf{x}_{jt}^\top \boldsymbol{\gamma}_{g_j} - \boldsymbol{\lambda}_{g_j, j}^\top \mathbf{f}_{g_j, t} \right)^2, \quad (3)$$

where $\boldsymbol{\lambda}_{g, i}$ is given by $\boldsymbol{\lambda}_{g, i} = \mathbf{F}_g^\top \mathbf{z}_i / T$ for $g = 1, \dots, G$.

We summarize the algorithm as illustrated above when the number of groups G and the number of group specific factors r_g , ($g = 1, \dots, G$) are given in Algorithm 1. With respect to the implementation of the algorithm, we have the following remarks.

Remark 1 To obtain a reliable estimation result, it is necessary to set appropriate initial values of group memberships and parameters in Algorithm 1, as the algorithm is not strictly convex. Specifically, we suggest the

Algorithm 1 Iterative estimation algorithm

-
- (i) Fix the number of groups G and the number of group specific factors r_g for $g = 1, \dots, G$. Initialize the group memberships $\mathcal{G}^{(0)} = (g_i^{(0)} : i \in [N])^T$ and the parameters $\Theta_g^{(0)} = (\boldsymbol{\beta}_g^{(0)}, \boldsymbol{\nu}_g^{(0)}, \boldsymbol{\gamma}_g^{(0)T})^T, g = 1, \dots, G$.
- (ii) For $m = 1, 2, \dots$, repeat the following iterations till converge
- (a) Given $\mathcal{G}^{(m-1)} = (g_i^{(m-1)} : i \in [N])^T$ and $\Theta_g^{(m-1)} = (\boldsymbol{\beta}_g^{(m-1)}, \boldsymbol{\nu}_g^{(m-1)}, \boldsymbol{\gamma}_g^{(m-1)T})^T, g = 1, \dots, G$, update the estimation of group specific factors and factor loadings $\{\Lambda_1^{(m)}, \mathbf{F}_1^{(m)}, \dots, \Lambda_G^{(m)}, \mathbf{F}_G^{(m)}\}$ via the principle components' estimates.
- (b) Given $\mathcal{G}^{(m-1)} = (g_i^{(m-1)} : i \in [N])^T$ and $\{\Lambda_1^{(m)}, \mathbf{F}_1^{(m)}, \dots, \Lambda_G^{(m)}, \mathbf{F}_G^{(m)}\}$, update the parameters $\Theta_g^{(m)} = (\boldsymbol{\beta}_g^{(m)}, \boldsymbol{\nu}_g^{(m)}, \boldsymbol{\gamma}_g^{(m)T})^T, g = 1, \dots, G$ by (2).
- (c) Given $\{\Lambda_1^{(m)}, \mathbf{F}_1^{(m)}, \dots, \Lambda_G^{(m)}, \mathbf{F}_G^{(m)}\}$ and $\Theta_g^{(m)} = (\boldsymbol{\beta}_g^{(m)}, \boldsymbol{\nu}_g^{(m)}, \boldsymbol{\gamma}_g^{(m)T})^T, g = 1, \dots, G$, update $\mathcal{G}^{(m)} = (g_i^{(m)} : i \in [N])^T$ by (3).
-

following approach. First, for each node i , we ignore the factor structure and the network structure, and only consider the $Y_{i(t-1)}$ and \mathbf{x}_{it} in the right side of (1). Using the least squares estimation method, we can obtain $\tilde{\boldsymbol{\alpha}}_i = \left(\sum_t \tilde{\mathbf{x}}_{it} \tilde{\mathbf{x}}_{it}^T \right)^{-1} \sum_t \tilde{\mathbf{x}}_{it} Y_{it}$, where $\tilde{\mathbf{x}}_{it} = (Y_{i(t-1)}, \mathbf{x}_{it}^T)^T$. Then we can apply the K -means algorithm on $\tilde{\boldsymbol{\alpha}}_i$ to cluster the N nodes into G groups. This gives the initial value of the group memberships $\mathcal{G}^{(0)} = (g_i^{(0)} : i \in [N])^T$. Next, given the initial values of the group memberships, we can obtain a rough estimate of the parameters $\Theta_g^{(0)} = (\boldsymbol{\beta}_g^{(0)}, \boldsymbol{\nu}_g^{(0)}, \boldsymbol{\gamma}_g^{(0)T})^T, (g = 1, \dots, G)$ by ignoring the factor structures. This serves as the initial values for the Algorithm 1. In simulation studies, this initial values contribute to fast convergence and superb estimation quality.

Remark 2 The principle components' estimate of the factors requires the eigen-decomposition of the $T \times T$ matrix $\mathbf{Z}_g^T \mathbf{Z}_g$, which can be computationally expensive when the time span T is large. To address this issue, Bai et al. (2002) proposes an alternative approach for estimating the factor structure. Specifically, it firstly estimates the factor loadings Λ_g , subject to the normalization of $\Lambda_g^T \Lambda_g / N = \mathbf{I}_{r_g}$, which is $\sqrt{N_g}$ times the first r_g eigenvectors of $\mathbf{Z}_g \mathbf{Z}_g^T \in \mathbb{R}^{N_g \times N_g}$. Suppose the estimate is given by $\hat{\Lambda}_g$, then we can obtain the estimate for \mathbf{F}_g as $\hat{\mathbf{F}}_g = \mathbf{Z}_g^T \hat{\Lambda}_g / N_g$. This alternative approach can speed up the estimation of the factor structure when $N_g < T$.

Remark 3 As pointed out by Zhu et al. (2022), the node i 's group membership g_i is involved in the objective function (2) not only through node i itself, but also through all those nodes that follows node i . As a result, the group membership updating equation (3) takes into consideration of two parts. The first part is the loss function of node i itself; the second part is the total loss function of nodes in $\{j : a_{ij} = 1\}$. Due to the sparse structure of real world networks, the updating procedure can be efficiently conducted.

2.3 Group and factor number estimation

In practice, the number of groups G and the number of group-specific factors r_g are usually unknown and need to be determined before conducting the estimation algorithm. To conduct the estimation algorithm, we first need to estimate the group number as well as the factor number. Motivated by Ando et al. (2017), we adopt a penalized information criterion (PIC) For estimation. The PIC is formulated as follows,

$$\begin{aligned} \text{PIC}(G, r_1, \dots, r_G) &= \frac{1}{NT} \sum_{i,t} \left(Y_{it} - \sum_j \beta_{g_i, g_j} w_{ij} Y_{j(t-1)} - \nu_{g_i} Y_{i(t-1)} - \mathbf{x}_{it}^T \boldsymbol{\gamma}_{g_i} - \boldsymbol{\lambda}_{g_i, i}^T \mathbf{f}_{g_i, t} \right)^2 \\ &+ \frac{1}{G} \sum_g C \times r_g \times \hat{\sigma}^2 \left(\frac{T + N_g}{TN_g} \right) \log(TN_g), \end{aligned} \tag{4}$$

where C is some positive constant and $\hat{\sigma}^2$ is an estimate of $E(e_{it}^2)$. The first part of PIC measures the goodness of model fitting while the second term is the penalty term applied on the number of groups and the number of group specific factors. By minimizing PIC we can select the appropriate values of G and r_g from a properly predefined range. In practice, the $\hat{\sigma}^2$ can be estimated by $\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^N \hat{e}_{it}^2$ when the possible largest G and r_g are specified, where \hat{e}_{it} is the corresponding residual value.

As shown by (4), the finite sample performance of the PIC relies largely on the specification of the tuning parameter C . To specify an appropriate value of C , we adopt the subsample based approach recommended by Ando et al. (2017), Hallin et al. (2007) and Alessi et al. (2010). Specifically, we define a set of time spans $T = T^{(0)} > T^{(1)} > \dots > T^{(S)} > 1$ with a pre-specified number S . Typically, $T^{(s)}$ is chosen as a grid value, e. g., $T^{(s)} = T - ms$ where m is a pre-specified integer number. Specifically, let $r_g^C(N, T^{(s)})$ denote the estimated factor number of the g th group by minimizing the PIC value with tuning parameter C using the subsample data $\mathcal{D}_s = \{(Y_{it}, \mathbf{x}_{it}) : i \in [N], t \in [T^{(s)}]\}$. Following Ando et al. (2017), we utilize the following empirical variance measure to select C as

$$V_C^2 = \sum_{g=1}^{G_{\max}} \left[\frac{1}{S} \sum_{s=1}^S \left(r_g^C(N, T^{(s)}) - S^{-1} \sum_{s'=1}^S r_g^C(N, T^{(s')}) \right)^2 \right], \tag{5}$$

where G_{\max} denotes the largest G value under consideration. Specifically, we set $r_g^C(N, T^{(s)}) = 0$ for $g > G^C(N, T^{(s)})$, where $G^C(N, T^{(s)})$ is the estimated number of groups with subsample \mathcal{D}_s and tuning parameter C . We summarize the model fitting procedure in the following.

Model Fitting Procedure

Step 1 Fix the tuning parameter C . For each candidate pairs of $G \in [G_{\max}]$ and $r_g \in [r_{\max}]$, estimate the model using algorithm 1, compute the PIC values. Select the (G, r_1, \dots, r_G) that minimizes PIC.

Step 2 Repeat Step 1 using the subsamples \mathcal{D}_s and calculate the corresponding V_C^2 values as in (5).

Step 3 Select the tuning parameter C that corresponds with stable V_C^2 values.

2.4 Statistical inference

In this section, we provide a statistical inference procedure for the estimated parameters Θ . As implied by the recent results of Ando et al. (2017) and Zhu et al. (2022), the group membership can be consistently estimated when $G = G_0$, where G_0 is the true group number. Therefore, in this subsection, we discuss the statistical inference of $\hat{\Theta}$ when the true group membership is given, i. e., $\mathcal{G} = \mathcal{G}^0$, where \mathcal{G}^0 is the true membership. For the g th group, we define $\mathbf{F}_g = (\mathbf{f}_{gt} : t \in [T])^T \in \mathbb{R}^{N_g \times r_g}$ and introduce the projection matrix

$$\mathbf{M}_g = \mathbf{I}_T - \mathbf{F}_g (\mathbf{F}_g^T \mathbf{F}_g)^{-1} \mathbf{F}_g^T = \mathbf{I}_T - \mathbf{F}_g \mathbf{F}_g^T / T.$$

Then the least squares estimator for Θ_g given $\mathbf{F} = \{\mathbf{F}_1, \dots, \mathbf{F}_G\}$ can be expressed as

$$\hat{\Theta}_g(\mathbf{F}) = \left(\sum_{i \in \mathcal{C}_g} \mathbb{X}_i^T \mathbf{M}_g \mathbb{X}_i \right)^{-1} \left(\sum_{i \in \mathcal{C}_g} \mathbb{X}_i^T \mathbf{M}_g \mathbb{Y}_i \right),$$

where $\mathbb{X}_i = (\mathcal{X}_{it} : t \in [T])^T \in \mathbb{R}^{T \times (G+p+1)}$ is previously defined. It readily follows Bai(2009) to obtain the asymptotic normality result for $\hat{\Theta}_g - \Theta_g^0$, where Θ_g^0 is the true value for Θ_g . Suppose the error term e_{it} is i. i. d over t and i with zero mean and variance σ^2 . We state the results in our case as follows. Let

$$\hat{\mathbf{D}}(\mathbf{F}_g) = \frac{1}{N_g T} \sum_{i: g_i = g} \left(\mathbb{X}_i^T \mathbf{M}_g \mathbb{X}_i - \frac{1}{N_g} \sum_{j: g_j = g} \mathbb{X}_i^T \mathbf{M}_g \mathbb{X}_j c_{g,j} \right),$$

where $c_{g,j} = \boldsymbol{\lambda}_{g,i}^T (\boldsymbol{\Lambda}_g^T \boldsymbol{\Lambda}_g)^{-1} \boldsymbol{\lambda}_{g,j}$. By Corollary 5.7 of Bai(2009), under certain regularity conditions, as $N, T \rightarrow \infty$ with $T/N \rightarrow 0$, we have

$$\sqrt{N_g T} \hat{\mathbf{D}}(\mathbf{F}_g)^{-1/2} (\hat{\boldsymbol{\Theta}}_g - \boldsymbol{\Theta}_g^0) \xrightarrow{d} N(0, \sigma^2 \mathbf{I}). \quad (6)$$

In practice, we implement the inference procedure by substituting $\hat{\mathbf{D}}(\mathbf{F}_g)$ as $\hat{\mathbf{D}}(\hat{\mathbf{F}}_g)$ and σ^2 as $\hat{\sigma}^2 = (NT)^{-1} \sum_{i,t} \hat{e}_{it}^2$ to implement the inference procedure. We examine the performance of the statistical inference procedure in the following section.

3 Numerical studies

3.1 Simulation

3.1.1 Simulation models and algorithm implementation To demonstrate the finite sample performance of the proposed method, we conduct simulation studies with various network structures and parameter settings using model (1). Specifically, the covariates \mathbf{x}_{it} , the group-specific factor $f_{g,t}$, and the factor loading $\boldsymbol{\lambda}_{g,i}$ are all independently and identically sampled from the standard normal distribution $\mathcal{N}(0,1)$. We investigate the following two network structures.

1) Stochastic Block Model(SBM). In the SBM network structure, the nodes are partitioned into K communities. Nodes from the same communities have higher probability to be connected, compared to nodes from different communities. Specifically, we set $P(a_{ij} = 1) = 2 \log(N)/N$ if node i and node j are from the same communities, and $P(a_{ij} = 1) = \log(N)/N$ otherwise. For different network sizes $N = 100, 200, 300$, we set the number of communities as $K = 5, 10, 15$, respectively.

2) Power-law Distribution Network(POW). In the POW network, the in-degrees $\left(d_i = \sum_{j=1}^N a_{ji} \right)$ of nodes follow a power-law distribution. Specifically, for each node i , we generate $P(d_i = d) = cd^{-\alpha}$, where c is a normalizing constant and α is the exponent parameter. We set $\alpha = 2.5$ as suggested by Clauset et al. (2009). This network well captures the characteristics that the majority of nodes have few followers while a small percentage of nodes have a large number of followers.

For each network structure, we consider two group settings with true number of groups $G_0 = 2$ and $G_0 = 3$, and sample the true memberships of nodes from a multinomial distribution with $(\pi_1, \pi_2) = (0.5, 0.5)$ for $G_0 = 2$ and $(\pi_1, \pi_2, \pi_3) = (0.3, 0.3, 0.4)$ for $G_0 = 3$ respectively. For convenience we set the number of factors $r_g^0 = 2$ for $g \in [G_0]$ and $p = 2$ for the dimension of \mathbf{x}_{it} across all group settings, and the true parameters are specified in Table 1 for $G_0 = 2$ and $G_0 = 3$ respectively. Here r_g^0 refers to the true factor number in the g th group. To compare the model performance under different types of noise terms, we follow Zou et al. (2017) to consider three different scenarios where the noise terms e_{it} are sampled from different distributions. In scenario 1, the noise terms are i. i. d sampled from standard normal distribution. In scenario 2, the noise terms are i. i. d sampled from a mixture normal distribution $\xi \cdot \mathcal{N}(0, 5/9) + (1 - \xi) \cdot \mathcal{N}(0, 5)$, where $P(\xi = 1) = 0.9$ and $P(\xi = 0) = 0.1$. In scenario 3, the noise terms are i. i. d sampled from a standardized exponential distribution $\text{Exp}(1) - 1$. We set different network sizes as $N = 100, 200, 300$ and time span $T = 100, 200, 300$ and run each model setting for $B = 200$ times.

3.1.2 Performance measurements and simulation results We utilize the proposed PIC criterion to select the number of groups G and the number of group specific factors r_g , which depends on the value of C . We set possible candidates for C as $C = 0.1k, k = 0, 1, \dots, 20$. To identify an appropriate value of C using V_C^2 , we set the sub-

Table 1 True parameters for $G_0 = 2$ and $G_0 = 3$

Groups	$G_0 = 2$						$G_0 = 3$					
	$\beta_{g, g'}^0$		ν_g^0		γ_g^0		$\beta_{g, g'}^0$		ν_g^0		γ_g^0	
g/g'	1	2	3	-	-	-	1	2	3	-	-	-
1	0.3	0.1	-	0.1	0.3	0.15	0.2	0.2	-0.1	-0.1	0.3	0.1
2	0.1	-0.1	-	-0.1	0.15	-0.1	0.1	0.3	-0.2	0.05	0.25	-0.1
3	-	-	-	-	-	-	0.15	0.1	-0.4	0.15	0.15	-0.3

sample size $T^{(s)} = T - s \times 10$ with $s = 0, 1, \dots, 4$. For each model setting, we first repeat the calculation of V_C^2 for 20 times to observe its behavior. Figure 1 plots V_C^2 values against C under the POW network with size $N = 200, T = 200, G_0 = 2$ and $r_g^0 = 2$ for $g = 1, 2$. The plot reveals that both extremely small and large values of C result in fluctuations of the V_C^2 values. However, a range of C values between 0.4 and 1.2 yields stable V_C^2 values in the 20 simulations. To facilitate our analysis, we set $C = 1$ for all model settings in the 200 simulation runs. This choice leads to satisfactory numerical results that are sufficient for illustrating the model's performance. The plot reveals that both extremely small and large values of C result in fluctuations of the V_C^2 values. However, a range of C values between 0.4 and 1.2 yields stable V_C^2 values in the 20 simulations. To facilitate our analysis, we set $C = 1$ for all model settings in the 200 simulation runs. This choice leads to satisfactory numerical results that are sufficient for illustrating the model's performance.

We first evaluate the performance of the proposed model selection method. In the b th simulation round, the estimated group number and factor numbers are denoted as $\hat{G}^{(b)}$ and $\hat{r}_g^{(b)}$ respectively. To measure the accuracy of estimated group numbers, we calculate $R_U^G = B^{-1} \sum_b I(\hat{G}^{(b)} < G_0)$, $R_C^G = B^{-1} \sum_b I(\hat{G}^{(b)} = G_0)$, and $R_O^G = B^{-1} \sum_b I(\hat{G}^{(b)} > G_0)$ as the percentage of under-estimated, correctly estimated, and over-estimated group numbers. Next, we evaluate the accuracy of the factor number estimation. For convenience, we only evaluate the cases when the group number is correctly estimated. Similar to the previous definitions, we define $R_U^f =$

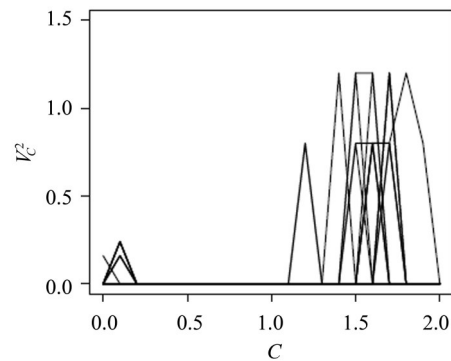


Fig. 1 Behavior pattern of values of V_C^2 on different values of C

$B_0^{-1} \sum_b \sum_{g=1}^{G_0} I(\hat{G}^{(b)} = G_0, \hat{r}_g^{(b)} < r_g^0)$, $R_C^f = B_0^{-1} \sum_b \sum_{g=1}^{G_0} I(\hat{G}^{(b)} = G_0, \hat{r}_g^{(b)} = r_g^0)$, $R_O^f = B_0^{-1} \sum_b \sum_{g=1}^{G_0} I(\hat{G}^{(b)} = G_0, \hat{r}_g^{(b)} > r_g^0)$, where $B_0 = \sum_b I(\hat{G}^{(b)} = G_0)$. The percentage (%) of under, correct and over-estimated number of groups \hat{G} and the number of factors \hat{r}_g for the SBM network and POW network are summarized in Table 2-4 for scenario 1 to scenario 3, respectively. First we look at the Table 2 for scenario 1. We can see that when the group structure is relatively simple ($G_0 = 2$), the proposed model selection method can select the correct model nearly all the time, except for a few cases of over-estimated number of groups \hat{G} when the network size N and time span T are both small ($N = 100, T = 100$). When $G_0 = 3$, the group structure is more complex and the proposed model selection method have both under- and over-estimation cases regarding \hat{G} and \hat{r}_g . However, one can clearly observe that as the network size N and time span T increases, the selection accuracy significantly increases. This indicates that our proposed model selection method can consistently select the true model.

The model selection results in scenario 2 and 3 are quiet similar to the results in scenario 1. Specifically, the scenario 2 and 3 have relatively higher selection accuracy in the number of groups \hat{G} , but slightly lower accuracy

Table 2 The percentage (%) of under, correct and over-estimated number of groups \hat{G} and the number of factors \hat{f}_g in scenario 1

G_0	N	T	POW						SBM					
			R_U^c	R_C^c	R_O^c	R_U^f	R_C^f	R_O^f	R_U^c	R_C^c	R_O^c	R_U^f	R_C^f	R_O^f
2	100	100	0	99	1	0	100	0	0	99.5	0.5	0	100	0
		200	0	100	0	0	100	0	0	100	0	0	100	0
		300	0	100	0	0	100	0	0	100	0	0	100	0
	200	100	0	100	0	0	100	0	0	100	0	0	100	0
		200	0	100	0	0	100	0	0	100	0	0	100	0
		300	0	100	0	0	100	0	0	100	0	0	100	0
	300	100	0	100	0	0	100	0	0	100	0	0	100	0
		200	0	100	0	0	100	0	0	100	0	0	100	0
		300	0	100	0	0	100	0	0	100	0	0	100	0
3	100	100	9	70.5	20.5	5.4	91.5	3.1	6.5	71	22.5	5.6	91.5	2.8
		200	6.5	87.5	6	4	94.5	1.5	7.5	87.5	5	3.2	96.2	0.6
		300	1.5	97	1.5	2.9	96.4	0.7	4.5	93.5	2	2.1	97.9	0
	200	100	0	85.5	14.5	0.8	98.4	0.8	0	85	15	0.2	98.8	1
		200	0	99.5	0.5	0	100	0	0	99.5	0.5	0	100	0
		300	0	100	0	0	100	0	0	100	0	0	100	0
	300	100	0	90.5	9.5	0	99.6	0.4	0	88	12	0.4	98.7	0.9
		200	0	100	0	0	100	0	0	99.5	0.5	0	100	0
		300	0	100	0	0	100	0	0	100	0	0	100	0

Table 3 The percentage (%) of under, correct and over-estimated number of groups \hat{G} and the number of factors \hat{f}_g in scenario 2

G_0	N	T	POW						SBM					
			R_U^c	R_C^c	R_O^c	R_U^f	R_C^f	R_O^f	R_U^c	R_C^c	R_O^c	R_U^f	R_C^f	R_O^f
2	100	100	0	98.5	1.5	0	100	0	0	99.5	0.5	0	100	0
		200	0	100	0	0	100	0	0	100	0	0	100	0
		300	0	100	0	0	100	0	0	100	0	0	100	0
	200	100	0	100	0	0	100	0	0	100	0	0	100	0
		200	0	100	0	0	100	0	0	100	0	0	100	0
		300	0	100	0	0	100	0	0	100	0	0	100	0
	300	100	0	100	0	0	100	0	0	100	0	0	100	0
		200	0	100	0	0	100	0	0	100	0	0	100	0
		300	0	100	0	0	100	0	0	100	0	0	100	0
3	100	100	5.5	79	15.5	7	89.9	3.2	5	77	18	5.4	91.1	3.5
		200	7.5	87.5	5	3.8	94.5	1.7	6.5	87.5	6	3	96.4	0.6
		300	2	96.5	1.5	3.8	95.3	0.9	4.5	93.5	2	4.1	95.2	0.7
	200	100	0	89	11	0.7	96.4	2.8	0	87.5	12.5	0.4	98.9	0.8
		200	0	100	0	0	100	0	0.5	98.5	1	0	100	0
		300	0	100	0	0	100	0	0	100	0	0	100	0
	300	100	0	89	11	0	99.3	0.7	0	91	9	0.2	99.6	0.2
		200	0	100	0	0	100	0	0	99	1	0	100	0
		300	0	100	0	0	100	0	0	100	0	0	100	0

Table 4 The percentage (%) of under, correct and over-estimated number of groups \hat{G} and the number of factors \hat{r}_g in scenario 3

G_0	N	T	POW						SBM					
			R_U^c	R_C^c	R_O^c	R_U^f	R_C^f	R_O^f	R_U^c	R_C^c	R_O^c	R_U^f	R_C^f	R_O^f
2	100	100	0	99	1	0	99.7	0.3	0	99.5	0.5	0	100	0
		200	0	100	0	0	100	0	0	100	0	0	100	0
		300	0	100	0	0	100	0	0	100	0	0	100	0
	200	100	0	100	0	0	100	0	0	100	0	0	100	0
		200	0	100	0	0	100	0	0	100	0	0	100	0
		300	0	100	0	0	100	0	0	100	0	0	100	0
	300	100	0	100	0	0	100	0	0	100	0	0	100	0
		200	0	100	0	0	100	0	0	100	0	0	100	0
		300	0	100	0	0	100	0	0	100	0	0	100	0
3	100	100	8	73.5	18.5	7.7	88.2	4.1	10.5	70	19.5	6.9	89.5	3.6
		200	6.5	87.5	6	3	95.6	1.3	8	90	2	2.8	96.3	0.9
		300	3	97	0	2.7	97.1	0.2	4.5	94	1.5	3.5	96.3	0.2
	200	100	0.5	87	12.5	0.8	98.3	1	0	92	8	0.5	98.6	0.9
		200	0	100	0	0	100	0	0	98.5	1.5	0	100	0
		300	0	100	0	0	100	0	0	100	0	0	100	0
	300	100	0	93.5	6.5	0	99.6	0.4	0	93	7	0.4	98.7	0.9
		200	0	100	0	0	100	0	0	100	0	0	100	0
		300	0	100	0	0	100	0	0	100	0	0	100	0

in the number of group specific factors \hat{r}_g . However, as N and T gets larger, the accuracy improvement trend in scenario 2 and 3 is just the same as in scenario 1. This indicates the robustness of our proposed model selection method against different types of noises.

Then, we evaluate the performance of the membership and parameter estimation when the number of groups are correctly selected. For the membership estimation, we report the average membership estimation accuracy as

$$AC_{\text{mem}} = B_0^{-1} \sum_{b=1}^{B_0} N^{-1} \sum_{i=1}^N I(\hat{g}_i^{(b)} = g_i^0),$$

where $\hat{g}_i^{(b)}$ is the estimated group membership i in the b th simulation round and g_i^0 is the true group membership.

For parameter estimation evaluation, we report the root mean squared error (RMSE) for $\hat{\beta}$, $\hat{\nu}$ and $\hat{\gamma}$.

Take $\hat{\beta}$ as an example, the RMSE is defined as $RMSE_{\beta} = B_0^{-1} \sum_{b=1}^{B_0} \left\| \hat{\beta}^{(b)} - \beta_0 \right\|_F$ after proper

label permutation, where $\hat{\beta}^{(b)}$ is the estimation for β in the b th simulation round and β_0 is the true parameter.

The RMSE for $\hat{\nu}$ and $\hat{\gamma}$ are similarly defined. Lastly, to evaluate the performance of the statistical inference using (6),

we construct a 95% confidence interval for each estimated parameter. Taking ν^0 for example, in the b th simulation run, the 95% interval for ν_g^0 is calculated as $CI_{\nu_g}^{(b)} = \left(\hat{\nu}_g^{(b)} - 1.96 \widehat{SE}_{\nu_g}^{(b)}, \hat{\nu}_g^{(b)} + 1.96 \widehat{SE}_{\nu_g}^{(b)} \right)$, where $\hat{\nu}_g^{(b)}$ and $\widehat{SE}_{\nu_g}^{(b)}$

are the estimated ν_g in the b th simulation and the estimated standard error of $\hat{\nu}_g^{(b)}$ based on (6) respectively. The coverage probability is then calculated as $CP_{\nu} = G_0^{-1} \sum_{g=1}^{G_0} B^{-1} \sum_{b=1}^B I(\nu_g^0 \in CI_{\nu_g}^{(b)})$. The CP's for β and γ are similarly defined.

The RMSE ($\times 1000$) for the estimation of β_0 , ν_0 and γ_0 , as well as their CP (%), in the parenthesis), and the AC_{mem} (%) for the SBM network and the POW network are summarized in Table 5–7 for scenario 1 to scenario

Table 5 The results for RMSE, CP and the AC_{mem} (%) in scenario 1

G_0	N	T	POW				SBM			
			$\hat{\beta}_0$	$\hat{\nu}_0$	$\hat{\gamma}_0$	AC_{mem}	$\hat{\beta}_0$	$\hat{\nu}_0$	$\hat{\gamma}_0$	AC_{mem}
2	100	100	29.71 (92.3)	14.9 (92.2)	14.57 (94.6)	99.5	29.24 (93.1)	14.64 (93.2)	13.95 (95.5)	99.6
		200	20.26 (93.6)	10.64 (93.0)	10.74 (94.1)	100	19.54 (94.1)	10.55 (92.2)	10.13 (94.9)	100
		300	16.57 (94.6)	8.35 (93.5)	8.47 (94.8)	100	16.45 (94.1)	8.29 (93.0)	8.40 (94.8)	100
	200	100	22.74 (93.6)	10.09 (93.5)	10.03 (94.5)	99.6	19.35 (92.2)	9.97 (93.2)	10.44 (93.8)	99.7
		200	15.48 (95.0)	6.63 (94.5)	7.11 (94.9)	100	12.96 (94.5)	6.61 (95.5)	7.40 (94.8)	100
		300	12.67 (95.6)	5.59 (94.2)	6.34 (93.0)	100	10.58 (94.2)	5.76 (93.8)	5.79 (95.1)	100
	300	100	19.14 (93.2)	8.31 (94.0)	8.40 (94.6)	99.5	15.24 (95.0)	8.38 (93.2)	8.56 (93.8)	99.6
		200	13.17 (94.0)	5.57 (95.0)	5.93 (95.2)	100	11.47 (93.9)	5.61 (95.2)	5.69 (94.8)	100
		300	10.96 (94.1)	4.47 (95.0)	4.83 (94.0)	100	8.78 (93.8)	4.48 (94.8)	4.74 (95.0)	100
3	100	100	71.51 (93.6)	21.85 (90.8)	27.03 (94.0)	95.4	78.73 (92.7)	25.59 (93.7)	28.72 (92.7)	95.3
		200	42.49 (94.6)	14.43 (94.0)	16.46 (93.8)	98.5	35.04 (93.4)	13.00 (94.3)	13.53 (94.0)	99.6
		300	29.40 (93.8)	11.51 (92.5)	11.07 (94.0)	99.6	27.26 (94.8)	11.57 (93.7)	10.83 (94.2)	99.9
	200	100	35.64 (94.1)	13.23 (95.4)	13.68 (93.2)	98.7	29.62 (94.3)	12.77 (94.6)	13.55 (94.0)	98.9
		200	22.58 (95.4)	8.91 (92.6)	9.09 (93.6)	99.8	20.35 (94.9)	8.49 (94.6)	8.88 (95.9)	99.9
		300	19.27 (94.5)	6.96 (94.2)	7.26 (95.2)	100	16.67 (94.2)	6.92 (95.0)	7.70 (92.8)	100
	300	100	30.43 (95.0)	10.07 (94.4)	10.85 (93.8)	98.9	25.91 (93.3)	11.08 (93.4)	11.07 (94.0)	98.8
		200	21.23 (94.3)	7.00 (94.7)	7.30 (94.3)	99.9	17.06 (94.3)	7.05 (94.3)	7.15 (95.2)	99.9
		300	17.44 (95.4)	5.80 (93.3)	6.10 (94.4)	100	13.96 (94.7)	5.68 (94.0)	5.89 (94.8)	100

3, respectively. First we look at the Table 5 for scenario 1. For the membership estimation, we can see that the proposed model estimation algorithm can achieve very high accuracy, even when the network size N and time span T are small. As for the parameter estimation, we can see that the RMSE of all estimated parameters decreases stably when the network size N and time span T increases. Lastly, regarding the statistical inference, we can see that the CP values in the parenthesis are all very close to the true value 95%, despite that when the $G = 3$ the difference between CPs and 95% are larger than when $G = 2$ on the whole.

The model estimation results in scenario 2 and 3 are also very similar to scenario 1, except that the membership estimation accuracy are slightly worse (less than 1%). Based on the performances we can conclude that our proposed model estimation algorithm can consistently estimate the model parameters and the memberships, and that the validity of statistical inference procedures in 2.6 are well supported.

3.2 An empirical case study with stocks from Chinese A-Share Market

In this section, we utilize the proposed model to analyze the returns of stocks from the Chinese A-Share market. We begin by presenting a brief description of the data. Next, we employ the PIC criterion to determine the appropriate number of groups and factors for the data. Finally, we conduct a model analysis using the proposed method.

3.2.1 Data description In this study, we collected $N = 766$ stocks from the Chinese A-Share market after performing necessary data cleaning procedures. Specifically, we select those stocks that have complete return and covariates information during the year from 2016 to 2021 from the China Stock Market and Accounting Research (CSMAR) database, which leads to $T = 72$ months in total. The response variable Y_{it} is the standardized return of each stock i at each month t .

First, since stocks from the same industry tend to exhibit higher correlation(Chan et al. , 1999), we construct

Table 6 The results for RMSE, CP and the AC_{mem} (%) in scenario 2

G_0	N	T	POW				SBM				
			$\hat{\beta}_0$	$\hat{\nu}_0$	$\hat{\gamma}_0$	AC_{mem}	$\hat{\beta}_0$	$\hat{\nu}_0$	$\hat{\gamma}_0$	AC_{mem}	
2	100	100	27.72 (94.5)	14.63 (94.4)	15.04 (94.2)	99.5	28.46 (93.8)	14.32 (94.5)	14.66 (94.5)	99.6	
		200	21.07 (93.4)	10.23 (93.8)	10.33 (93.9)	100	18.94 (94.5)	10.21 (93.0)	10.52 (93.2)	100	
		300	15.93 (94.9)	7.63 (95.5)	8.35 (95.0)	100	15.88 (94.8)	7.68 (95.2)	8.26 (94.9)	100	
	200	100	21.27 (95.5)	10.38 (94.2)	9.96 (95.4)	99.6	18.04 (94.8)	10.11 (94.8)	10.17 (93.6)	99.6	
		200	15.82 (94.4)	6.90 (96.2)	7.28 (93.8)	100	12.61 (94.4)	7.04 (95.2)	7.44 (94.2)	99.9	
		300	12.92 (94.2)	5.42 (96.2)	5.92 (95.5)	100	10.21 (95.2)	5.26 (96.5)	5.84 (94.0)	100	
	300	100	18.49 (94.5)	8.37 (93.2)	8.45 (94.5)	99.5	16.11 (93.4)	8.41 (94.2)	8.32 (94.6)	99.6	
		200	12.83 (95.2)	6.03 (92.0)	6.03 (93.4)	100	11.04 (94.0)	6.00 (92.5)	5.90 (95.4)	100	
		300	10.72 (94.5)	4.62 (95.0)	4.69 (95.5)	100	8.68 (94.4)	4.74 (93.5)	4.84 (94.2)	100	
	3	100	100	88.26 (92.4)	29.77 (93.2)	36.61 (94.0)	94.7	77.00 (93.0)	24.99 (93.2)	28.24 (92.8)	95.8
			200	35.86 (93.4)	14.47 (91.1)	13.90 (95.9)	98.7	48.29 (92.9)	15.41 (92.8)	17.45 (95.0)	98.8
			300	27.35 (94.5)	11.27 (93.9)	11.08 (94.0)	99.4	29.92 (93.9)	11.11 (95.3)	11.38 (93.5)	99.5
200		100	41.62 (93.3)	14.12 (92.8)	15.53 (93.3)	98	29.36 (94.2)	12.92 (94.5)	13.65 (93.2)	98.9	
		200	23.19 (94.9)	8.51 (94.5)	9.03 (94.5)	99.9	20.61 (93.9)	8.51 (95.3)	8.76 (95.7)	99.9	
		300	18.80 (95.1)	6.71 (96.5)	7.08 (95.7)	100	17.07 (94.2)	6.67 (95.8)	7.55 (94.8)	100	
300		100	29.91 (94.4)	9.75 (94.3)	10.81 (92.6)	98.9	26.40 (93.1)	10.11 (94.7)	10.49 (95.3)	99	
		200	21.31 (94.8)	7.05 (93.7)	7.20 (95.2)	99.9	16.83 (95.4)	7.20 (93.6)	7.25 (94.7)	99.8	
		300	17.83 (94.2)	5.47 (96.3)	5.81 (95.0)	100	13.68 (94.9)	5.37 (95.5)	5.83 (94.4)	100	

Table 7 The results for RMSE, CP and the AC_{mem} (%) in scenario 3

G_0	N	T	POW				SBM				
			$\hat{\beta}_0$	$\hat{\nu}_0$	$\hat{\gamma}_0$	AC_{mem}	$\hat{\beta}_0$	$\hat{\nu}_0$	$\hat{\gamma}_0$	AC_{mem}	
2	100	100	26.87 (95.9)	14.72 (91.9)	15.26 (93.7)	99.6	28.23 (93.7)	14.68 (92.7)	14.69 (94.1)	99.5	
		200	19.89 (93.4)	10.27 (93.8)	10.49 (93.9)	100	18.93 (94.2)	9.96 (94.8)	9.81 (95.4)	100	
		300	16.33 (94.1)	7.86 (95.0)	8.45 (96.0)	100	15.86 (94.9)	7.74 (95.0)	8.02 (95.9)	100	
	200	100	21.22 (94.9)	10.52 (92.0)	9.98 (94.8)	99.6	18.59 (93.9)	10.54 (92.2)	10.04 (95.0)	99.6	
		200	15.84 (94.9)	7.07 (94.0)	7.34 (93.9)	100	12.72 (94.6)	6.67 (95.2)	7.49 (93.1)	100	
		300	13.40 (93.6)	5.51 (95.8)	5.94 (93.6)	100	10.55 (95.5)	5.78 (94.8)	5.63 (94.1)	100	
	300	100	18.81 (94.5)	8.63 (92.8)	8.18 (95.0)	99.5	16.44 (92.9)	8.43 (93.2)	8.14 (95.4)	99.7	
		200	13.17 (94.0)	5.42 (95.2)	6.08 (93.4)	100	10.82 (96.0)	5.50 (95.0)	6.14 (93.1)	100	
		300	10.93 (93.0)	4.58 (93.2)	4.75 (94.9)	100	8.79 (94.4)	4.55 (93.8)	4.73 (95.0)	100	
	3	100	100	81.60 (91.6)	26.54 (90.4)	32.01 (94.2)	93.7	89.44 (93.7)	26.90 (90.6)	32.87 (93.1)	94.7
			200	33.77 (94.2)	13.56 (94.8)	13.56 (95.1)	98.9	42.16 (93.9)	14.55 (93.6)	15.67 (93.9)	99.2
			300	27.12 (93.7)	10.55 (94.0)	11.21 (93.4)	99.9	27.17 (94.8)	10.93 (92.3)	10.87 (94.6)	99.8
200		100	35.03 (94.1)	12.84 (93.3)	13.79 (93.8)	98.5	29.83 (93.9)	13.42 (92.9)	13.47 (94.9)	98.7	
		200	22.99 (94.8)	8.90 (93.5)	9.00 (95.5)	99.9	21.19 (93.2)	8.82 (94.1)	9.28 (93.6)	99.9	
		300	20.35 (92.8)	7.03 (94.2)	7.53 (94.5)	100	17.20 (94.1)	7.01 (94.3)	7.35 (94.2)	100	
300		100	31.75 (92.8)	11.01 (91.8)	10.53 (93.8)	98.9	24.69 (95.4)	10.99 (91.6)	11.18 (93.9)	98.8	
		200	21.57 (94.3)	6.78 (96.0)	7.46 (94.3)	99.9	16.88 (95.3)	6.81 (94.2)	7.28 (94.8)	99.9	
		300	17.09 (95.3)	5.73 (94.0)	5.76 (96.0)	100	13.73 (95.2)	5.65 (95.0)	6.10 (93.7)	100	

the stock network A based on their industry classification. Specifically, $a_{ij} = 1$ if stock i and stock j belong to the same industry, and $a_{ij} = 0$ if otherwise. The diagonal elements of A are set to be 0. In addition, we also collected $p = 4$ covariates from the financial statement of the firms, which are SIZE(the logarithm of the firm's market value), BM(book to market ratio), CASH(cash flow of the firm) and LEV(the firm's leverage ratio). Those covariates are standardized with mean 0 and standard deviation 1. We included these covariates because they have significant potential explanatory power on stock returns (Chan et al. , 1998), and they are widely used in existing literature(Zou et al. , 2017; Fan et al. , 2022).

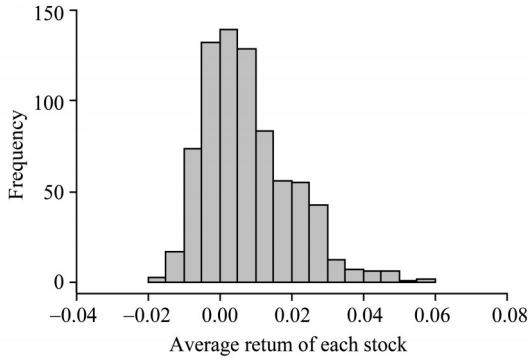


Fig. 2 Histogram of average return of $p = 766$ stocks

We first conduct some basic data descriptive analysis as follows. We visualize the histogram of the average return of each stock $\frac{1}{T} \sum_t Y_{it}, i = 1, \dots, N$ in Figure 2. We can see that the majority of the stocks gained an average return between -1% and 3% , with only a small proportion of stocks having an average return over 3% or less than -1% . Next, we plot the time series of average stock returns $\frac{1}{N} \sum_i Y_{it}, t = 1, \dots, T$ in Figure 3. We

observe two peaks occurring in the third month in the year of 2016 and in the second month in the year of 2019. Regarding the industry distribution of the $N = 766$ stocks, they are sourced from 47 unique industries. The pharmaceutical industry has the highest representation with 83 stocks in the industry.

3.2.2 Model estimation and evaluation Next we fit our model on this stock data. To this end, we consider subsample sizes $T^{(s)} = T - s$, where $s = 0, 1, \dots, 5$, and compute V_C^2 for $C = 0.1k$, where $k = 0, 1, \dots, 20$. The resulting values of V_C^2 are plotted in Figure 4. We observe that V_C^2 is relatively small in the range $C \in [1.1, 2.0]$. Therefore, we choose $C = 1.2$ for the subsequent analysis.

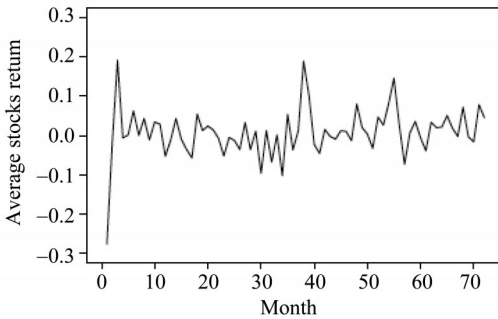


Fig. 3 The time series of average stock returns over $T = 72$ months

We then use the PIC to select the group number and factor numbers. This results in $\hat{G} = 6$ and $(\hat{r}_1, \dots, \hat{r}_6) = (1, 1, 1, 1, 1, 1)$.

Table 8 presents the estimation results. We first examine the group membership assignment, where we observe that each of the six groups has a size around 100, with Group 1 having the largest size of 178 stocks and Group 5 having the smallest size of 70 stocks. The stock allocation across the six groups is relatively balanced, with each group having a specific factor. These findings suggest a strong grouped factor structure in the stock market, where the stock returns are influenced by their respective group-specific factors. Moving on to the group-wise network effects, we observe that most groups have negative inner-group network effects, indicating a negative correlation among the stocks within each group. Regarding the momentum effects, we find that most groups have negative coefficients, except for Group 5, which reflects the well-known mean-reversion phenomenon in the stock mar-

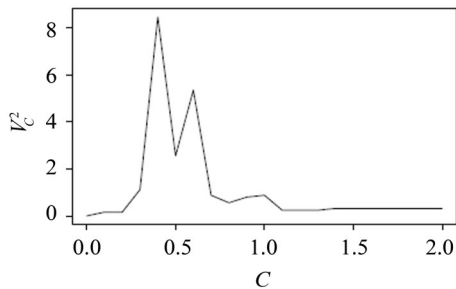


Fig. 4 Behavior of values of V_C^2 on different values of C , calculated from subsamples of real data

ket. We observe that most groups have negative inner-group network effects, indicating a negative correlation among the stocks within each group. Regarding the momentum effects, we find that most groups have negative coefficients, except for Group 5, which reflects the well-known mean-reversion phenomenon in the stock mar-

ket. Finally, we examine the covariate effects. SIZE has a positive coefficient on all six groups, while BM has a negative coefficient on all six groups. CASH has a positive coefficient on Group 1 and 5, and a negative coefficient on the other groups. LEV has a small negative coefficient on Group 1 and 2, a negative coefficient on Group 3 and 4, and a positive coefficient on Group 5 and 6. In summary, the Chinese A-share market exhibits a strong group structure in stock returns, and their behavior patterns vary significantly depending on the group to which they belong.

Table 8 The estimated parameters as well as their standard errors (in the parenthesis), the group size, and the number of group specific factors for each group

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Group 1	-0.018 (0.029)	-0.094 (0.06)	-0.109 (0.081)	0.264* (0.062)	-0.742* (0.143)	1.399* (0.103)
Group 2	0.585* (0.057)	-0.046 (0.031)	0.143 (0.096)	-0.045 (0.049)	0.599* (0.16)	-0.064 (0.076)
Group 3	-0.318* (0.081)	-0.593* (0.1)	-0.256* (0.145)	0.131 (0.091)	-0.746* (0.182)	-0.451* (0.157)
Group 4	0.227* (0.058)	0.031 (0.048)	0.327* (0.09)	0.028 (0.051)	0.148 (0.117)	-0.618* (0.097)
Group 5	-0.022 (0.112)	0.885* (0.139)	0.512* (0.132)	0.036 (0.069)	-0.41* (0.112)	0.454* (0.109)
Group 6	-0.899* (0.084)	0.219* (0.077)	-0.7* (0.131)	-0.195* (0.077)	-0.192 (0.119)	-0.346* (0.078)
Momentum	-0.097* (0.009)	-0.043* (0.009)	-0.072* (0.013)	-0.1* (0.009)	0.065* (0.015)	-0.082* (0.01)
SIZE	0.008* (0.001)	0.005* (0.001)	0.011* (0.002)	0.009* (0.001)	0.017* (0.002)	0.014* (0.002)
BM	-0.011* (0.001)	-0.017* (0.001)	-0.014* (0.001)	-0.01* (0.001)	-0.026* (0.002)	-0.02* (0.002)
CASH	0.001 (0.001)	-0.001 (0.001)	-0.011* (0.006)	-0.005* (0.001)	0.02* (0.011)	-0.005 (0.004)
LEV	0.000 (<0.001)	0.000 (0.001)	-0.002* (0.001)	-0.004* (0.002)	0.003 (0.002)	0.011* (0.003)
N_g	178	167	81	164	70	106
r_g	1	1	1	1	1	1

* means 5% significant.

4 Conclusion

In this work, we study the network autoregression model with factor structure. It introduces a latent group structure so that it can characterize the nodes' heterogeneity patterns. We propose an iterative algorithm to estimate the parameters as well as the group structure simultaneously. A PIC criterion is developed to select the number of groups and the number of group specific factors. We also perform statistical inference on the estimated parameters and conduct a number of simulation studies to examine the finite-sample performances. Lastly, we apply our model to a real stock dataset to demonstrate its usefulness.

Here we briefly discuss some future research topics. First, we assume in our model that the same group structure applies to both the parameters and the factors. It is interesting to allow for different group structures for regression coefficients and latent factors. Second, rigorous theoretical investigation can be provided to the statistical inference. Lastly, besides the group specific factors, the inclusion of common factors may improve the interpretation power of the model.

References:

- ALESSI L, BARIGOZZI M, CAPASSO M, 2010. Improved penalization for determining the number of factors in approximate factor models[J]. *Stat Probab Lett*, 80(23/24): 1806–1813.
- ANDO T, BAI J, 2017. Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures[J]. *J Amer Stat Assoc*, 112(519): 1182–1198.

- BAI J, 2003. Inferential theory for factor models of large dimensions[J]. *Econometrica*, 71(1): 135–171.
- BAI J, 2009. Panel data models with interactive fixed effects[J]. *Econometrica*, 77(4): 1229–1279.
- BAI J, NG S, 2002. Determining the number of factors in approximate factor models[J]. *Econometrica*, 70(1): 191–221.
- CHAN L K C, KARCESKI J, LAKONISHOK J, 1998. The risk and return from factors[J]. *J Financial Quant Anal*, 33(2): 159–188.
- CHAN L K C, KARCESKI J, LAKONISHOK J, 1999. On portfolio optimization: Forecasting covariances and choosing the risk model[J]. *Rev Financ Stud*, 12(5): 937–974.
- CLAUSET A, SHALIZI C R, NEWMAN M E J, 2009. Power-law distributions in empirical data[J]. *SIAM Rev*, 51(4): 661–703.
- FAMA E F, FRENCH K R, 1993. Common risk factors in the returns on stocks and bonds[J]. *J Financial Econ*, 33(1): 3–56.
- FAN J, LIAO Y, MINCHEVA M, 2011. High dimensional covariance matrix estimation in approximate factor models[J]. *Ann Stat*, 39(6): 3320–3356.
- FAN X, LAN W, ZOU T, et al, 2022. Covariance model with general linear structure and divergent parameters[J]. *J Bus Econ Stat*: 1–13.
- HALLIN M, LIŠKA R, 2007. Determining the number of factors in the general dynamic factor model[J]. *J Amer Stat Assoc*, 102(478): 603–617.
- HOU K, XUE C, ZHANG L, 2015. A comparison of new factor models[R/OL]. Fisher College of Business Working Paper, (2017–04–19)[2023–04–13]. <https://dx.doi.org/10.2139/ssrn.2520929>.
- LEWIS K, KAUFMAN J, GONZALEZ M, et al, 2008. Tastes, ties, and time: A new social network dataset using Facebook.com[J]. *Soc Netw*, 30(4): 330–342.
- LIN C C, NG S, 2012. Estimation of panel data models with parameter heterogeneity when group membership is unknown[J]. *J Econom Methods*, 1(1): 42–55.
- LIU R, SHANG Z, ZHANG Y, et al, 2020. Identification and estimation in panel models with overspecified number of groups[J]. *J Econom*, 215(2): 574–590.
- MATTHEWS K, 2013. Risk management and managerial efficiency in Chinese banks: A network DEA framework[J]. *Omega*, 41(2): 207–215.
- STOCK J H, WATSON M W, 2011. *Dynamic factor models*[D]. Cambridge: Harvard University.
- ZHU X, PAN R, LI G, et al, 2017. Network vector autoregression[J]. *Ann Stat*, 45(3): 1096–1123.
- ZHU X, XU G, FAN J, 2022. Simultaneous estimation and group identification for network vector autoregressive model with heterogeneous nodes[EB/OL]. arXiv:2209.12229, (2022–09–25)[2023–04–13]. <https://arxiv.org/abs/2209.12229>.
- ZOU T, LAN W, WANG H, et al, 2017. Covariance regression analysis[J]. *J Amer Stat Assoc*, 112(517): 266–281.

(责任编辑 冯兆永)