Statistica Sinica Preprint No: SS-2024-0170	
Title	Network Assisted Approximate Factor Model Estimation
Manuscript ID	SS-2024-0170
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0170
Complete List of Authors	Yuzhou Zhao, Xinyan Fan and Bo Zhang
Corresponding Authors	Xinyan Fan
E-mails	1031820039@qq.com

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

Network Assisted Approximate Factor Model Estimation

Yuzhou Zhao, Xinyan Fan* and Bo Zhang

Center for Applied Statistics and School of Statistics, Renmin University of China

Abstract: The factor models are powerful tools for uncovering patterns of similarity or co-movement among individuals, and they have been successfully applied in the fields of finance and biology. However, the classical approximate factor model encounters limitations when dealing with small sample sizes. To overcome this challenge, we leverage auxiliary network information and propose a novel joint quasi-maximum likelihood estimation, which can use the network information flexibly and allow network heterogeneity. The theoretical properties of these estimators are rigorously established. We obtain a new convergence rate, which is faster than the rate of classical maximum likelihood estimators when the sample size is small. Numerous numerical studies have been conducted to evaluate the performance of the proposed methods.

Key words and phrases: approximate factor model, high dimensionality, latent space model, network structure, penalized maximum likelihood

*Corresponding author.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

1. Introduction

The factor models are powerful tools for uncovering patterns of similarity or co-movement among individuals, which have been successfully applied in financial engineering, economic analysis, and biological technology (Fama and French, 1992; Chamberlain and Rothschild, 1983; Mayrink and Lucas, 2013). As one of the most commonly used factor models, the approximate factor model is appealing as it allows the idiosyncratic errors to be cross-sectionally correlated. There are two main strategies for the estimation of factor models: Principal Component (PC) based estimation, and maximum likelihood (ML) based estimation. PC-based method minimizes the sum of squares of response prediction errors (Bai, 2003; Fan et al., 2013). ML-based method maximizes a Gaussian-type log-likelihood function to obtain the factor loadings (Bai and Li, 2012; Bai and Liao, 2016). Compared to the PC-based method, the ML-based method is more efficient under cross-sectional heteroskedasticity structures with unknown dependence structures (Bai and Liao, 2016).

Despite the great success in theory, methodology, and applications, there are still limitations in the estimation of factor models. One major constraint is the necessity for a large sample size. Specifically, the estimation of factor loadings is constrained to a rate no faster than $O_p(T^{-1/2})$,

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

where T represents the sample size (Bai and Li, 2016). To alleviate the requirement for a large sample size, one popular approach is to incorporate auxiliary information (such as explanatory variables, spatial information, and network). For example, Fan et al. (2016) pointed out that the factor loadings are often highly correlated with explanatory variables, and projected (smoothed) data matrix onto a given linear space spanned by explanatory variables to estimate the factor loadings. For another example, Huang and Yang (2010) assumed that the factor loadings corresponding to the same cluster of explanatory variables are all the same.

In addition to explanatory variables, networks among individuals represent a distinct and crucial form of auxiliary information. Homophily and heterophily are common phenomena in networks. Homophily refers to greater similarity among connected individuals, while heterophily suggests greater dissimilarity (McPherson et al., 2001; Xie et al., 2016). Both phenomena highlight that networks provide additional information for describing the similarity between individuals. Consequently, properly incorporating network data can substantially improve the performance of factor models. To illustrate the role of networks in factor models, we present three examples. First, consider financial data. Factor models are powerful tools for studying the co-movement of stock returns. Networks between

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

stocks, such as co-holding relationships, also capture the co-movement of stocks (Anton and Polk, 2014; Lin and Qiu, 2024), providing valuable auxiliary information to enhance the factor model. Second, consider biological gene data. Gene networks, based on interactions or co-occurrence (Wong et al., 2004; Yi et al., 2022), are essential for identifying similarities between genes. Integrating these networks into the factor model allows for a more effective derivation of low-dimensional representations. Lastly, consider environmental pollution data. Factor models can be employed to capture the common variation of air pollutants across different regions. Networks constructed from spatial geographic locations or climate conditions can further assist in analyzing the similarities in pollution levels across regions (Fountalis et al., 2014; Von Ferber et al., 2009). To incorporate network information, Yu et al. (2020) excavated the prior network information through the Laplacian penalty and Projection penalty. However, their models are not trouble-free. For example, the Laplacian penalty neglects the degree heterogeneity. Thus, such a model may be not suitable for degree heterogeneous networks. Meanwhile, the Laplacian and Projection penalties are applied to the vector of factor loadings, disregarding variations in network effects across different loadings. Accordingly, there is a need to develop a new factor model with network association.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

In this paper, we propose a novel network assisted approximate factor model (NAAF). Inspired by Fan et al. (2016) and Linton and Connor (2000), we assume that the factor loadings are functions of some latent variables but not the observable explanatory variables. The latent variables can be the characteristics of individuals or some important but unobservable explanatory variables. For simplicity, the functions between the latent vector and the factor loadings are assumed to be linear. In addition to the impact on the factor loadings, the latent variables are also assumed to determine the network structure. Following Hoff et al. (2002); Krivitsky et al. (2009), we model the network using a latent space model, and assign individuals in the network latent "locations" that determine their connection pattern. We assume that the latent "locations" of individuals are also linear functions of the latent variables that determine the factor loadings. The spaces spanned by the columns of the factor loading matrix and the individuals' latent location matrix in the network are assumed to be the same. Under this assumption, we incorporate the network information by jointly estimating the factor loadings and the latent locations of individuals by the penalized quasi-maximum likelihood estimator. Specifically, we pursue parameters that maximize the balanced quasi-maximum likelihood function of the factor model and latent space model and make the covariance matrix

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

of idiosyncratic errors sparsity via penalization.

The main contributions can be summarized as follows. First, we develop a novel framework for jointly analyzing the factor model and network information, which can be flexibly extended to other factor-based methods. The proposed method enriches the factor models with the association of auxiliary information. Second, we rigorously establish the statistical properties of estimators. The model's identifiability has been demonstrated. The explicit convergence rates of estimated factor loadings and the covariance of response are studied. Under mild conditions, the factor loadings convergence rate can be improved from $O_p(T^{-1/2})$ to $O_p((T/\log(p))^{-3/4})$. The proof of theoretical properties faces great challenges, which stem from the joint estimation of the factor loadings and "latent" locations, and the complexity of the quasi-likelihood function. Third, an efficient alternative updating algorithm is developed to address the resultant optimization task. At last, numerical experiments on both simulated and real examples indicate that the proposed method is superior to the existing ML–based methods and PC–based methods.

The paper is organized as follows. Section 2 introduces the NAAF, proposes the penalized quasi-maximum likelihood function, and develops an efficient algorithm to tackle the computational challenge. Section 3 es-

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

tablishes the statistical properties of NAAF estimators. Simulation studies and an empirical example are given in Sections 4 and 5, respectively. Section 6 concludes the article with short discussions. All theoretical proofs are relegated to the supplementary material.

2. Methodology

2.1 Model and Parameter Estimation

Let us consider p individuals for T periods. Denote $Y = (Y_1, \ldots, Y_T) = (Y_{it})_{p\times T} \in \mathbb{R}^{p\times T}$ as the large panel data, where Y_{it} is the response value of i-th individual at time t, for $i = 1, \cdots, p$ and $t = 1, \ldots, T$. The volatilities of response variables are driven by a few latent common factors and idiosyncratic errors,

$$Y_t = \mu_0 + B_0 f_t + e_t, \text{ for } t = 1, \ldots, T, \tag{2.1}$$

where $B_0 = (b_{01}, \ldots, b_{0p})^\top \in \mathbb{R}^{p\times r}$ is the factor loadings matrix and $b_{0i} = (b_{0i,1}, \ldots, b_{0i,r})^\top$ is the corresponding loading vector of individual i, $f_t = (f_{1t}, \ldots, f_{rt})^\top$ is the unobservable factor vector, r is the factor number and assumed to be known in our model, $\mu_0 \in \mathbb{R}^p$ is the mean vector of Y_t, and $e_t \in \mathbb{R}^p$ is the idiosyncratic error vector with mean zero and covariance matrix Σ_{e0}, which allowed being cross-sectional correlated. To make equation

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

(2.1) identifiable, we assume that $B_0^\top \Sigma_{e0}^{-1} B_0$ is diagonal, $T^{-1}\sum_{t=1}^{T} f_t f_t^\top = I_r$ and $T^{-1}\sum_{t=1}^{T} f_t = 0$. Similar identifiable conditions can also be found in (Bai and Liao, 2016).

Suppose a network is collected alongside. The network can be represented by an adjacency matrix $A = (A_{ij})_{p\times p} \in \{0,1\}^{p\times p}$, where $A_{ij} = A_{ji} = 1$ if there exists an edge between two individuals (i,j), and $A_{ij} = A_{ji} = 0$ otherwise. We use the latent space model to analyze the network. We assume that the (i,j)-th elements of adjacency matrix A are independently generated as follows:

$$\Pr(A_{ij} = 1) = P_{ij} \text{ and } \operatorname{logit}(P_{ij}) = (\alpha_i^* + \alpha_j^* + \beta_i^\top I_{q_1,q_2}\beta_j), \tag{2.2}$$

for $i,j = 1,\ldots,p$, where $\operatorname{logit}(x) = \log\{x/(1-x)\}$ for any $x \in \mathbb{R}$, $\beta_i \in \mathbb{R}^r$ is the latent vector corresponding the i-th individual's latent location, $I_{q_1,q_2} = \operatorname{diag}(I_{q_1}, -I_{q_2})$ with $q_1 + q_2 = r$, and α_i^* is the heterogeneity parameter of ith individual, for $i = 1,2,\ldots,p$. Denote $\Gamma = (\beta_1,\ldots,\beta_p)^\top$. For ease of presentation, we write $A \sim Ber(P)$, where $P = (P_{ij})_{p\times p} \in \mathbb{R}^{p\times p}$. In Equation (2.2), the function form of β_i and β_j is similar to the model of Rubin-Delanchy et al. (2022), although they considered P_{ij} rather than $\operatorname{logit}(P_{ij})$. The connection probability of nodes i and j increases with the similarity of the first q_1 elements of β_i and β_j, and decreases with the similarity of their last q_2 elements. The model accounts for only homophily

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

in the network when $q_2 = 0$, and only heterophily when $q_1 = 0$. Equation (2.2) allows for the coexistence of both homophily and heterophily.

To establish the connection between the factor model and the network model, we assume that characteristics of individuals can be represented by a latent matrix $Z \in \mathbb{R}^{p\times r}$. That is, the i-th individual can be represented by the i-th row of Z, $Z_{i\cdot}$. We further model the factor loading matrix $B_0 = ZW_1$ and the individual location matrix $\Gamma = ZW_2$, for some transition matrices W_1 and W_2 with full rank. The linear function is used to represent the relationship between B_0 and Z and that Γ and Z for simplicity. This assumption is reasonable to some extent. For example, let Y_t be the activity measure of all individuals at time t on social media, and A represent the friendship network. The loading factors and latent locations in the network are all dependent on the individuals' hobbies and characteristics. Since W_1 and W_2 are invertible, there exists W, such that $\Gamma = B_0W$. We denote $\Omega_0 = WI_{q_1,q_2}W^\top$. Then Equation (2.2) can be rewritten as

$$\text{logit}(P) = B_0\Omega_0B_0^\top + \alpha^*\mathbf{1}_p^\top + \mathbf{1}_p\alpha^{*\top}, \tag{2.3}$$

where $\text{logit}(P) = (\text{logit}(P_{ij}))_{p\times p}$, and $\alpha^{*\top} = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_p^*)^\top$. Note that Equation (2.3) is not identifiable. For example, let $\breve{B} = B_0 + \mathbf{1}_p\iota^\top$, $\breve{\alpha} = \alpha^* - B_0\Omega_0\iota - (1/2)\mathbf{1}_p\iota^\top\Omega_0\iota$ for some constant vector $\iota \in \mathbb{R}^p$. Then, we have $\breve{B}\Omega_0\breve{B}^\top + \breve{\alpha}\mathbf{1}_p^\top + \mathbf{1}_p\breve{\alpha}^\top = B_0\Omega_0B_0^\top + \alpha^*\mathbf{1}_p^\top + \mathbf{1}_p\alpha^{*\top}$. To make (2.3)

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

identifiable, we revise (2.3) as

$$\text{logit}(P) = J_p B_0 \Omega_0 B_0^\top J_p + \alpha_0 \mathbf{1}_p^\top + \mathbf{1}_p \alpha_0^\top,$$

with $J_p = I_p - \mathbf{1}_p \mathbf{1}_p^\top / p$, following Zhang et al. (2022). Further elaboration on the identifiability can be found in Theorem 1 of Section 3.

Remark 1. In this paper, we assume $\Gamma = B_0 W$ for some square matrix W, which implies that the factor number and the dimension of the latent vector are the same. In a more general model, the dimensionality of the latent space and the number of factors may not be equal. Let r be the dimension of f_t and k the dimension of β_i. Our model can accommodate the case $k < r$. In that case, W is an $r \times k$ matrix, and $\Omega_0 = W I_{q_1,q_2} W^\top$ is singular. Our model can also be applied to the case where $k > r$ with a extension. Recall that the dimensions of Γ are $p \times k$. One can assume $\Gamma = (B_0 W, \Gamma_0)_{p\times k} \Phi_\pi$, where Γ_0 is $p \times (k - r)$ matirx, and Φ_π is a column permutation matrix. Here, Φ_π is designed to ensure that the node connection probabilities increase when the first q_1 latent variables are similar, while the remaining q_2 latent variables are dissimilar. The latent space model can be reformulated as $\text{logit}(P) = J_p B_0 \Omega_0 B_0^\top J_p + \Upsilon_0 I_{q_{01},q_{02}} \Upsilon_0^\top + \alpha_0 \mathbf{1}_p^\top + \mathbf{1}_p \alpha_0^\top$, where $\Omega_0 = W I_{q_{b1},q_{b2}} W^\top$, $\Upsilon_0 = J_p \Gamma_0$ is the the matrix of latent variables that cannot be captured by the factor loading matrix B_0, and $q_{b1}, q_{b2}, q_{01}, q_{02}$ are constants satisfying $q_{b1} + q_{01} = q_1$, $q_{b2} + q_{02} = q_2$.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

To estimate the factor loadings with the assistance of the network, we consider the penalized negative quasi-log likelihood estimator,

$$(\widehat{B}, \widehat{\Sigma}_e, \widehat{\Omega}, \widehat{\alpha}) = \text{argmin}_{B,\Omega,\alpha,\Sigma_e \succ 0} L(B, \Sigma_e, \Omega, \alpha),$$

$$L(B, \Sigma_e, \Omega, \alpha) = L_Y(B, \Sigma_e) + \lambda T^{-1} L_A(B, \Omega, \alpha) + P_T(\Sigma_e). \tag{2.4}$$

In equation (2.4), $L_Y(B, \Sigma_e)$ and $L_A(B, \Omega, \alpha)$ are proportional to the negative quasi-log likelihood functions corresponds to Y and A with parameters B, Σ_e, Ω, and α, respectively, that is

$$L_Y(B, \Sigma_e) = \log\{\det(BB^\top + \Sigma_e)\} + \text{tr}\{S_y(BB^\top + \Sigma_e)^{-1}\},$$

$$L_A(B, \Omega, \alpha) = -\sum_{1 \le i < j \le p} [A_{ij}\Theta_{A,ij} - \log\{1 + \exp(\Theta_{A,ij})\}],$$

where S_y is the sample covariance matrix of Y, $\Theta_A = J_p B \Omega B^\top J_p + \alpha \mathbf{1}_p^\top + \mathbf{1}_p \alpha^\top$, λ is a tuning parameter, and $P_T(\Sigma_e)$ is a penalty function on Σ_e. With $\widehat{B}$ and $\widehat{\Sigma}_e$, we can estimate $\widehat{f_t}$ via generalized least squares.

The objective function (2.4) comprises three terms. With the first two terms, we pursue the best fitting of the data matrix Y and the adjacency matrix A. The tuning parameter λ balances the importance of $L_Y(B, \Sigma_e)$ and $L_A(B, \Omega, \alpha)$, which is data-dependent. When $\lambda = 1$, the $L_Y(B, \Sigma_e) + \lambda T^{-1} L_A(B, \Omega, \alpha)$ is directly proportional to the joint negative log-likelihood function of (Y, A). However, in practice, $\lambda = 1$ is not always the best choice. Since the assumption that $\Gamma = B_0 W$ is strong, a data-driven tuning

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

parameter enhances the model's robustness when these assumptions are not precisely met. In numerical studies, the selection of λ depends on various factors, such as the scale of Ω_0. When Ω_0 is small, the network information is limited. Taking the network into consideration may introduce extra errors as well as information. Thus, in such cases, the choice, $\lambda < 1$, is better than that $\lambda = 1$. The third term $P_T(\Sigma_e)$ encourages the sparsity of $\widehat{\Sigma}_e$. In this paper, we consider the lasso penalty (Tibshirani, 1996), such that $P_T(\Sigma_e) = \rho_{p,T} \sum_{i \neq j} |\Sigma_{e,ij}|$ with a tuning parameter $\rho_{p,T}$. The same penalty has also been found in Bai and Liao (2016).

2.2 Computing Algorithm

We develop an efficient alternative updating algorithm based on the gradient descent method. Let $\Pi = (B, \Omega, \alpha)$. Let $B^{(k-1)}, \Sigma_e^{(k-1)}, \Omega^{(k-1)}, \alpha^{(k-1)}$ and $\Pi^{(k-1)}$ be the parameters obtained in the $(k-1)$-th iteration. The kth iteration consists of two steps. In first step, we update $\Sigma_e^{(k)} = \text{argmin}_{\Sigma_e} L(\Sigma_e, \Pi^{(k-1)})$, which is optimized using the method proposed by (Bien and Tibshirani, 2011). Recall that $\Sigma^{(k-1)} = B^{(k-1)}(B^{(k-1)})^\top + \Sigma_e^{(k-1)}$. We substitute the concave term $\log\{\det(\Sigma^{(k-1)})\}$ with the tangent plane $\text{tr}\{(\Sigma^{(k-1)})^{-1}(\Sigma_e - \Sigma_e^{(k-1)})\}$ and then minimize $\text{tr}\{(\Sigma^{(k-1)})^{-1}(\Sigma_e - \Sigma_e^{(k-1)})\} + \text{tr}\{S_y(B^{(k-1)}(B^{(k-1)})^\top + \Sigma_e)^{-1}\} + P_T(\Sigma_e)$. In the second step, we aim to optimize $\Pi^{(k)} = \text{argmin}_{\Pi} L(\Sigma_e^{(k)}, \Pi)$.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

We use the gradient descent method to update $\Pi^{(k)}$. Furthermore, we can set $\widehat{\Omega}$ to be a diagonal matrix, and remove the restriction $\widehat{B}^\top \widehat{\Sigma}_e^{-1} \widehat{B}$ is diagonal in the iteration. Finally, we take $\widehat{B}O$ as the estimator such that $O^\top \widehat{B}^\top \widehat{\Sigma}_e^{-1} \widehat{B}O$ is a diagonal matrix. The algorithm details are provided in Algorithm 1.

Algorithm 1 requires initial inputs of several hyperparameters. In selecting the number of factors, we initially disregard network information. The information criterion is used to determine r (Bai and Ng, 2002). Specifically, $\widehat{r} = \text{argmin}_k \log\{(pT)^{-1} \min_B ||Y - B\widehat{F}_k||_F^2\} + k(p+T)\log\{pT/(p+T)\}/(pT)$, where $\widehat{F}_k$ is the PCA estimator of factors when the number of factors is k. Step sizes $(s_B, s_\alpha, s_\Omega)$ and η are user-specified small constants. In this paper, we set the s and η change as p and T vary. The initial values of $B^{(0)}$, $\Sigma_e^{(0)}$, $\alpha^{(0)}$, and $\Omega^{(0)}$ are obtained as follows. First, we analyze Y to obtain $\tilde{B}^{(0)}$ and $\Sigma_e^{(0)}$ by the POET method, which is a PC-based method to estimate the approximate factor model proposed by Fan et al. (2013). Then, we use the project gradient descent algorithm to obtain an approximate estimate $\tilde{\Theta}_A$ and set $\Omega^* = (\tilde{B}^{(0)\top} J_p \tilde{B}^{(0)})^{-1} \tilde{B}^{(0)\top} J_p \tilde{\Theta}_A J_p \tilde{B}^{(0)} (\tilde{B}^{(0)\top} J_p \tilde{B}^{(0)})^{-1}$. Finally, we find an orthogonal matrix U such that $U\Omega^* U^\top$ is diagonal, and set $\Omega^{(0)} = U\Omega^* U^\top$ and $B^{(0)} = \tilde{B}^{(0)} U^\top$. Cross-validation is used to select the tuning parameters λ and $\rho_{p,T}$, following Bai and Liao (2016). For more

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

Algorithm 1

1: Input factor number r, step sizes $s = (s_B, s_\Omega, s_\alpha)$, small number η, hyperparameter ρ_{pT} and λ.

2: Set $k = 0$. Initialize $(B^{(0)}, \Sigma_e^{(0)}, \Omega^{(0)}, \alpha^{(0)})$.

3: Update $k = k + 1$.

(a) Update Σ_e: Let $\widehat{\Sigma}^{(k-1)} = B^{(k-1)}(B^{(k-1)})^\top + \Sigma_e^{(k-1)}$, and

$$\Phi = \Sigma_e^{(k-1)} - \eta \left\{ (\widehat{\Sigma}^{(k-1)})^{-1} - (\widehat{\Sigma}^{(k-1)})^{-1} S_y (\widehat{\Sigma}^{(k-1)})^{-1} \right\}.$$

Set $\Sigma_e^{(k)} = (\Sigma_{e,ij}^{(k)})$, where $\Sigma_{e,ij}^{(k)} = \mathcal{S}(\Phi_{ij}, \eta\rho_{p,T})\mathrm{I}(i \neq j) + \Phi_{ij}\mathrm{I}(i = j)$, $\mathcal{S}(a, \eta\rho_{p,T}) = \mathrm{sign}(a)(|a| - \eta\rho_{p,T})^+$, and $x^+ = \max(x, 0)$ for any $x \in \mathbb{R}$ is the positive part of x.

(b) Update Π:

(i) Initialize $n = 0$ and $\Pi^{(k,0)} = \Pi^{(k-1)}$.

(ii) Update $n = n + 1$. Let

$$B^{(k,n)} = B^{(k,n-1)} - s_B \frac{\partial L}{\partial B}(B^{(k,n-1)}, \widehat{\Sigma}_e^{(k)}, \Omega^{(k,n-1)}, \alpha^{(k,n-1)}),$$
$$\alpha^{(k,n)} = \alpha^{(k,n-1)} - s_\alpha \frac{\partial L}{\partial \alpha}(B^{(k,n-1)}, \widehat{\Sigma}_e^{(k)}, \Omega^{(k,n-1)}, \alpha^{(k,n-1)}),$$
$$\Omega^{(k,n)} = \Omega^{(k,n-1)} - s_\Omega \frac{\partial L}{\partial \Omega}(B^{(k,n-1)}, \widehat{\Sigma}_e^{(k)}, \Omega^{(k,n-1)}, \alpha^{(k,n-1)}).$$

(iii) Repeat (ii) until convergence. Set $\Pi^{(k)} = \Pi^{(k,n)}$.

4: Repeat Step 3 until convergence.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

details, please refer to the Section S.8 in the supplementary materials. In numerical studies, the algorithm is computationally affordable. For example, when $p = 200, T = 300$, the ten times average computational time is 37.53 seconds using a laptop with Apple M1 Pro and 16 GB memory.

2.3 Benefits of Network

We note that if λ in equation (2.4) is set to 0, our method reduces to the classical approximate factor model. The factor loadings estimated based on the penalized likelihood function of Y, $L_Y(B, \Sigma_e) + P_T(\Sigma_e)$, will also yield consistent estimates. However, the convergence rate of the factor loadings is inherently limited to $O_p(T^{-1/2})$. In this paper, we introduce $\lambda T^{-1} L_A(B, \Omega, \alpha)$ to leverage the information from the network. Next, we clarify the benefits of incorporating network information when T is small.

Define $\bar{\Gamma} = J_p \Gamma$. Then, $\bar{\Gamma} I_{q_1,q_2} \bar{\Gamma}^\top = J_p B_0 \Omega_0 B_0^\top J_p$. Recall that $\Gamma = B_0 W$. For simplicity, we assume that W is nonsingular in this subsection. Then, we have $B_0 = (\mathbf{1}_p, \bar{\Gamma}) \bar{W}$ for some matrix $\bar{W} \in \mathbb{R}^{(r+1)\times r}$. Consequently, if $\bar{\Gamma}$ is known, the estimation of B_0 is reduced to the estimation of the coefficient matrix $\bar{W}$. The number of parameters requiring estimation decreases significantly from $O(p)$ to $O(1)$. This reduction in the number of parameters leads to faster convergence rates for the factor model. A

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

similar result can be found in (Fan et al., 2016). In this paper, $\bar{\Gamma}$ cannot be observed directly. Fortunately, we can recover the basis vector accurately using the network model. For example, when $q_2 = 0$, the convergence rate of the estimator of $\bar{\Gamma}$ is $O_p(p^{-1/2})$ (Zhang et al., 2020). By introducing the loss function $\lambda T^{-1} L_A(B, \Omega, \alpha)$, when λ is large, the estimation of the basis for the column space of B_0 becomes more accurate in the case of small T and large p, thereby improving the estimation of B_0.

It is important to highlight that the method proposed in this paper differs notably from the following two-step approach. In the two-step method, one can first estimate $\bar{\Gamma}$ only using the network information, then estimate B_0 and Σ_{e0} by optimizing $L_Y(B, \Sigma_e) + P_T(\Sigma_e)$ under the constraint $J_p B = \widehat{\bar{\Gamma}} \widehat{W}$, where $\widehat{W}$ is a parameter matrix and $\widehat{\bar{\Gamma}}$ is the estimator of $\bar{\Gamma}$. Compared to the two-step approach, our joint likelihood method is more flexible due to the introduction of λ. The tuning parameter λ balances the information derived from the network and that from the panel data. In the small p and large T cases, our method allows for a smaller value of λT^{-1}, which facilitates greater utilization of the information contained in the panel data, whereas when p is large and T is small, a larger value of λT^{-1} is preferred to incorporate the network information better. Additionally, our method selects smaller values of λ when the network generation does not

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

align with the hypothesis, resulting in more robust outcomes. Numerical results also confirm that the proposed method achieves smaller estimation errors than the two-step approach. For more details, refer to Section 4.

3. Theoretical Properties

For any matrix $X = (x_{ij}) \in \mathbb{R}^{k_1 \times k_2}$, denote $\sigma_k(X)$, $\sigma_{\max}(X)$, and $\sigma_{\min}(X)$ as the kth singular value, the largest singular value, and the smallest singular value of X, respectively. Let $||X||_F = \sqrt{\text{tr}(X^\top X)}$, $||X||_2 = \sigma_{\max}(X)$ and $||X||_1 = \max_{j \le k_2} \sum_i |x_{ij}|$ be the Frobenius norm, spectral norm, and maximum absolute column-sum norm, respectively. For given estimator $\widehat{\Sigma}$ of some covariance matrix Σ, we use the norm $||\widehat{\Sigma} - \Sigma||_\Sigma = p^{-1/2}||\Sigma^{-1/2}(\widehat{\Sigma} - \Sigma)\Sigma^{-1/2}||_F$ to evaluate the accuracy of estimator following Fan et al. (2013). Denote the true parameters as $B_0, \Sigma_{e0}, \Omega_0, \alpha_0$, and Θ_{A0}. Recall that $\Theta_A = \Theta_A(B, \Omega, \alpha) = J_p B \Omega B^\top J_p + \alpha \mathbf{1}_p^\top + \mathbf{1}_p \alpha^\top$. The true covariance matrix of Y_t and its estimator are $\Sigma_{Y0} = B_0 B_0^\top + \Sigma_{e0}$ and $\widehat{\Sigma}_Y = \widehat{B}\widehat{B}^\top + \widehat{\Sigma}_e$, respectively. We denote $F = (f_1, \ldots, f_T) \in \mathbb{R}^{r \times T}$, and $\mathcal{E} = (e_1, \ldots, e_T) \in \mathbb{R}^{p \times T}$. Define $J_U, J_L \subseteq \{(i,j) : i \le p, j \le p\}$ such that $J_U \cap J_L = \varnothing$ and $J_U \cup J_L = \{(i,j) : i \le p, j \le p\}$. Let J_L be the set of the indices for small elements of Σ_e in absolute value, and J_U contain the indices for large elements. Denote $D_p := \#\{(i,j) : (i,j) \in J_U, i \ne j\}$ be

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

the cardinal number of J_U. For any $N \in \mathbb{R}^+$, denote $\mathcal{F}_{-\infty}^0$ and $\mathcal{F}_N^\infty$ are σ-algebras generated by $\{(f_t, e_t), -\infty < t \leq 0\}$ and $\{(f_t, e_t), N \leq t < +\infty\}$ respectively. Define $\varphi(N) = \sup_{D \in \mathcal{F}_{-\infty}^0, G \in \mathcal{F}_N^\infty} |P(D)P(G) - P(DG)|$, and $\rho(N) = \sup_{g_1 \in \mathcal{L}^2(\mathcal{F}_{-\infty}^0), g_2 \in \mathcal{L}^2(\mathcal{F}_N^\infty)} |corr(g_1, g_2)|$, where $\mathcal{L}^2(\mathcal{F}_{-\infty}^0)$ is the set of all $\mathcal{F}_{-\infty}^0$ measurable functions with finite second order moments, and $\mathcal{L}^2(\mathcal{F}_N^\infty)$ has a similar definition. To study the properties of NAAF, we introduce the following seven technical assumptions.

Assumption 1. (1) Assume that $\{f_t, e_t\}$s are strictly stationary. In addition, $E(e_{it}) = E(e_{it}f_{jt}) = 0$ for all $i \leq p, j \leq r$ and $t \leq T$.

(2) There exist constants $c, C, c_1, C_1, c_2 > 0$ such that $c \leq \sigma_{\min}(\Sigma_{e0}) \leq \sigma_{\max}(\Sigma_{e0}) \leq C$, $c_1 \leq \sigma_{\min}(p^{-1}B_0^\top B_0) \leq \sigma_{\max}(p^{-1}B_0^\top B_0) \leq C_1$, and $\max\{\max_{j \leq p}\{||b_{0j}||_2\}, ||\Sigma_{e0}||_1, ||\Sigma_{e0}^{-1}||_1\} \leq c_2$, where b_{0j} is the vector corresponding to the j-th row of B_0.

(3) There exist $r_1, r_2 > 0$ and $a_1, a_2 > 0$, such that for any $s > 0$, $i \leq p$ and $j \leq r$, $\Pr(|e_{it}| > s) \leq \exp(-(s/a_1)^{r_1})$ and $\Pr(|f_{jt}| > s) \leq \exp(-(s/a_2)^{r_2})$.

(4) There exist $r_3 > 0$ and $a_3 > 0$ satisfying: for all $N \in \mathbb{Z}^+$,$\varphi(N) \leq \exp(-a_3 N^{r_3})$, and $r_4^{-1} := 3r_1^{-1} + r_3^{-1} > 1, r_5^{-1} := 3r_2^{-1} + r_3^{-1} > 1$.

Assumption 2. (1) The adjacency matrix, A, is independent of (e_t, f_t)s.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

(2)There exist constants $m, M, c_3, C_v > 0$ such that for all large p,

$$m \leq \sigma_{\min}(p^{-1}B_0^\top J_p B_0) \leq \sigma_{\max}(p^{-1}B_0^\top J_p B_0) \leq M, \text{ and } \sigma_r(\Omega_0) \geq c_3 v_p,$$

where $v_p^{1+\varepsilon} \gg p^{-1/2}$ for some small positive ε. Further, we assume $c_v \leq \sigma_1(\Omega) \leq C_v$.

(3)There exist M_1 such that $\max_{i\leq p, j\leq p} |\Theta_{A0,ij}| \leq M_1$.

Assumption 3. Assume that

(1) $(i,i) \in J_U$ for all $i \leq p$, and $D_p \ll \min\{p\sqrt{T/\log(p)}, p^2/\log(p), p^2 v_p^{2(1+\varepsilon)}\}$,

(2) $K_T := \sum_{(i,j)\in J_L} |\Sigma_{e,ij}| = o(p)$.

Assumption 4. (1) Assume that $p^{-1}B_0^\top \Sigma_{e0}^{-1} B_0$ is diagonal, and there exists a positive definite matrix H_1 with r distinct eigenvalues, such that $p^{-1}B_0^\top \Sigma_{e0}^{-1} B_0 \to H_1$.

(1') Assume that $p^{-1}B_0^\top (\Sigma_e^*)^{-1} B_0$ is diagonal, with $(\Sigma_e^*)_{ij} = \Sigma_{e0,ij} 1_{\{i=j\}}$. There exists a positive definite matrix H_2 with r distinct eigenvalues, such that and $p^{-1}B_0^\top (\Sigma_e^*)^{-1} B_0 \to H_2$.

Assumption 5. The tuning parameter $\rho_{p,T}$ in the penalty function satisfies

$$\sqrt{\log(p)/T} + \log(p)/p + 1/(pv_p^{2(1+\varepsilon)}) \ll \rho_{p,T} \ll \min\{p/D_p, \sqrt{p/D_p}, p/K_T\}.$$

Assumption 6. (1) There exist positive constants a_3', r_3' such that $\rho(N) \leq \exp(-a_3' N^{r_3'})$.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

(2) Assume that $(pT)^{-1}\sum_{i=1}^{p}\sum_{t=1}^{T}\sum_{s=1}^{T} f_t f_s^{\top} cov(e_{it}, e_{is}) \geq m_1$ almost surely, for some constant m_1.

Assumption 7. There exists a sufficiently large constant K, such that

(1) $||f_t||_2 \leq K$ almost surely for $t = 1, \ldots, T$, and $\{f_t\}$ independent of $\{e_t\}$;

(2) $E\left(e_{it}e_{js}\right) = \gamma_{ij,ts}$, with $(pT)^{-1}\sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{t=1}^{T}\sum_{s=1}^{T}|\gamma_{ij,ts}| \leq K$;

(3) $E\left[\left\|(pT)^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T}(\Sigma_{e0,ii})^{-1}b_{0i}\left\{e_{it}e_{jt} - E\left(e_{it}e_{jt}\right)\right\}\right\|_2^2\right] \leq K$, for $j = 1, \ldots, p$; and

(4) $E\left[\left\|(pT)^{-1/2}\sum_{i=1}^{p}\sum_{t=1}^{T}(\Sigma_{e0,ii})^{-2}b_{0i}b_{0i}^{\top}\left(e_{it}^2 - \Sigma_{e0,ii}\right)\right\|_F^2\right] \leq K$.

Assumption 1 is a commonly used assumption in the approximate factor model, which can be found in (Fan et al., 2013; Bai and Liao, 2016). Assumption 1 (1) states a stationary relationship and assumes non-correlation between f_t and e_t. Assumption 1 (2) assumes the eigenvalues of Σ_{e0} and $p^{-1}B_0^{\top}B_0$ are bounded, which leads to r spiked eigenvalues of S_y. Assumptions 1 (3) and (4) allow us to apply the Bernstein-type inequality (Merlevède et al., 2011). Assumption 2 is an assumption about the network. Similar assumptions can be found in Zhang et al. (2022, 2020). Assumption 2 (1) is a common independent assumption. Assumptions 2 (2) and (3) are introduced to show that the factor model and latent space model share enough information as $B^{\top}J_pB$ and $\sigma_r(\Omega)$ are not allowed to be too small.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

Assumption 3 is the sparsity assumption of covariance matrix Σ_{e0} which has been used in Bai and Liao (2016). Assumption 4 is commonly used in the latent space model that assumes the network is dense. Assumption 6 (1) is the ρ-mixing condition (Bradley, 2005), which is similar to a strong mixing condition that describes the dependence of two σ-algebra. Assumption 6 (2) aims to bound the variance of most $T^{-1/2}\sum_{t=1}^{T} f_t e_{jt}$ for $j = 1, 2, \ldots, p$ away from 0. Assumption 7 (1) gives the uniform bound of f_t, and Assumptions 7 (2)-(4) are borrowed from Bai and Li (2016). Assumptions 6 and 7 are not necessary for building the convergence results of our method. They are introduced to show that under special cases, the NAAF has a faster convergence rate than ML. Under the above assumptions, we define the parameter space,

$$\begin{aligned}\Xi_\delta =&\{(B, \Sigma_e, \Omega, \alpha) : \delta_1^{-1} \leq \sigma_r(p^{-1}B^\top B) \leq \sigma_1(p^{-1}B^\top B) \leq \delta_1,\\ &\delta_2^{-1} \leq \sigma_r(p^{-1}B^\top J_p B) \leq \sigma_1(p^{-1}B^\top J_p B) \leq \delta_2, B^\top \Sigma_e^{-1} B \text{ is diagonal},\\ &\max_{i,j} |\Theta_{A,ij}| \leq \delta_3, \text{ and } \max\{||\Sigma_e||_1, ||\Sigma_e^{-1}||_1, ||\Sigma_e||_2, ||\Sigma_e^{-1}||_2\} \leq \delta_4, \sigma_r(\Omega) \geq \delta_5 v_p\},\end{aligned}$$

and

$$\Xi_{\Sigma,\delta} = \{(B, \Sigma_e, \Omega, \alpha) : ||\Sigma_e||_0 \leq p + C(\delta)p^2 v_p^{-4}\},$$

for some large enough constants $\delta_1, \delta_2, \delta_3, \delta_4 > 0$, where $C(\delta)$ is a function of δ with relatively small values, and v_p is defined in Assumption 2.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

Theorem 1. *There exists* $C(\delta) > 0$*, such that for* $(B, \Sigma_e, \Omega, \alpha)$ *and* $(B_\star, \Sigma_{e\star}, \Omega_\star, \alpha_\star) \in \Xi_\delta \cap \Xi_{\Sigma,\delta}$ *satisfying* $BB^\top + \Sigma_e = B_\star B_\star^\top + \Sigma_{e\star}$*, and* $\Theta_A(B, \Omega, \alpha) = \Theta_A(B_\star, \Omega_\star, \alpha_\star)$*, there exists an orthogonal matrix* O *such that* $(B, \Sigma_e, \Omega, \alpha) = (B_\star O, \Sigma_{e\star}, O^\top \Omega_\star O, \alpha_\star)$*.*

According to Theorem 1, the models (2.2) and (2.3) can be identifiable if Σ_e is sparse. Next, we consider the consistency of $\widehat{B}$ and $\widehat{\Sigma}_e$, where $(\widehat{B}, \widehat{\Sigma}_e, \widehat{\Omega}, \widehat{\alpha}) = \text{argmin}_{\{(B,\Sigma_e,\Omega,\alpha)\in\Xi_\delta\cap\Xi_{\Sigma,\delta}\}} L(B, \Sigma_e, \Omega, \alpha)$. We first derive a general convergence result in Theorem 2.

Theorem 2. *Suppose Assumptions 1–3, 4(1), and 5 hold. For* $\lambda < +\infty$*, and* $\{\log(p)\}^{2/r_m - 1} \ll T$*, we have* $p^{-1}||\widehat{\Sigma}_e - \Sigma_{e0}||_F^2 = o_p(1)$*,* $\min_{OO^\top = O^\top O = I_r} p^{-1}||\widehat{B}O - B_0||_F^2 = o_p(1)$*,* $\min_{OO^\top = O^\top O = I_r} T^{-1}||O\widehat{F} - F||_F^2 = o_p(1)$*, and* $(pT)^{-1}||\widehat{B}\widehat{F} - B_0 F||_F^2 = o_p(1)$*, where* $r_m^{-1} = \max\{r_4^{-1}, r_5^{-1}\}$*. In addition, if* $\sigma_r(\Omega) = 0$*, the above convergence result holds for* $\lambda \ll d_{p,T} := \max\{\log(p)T/p, \sqrt{\log(p)T}\}$*.*

Theorem 2 shows that NAAF estimators are convergent for any non-random $\lambda < +\infty$, when p goes to infinity not faster than $\exp(T^\iota)$ for some ι, when the $\sigma_r(\Omega) \gg v_p$. If $\sigma_r(\Omega) = 0$ (as discussed in Remark I), choosing an appropriate λ can also ensure the convergence of the estimates. The convergence results are the same as that in Bai and Liao (2016) for the classical approximate factor model. To obtain the convergence rate, we provide Theorem 3.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

Theorem 3. *Suppose Assumptions 1–3 hold. Assume that* $\{\log(p)\}^{2/r_m-1} \ll T \ll p^{4/5}$, $\max\{K_T, v_p^{-1}\} = O(1)$, $D_p \asymp p$, $\lambda \gg d_{p,T}$, *and* $\rho_{p,T} = (\log(p)/T)^{1/4}$. *Then we have*

$$\min_{OO^\top = O^\top O = I_r} p^{-1}||\widehat{B}O - B_0||_F^2 = O_p((\log(p)/T)^{5/4}),$$

$$\text{and } ||\widehat{\Sigma}_Y - \Sigma_{Y_0}||_{\Sigma_{Y_0}} = O_p(p^{1/2}(\log(p)/T)^{5/4} + (\log(p)/T)^{1/4}).$$

According to Theorem 3, the convergence rate of $p^{-1/2}\|\widehat{B}O - B_0\|_F$ is faster is faster than $T^{-1/2}$ for $\log(p) \ll T^{1/5}$. The introduction of network information improves the performance of the factor model in estimating factor loadings. The assumptions on parameters such as D_T, v_p, ρ_T and others in the theorem are made to clarify the convergence rates of the factor loadings and the covariance matrix. For more general cases, please refer to Theorem S.1 in Section S.1 of the supplementary materials

Theorem 4. *Suppose Assumptions 1–3 hold and* Σ_{e0} *is diagonal. Assume that* $\{\log(p)\}^{2/r_m-1} \ll T$. *For* $\lambda \gg d_{p,T}$, *and* $\rho_{p,T} = +\infty$, *we have*

$$\min_{OO^\top = O^\top O = I_r} p^{-1}||\widehat{B}O - B_0||_F^2 = O_p(\log^2(p)/(pT) + (\log(p)/T)^{3/2} + p^{-1}v_p^{-2}).$$

In Theorem 4, we consider the classical factor model in which Σ_{e0} is diagonal. To make the estimate of Σ_{e0} is diagonal, we set $\rho_{p,T} = +\infty$. According to Theorem 4, for $\{\log(p)\}^{\max\{3,2/r_m-1\}} \ll T \ll pv_p^2$, the NAAF has

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

faster convergence rate for factor loadings than $O_p(T^{-1/2})$. Furthermore, when $T \ll \min\{pv_p^2, p^2/\log p\}$, the convergence rate is $O_p((\log(p)/T)^{3/4})$. Since $\{\log(p)\}^{\max\{3,2/r_m-1\}}$ grows extremely slow, Theorem 4 indicates that, in most small T and large p cases, taking the network into consideration has better performance. Meanwhile, the convergence rate in Theorem 4 is faster than that in Theorem 3 , which can be attributed to the non-diagonal assumption of Σ_{e0}. At last, we show that the convergence rate of factor loadings from the ML-based method proposed by Bai and Liao (2016) is not faster than $O_p(T^{-1/2})$.

Proposition 1. *Suppose Assumptions 1–3, 4(1'), 6 and 7 hold. Assume that $p \gg T$. There exist constants $k_1, k_2 > 0$ such that*

$$\liminf_{p,T\to\infty} \Pr\Big(\min_{OO^\top = O^\top O = I_r} p^{-1}||\widehat{B}_{ML}O - B_0||_F^2 \geq T^{-1}k_1\Big) \geq k_2,$$

where $(\widehat{B}_{ML}, \widehat{\Sigma}_{e,ML}) = \operatorname{argmin}_{(B,\Sigma_e)\in\Xi_\delta} L_Y(B, \Sigma_e)$, with the constraint that $\widehat{\Sigma}_e$ is diagonal.

Proposition 1 implies that under $p \gg T$ and some special conditions, the lower bound of factor loadings' ML estimator in Bai and Li (2016) convergence rate is no less than $O_p(T^{-1/2})$.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

4. Simulation Study

In this section, we assess the numerical performance of NAAF and compare it against widely used ML-based and PC-based methods, including the classical principal component analysis (PCA) method, the popular POET method proposed by Fan et al. (2013), the penalized maximum likelihood (PML) method proposed by Bai and Liao (2016), the network assisted PCA based on the Laplace penalty (PC-L) proposed by Yu et al. (2020), the penalized likelihood method based on the normalized Laplacian penalty (ML-nL), and the two-step methods (TSM) mentioned in Subsection 2.3. For more details on the comparison methods, please refer to Section S.9 in the supplementary materials. In Section 4.1, we will briefly introduce the simulation settings. In Section 4.2, the simulation results of different settings and different methods are presented.

4.1 Simulation Settings

We set the number of factors $r = 4$ and assume that r known. To generate Y_t and A, we first generate $\Sigma_{e0}, B_0, \Omega_0$, f_t and e_t. First, we follow Fan et al. (2011) to generate Σ_{e0}. To generate B_0, we first generate $\tilde{B}_0$, such that $(\tilde{B}_0)_{ij} \sim N(0,1)$ independently for $i = 1, \ldots, p$ and $j = 1, \ldots, r$. Let $\Omega_0^* = \text{diag}(6, 5, -5, -6)$. Then, we set $B_0 = \tilde{B}_0 V$, where V is orthogonal

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

matrix such that $V^\top \tilde{B}_0^\top \Sigma_{e0}^{-1} \tilde{B}_0 V$ is diagonal. We set $\Omega_0 = V^\top \Omega_0^* V$. The factors, f_ts, are generated from the AR model: $f_t = 0.2 f_{t-1} + \sqrt{1-0.04} d_t$ where $f_0 = d_0$ and $d_t \sim N(0, I_r)$ independently for $t = 0, 1, 2, ..., T$. Errors e_t with mean 0 and covariance matrix Σ_{e0} are independently drawn from multivariate mixed Gaussian distribution. We generate Y_t by Equation (2.1), for $t = 1, \ldots, T$. To generate A, we consider three examples.

Example I: We independently generate $\alpha_{0,i} \sim \mathcal{U}(-6, -5)$ for $i = 1, \ldots, p$ where $\mathcal{U}$ represents the uniform distribution. Then, we generate A by Equation (2.3). We aim to examine the performance of NAAF when T is relatively small and moderate, respectively. For the case that T is small, we set $T \in \{50, 100\}$ and $p \in \{50, 100, 150\}$. For the case that T is moderate, we set $T \in \{300, 500\}$ and $p \in \{100, 150, 200\}$.

Example II: We consider the effect of the density of the network on the performance of NAAF. We set $\alpha_{0,i} \sim \mathcal{U}(0, 1) - c$ for $i = 1, \ldots, p$ and $c \in \{6, 7, 8, 9\}$. As c increases, the network gets more sparse. The adjacency matrix A is generated by Equation (2.3). In this example, we set $T \in \{300, 500\}$ and $p \in \{100, 150\}$.

Example III: We examine the robustness of NAAF. We generate A in violation of Equation (2.3). Consider three cases: Case (a): generate $B_1 = B_0 + 0.2\mathbf{v}_1$ where $\mathbf{v}_{1,ij} \sim N(0, 1)$ with probability 0.3 and 0 with

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

probability 0.7. Then, we generate A as follows: $A \sim Ber(P), \text{logit}(P) = J_p B_1 \Omega_0 B_1^\top J_p + \alpha_0 \mathbf{1}_p^\top + \mathbf{1}_p \alpha_0^\top$. Case (b): generate $B_2 = B_0 + 0.2\mathbf{v}_2$ where $\mathbf{v}_{2,ij} \sim N(0,1)$. Then, let $A \sim Ber(P)$ with $\text{logit}(P) = J_p B_2 \Omega_0 B_2^\top J_p + \alpha_0 \mathbf{1}_p^\top + \mathbf{1}_p \alpha_0^\top$. The entries of α_0 are also drawn from $\mathcal{U}(-6,-5)$ independently in Cases (a) and (b). Case (c): $A \sim Ber(P)$, where $P_{ij} = \exp(-||b_{0i} - b_{0j}||^2/2)$. Cases (a) and (b) are used to examine NAAF when the assumption that $B_0 \Omega_0 B_0^\top = \Gamma I_{q_1,q_2} \Gamma^\top$ does not hold, where Γ is denoted below Equation (2.2). Case (c) is used to examine NAAF when the latent space model is invalid. We set $T \in \{300, 500\}$ and $p = 100$ in this example.

4.2 Simulation Results

In this subsection, we provide the simulation results. To assess the performance of estimators, we consider the following four measures: (i)$\text{ME}_B = \min_{OO^\top = O^\top O = I_r} p^{-1} ||\widehat{B}O - B_0||_F^2$; (ii)$\text{ME}_{\Sigma_Y} = ||\widehat{\Sigma}_Y - \Sigma_{Y0}||_{\Sigma_{Y0}}$; (iii) $\text{ME}_{\Sigma_e} = ||\widehat{\Sigma}_e - \Sigma_{e0}||_{\Sigma_{e0}}$; and (iv) $\text{ME}_F = \min_{OO^\top = O^\top O = I_r} T^{-1} ||F - O\widehat{F}||_F^2$ For each setting, we conduct 100 realizations.

The results of Example I are provided in Tables 1 and 2. For ME_B, ME_{Σ_Y}, ME_{Σ_e} and ME_F, the NAAF is competitive, especially when T is small. The estimation accuracy can be improved by taking the network information into consideration. The results of Example II are provided in

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

Table 3. As the network becomes denser, the information provided by the network gets more rich, and NAAF has better performance. The results of Example III are reported in Table S.1 in the supplementary materials. As shown in Table S.1, NAAF is robust. Even when Equation (2.3) is invalid, taking network information into consideration still increases the accuracy of the model in terms of ME_B and ME_{Σ_Y}.

5. Real Data Analysis

The study of co-movement in stock returns is crucial in finance. The factor model is an effective tool for capturing this co-movement structure and estimating the covariance matrix. We collect daily returns of the constituent stocks of the CSI 300 Index from the RESSET database, spanning the period from January 1, 2021, to April 30, 2023, including 563 trading days. A limited set of stocks is selected for analysis to avoid a sparse network and ensure computational efficiency. Stocks with incomplete data are excluded, resulting in a final sample of 246 stocks for analysis. Two common strategies are employed to construct networks between stocks. The first is based on industry classification, where a pair of stocks is connected if they share the same industry label (Yu et al., 2020). The second is based on fund co-holding, where two stocks are connected if they are heavily held by the

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

Table 1: Simulation results of Example I, with $(p, T) \in \{100, 150, 200\} \times \{300, 500\}$. Each cell shows the mean$\times$10 (standard error $\times$10).

p	T	ME	NAAF	PML	TSM	PCA	POET	PC-L	ML-nL
100	300	B	0.54(0.05)	0.82(0.07)	2.24(0.23)	0.85(0.07)	0.85(0.07)	0.82(0.07)	0.74(0.06)
		Σ_Y	1.71(0.07)	2.04(0.08)	3.87(0.31)	5.80(0.08)	2.09(0.08)	5.80(0.08)	1.92(0.08)
		Σ_e	0.94(0.06)	0.94(0.06)	1.14(0.12)	5.93(0.07)	1.06(0.09)	5.93(0.07)	0.94(0.06)
		F	2.08(0.09)	2.09(0.09)	2.09(0.09)	2.37(0.10)	2.37(0.10)	2.37(0.10)	2.24(0.11)
	500	B	0.38(0.03)	0.51(0.04)	2.23(0.23)	0.53(0.04)	0.53(0.04)	0.52(0.04)	0.45(0.04)
		Σ_Y	1.40(0.05)	1.56(0.06)	3.83(0.30)	4.49(0.06)	1.63(0.06)	4.49(0.06)	1.49(0.05)
		Σ_e	0.76(0.05)	0.76(0.05)	1.00(0.11)	4.78(0.05)	0.91(0.08)	4.78(0.05)	0.76(0.05)
		F	2.07(0.07)	2.07(0.07)	2.08(0.08)	2.36(0.09)	2.36(0.09)	2.36(0.09)	2.17(0.08)
150	300	B	0.46(0.03)	0.77(0.04)	1.65(0.15)	0.79(0.04)	0.79(0.04)	0.78(0.04)	0.73(0.04)
		Σ_Y	1.65(0.06)	2.06(0.05)	3.45(0.26)	7.09(0.07)	2.08(0.05)	7.08(0.07)	1.97(0.06)
		Σ_e	0.91(0.05)	0.92(0.05)	1.02(0.11)	7.08(0.06)	0.96(0.05)	7.08(0.06)	0.92(0.05)
		F	1.25(0.05)	1.25(0.05)	1.26(0.05)	1.46(0.06)	1.46(0.06)	1.46(0.06)	1.33(0.06)
	500	B	0.32(0.02)	0.47(0.03)	1.64(0.15)	0.48(0.03)	0.48(0.03)	0.48(0.03)	0.44(0.03)
		Σ_Y	1.33(0.04)	1.55(0.04)	3.41(0.26)	5.48(0.05)	1.58(0.05)	5.48(0.05)	1.51(0.04)
		Σ_e	0.72(0.04)	0.73(0.04)	0.86(0.12)	5.60(0.05)	0.79(0.05)	5.60(0.05)	0.73(0.04)
		F	1.24(0.05)	1.24(0.05)	1.24(0.05)	1.44(0.05)	1.44(0.05)	1.44(0.05)	1.27(0.05)
200	300	B	0.42(0.02)	0.78(0.04)	1.50(0.15)	0.79(0.04)	0.79(0.04)	0.78(0.04)	0.75(0.04)
		Σ_Y	1.62(0.05)	2.12(0.05)	3.30(0.24)	8.18(0.07)	2.13(0.05)	8.17(0.07)	2.05(0.05)
		Σ_e	0.94(0.05)	0.95(0.05)	1.01(0.07)	8.11(0.06)	0.96(0.05)	8.11(0.06)	0.95(0.05)
		F	0.97(0.03)	0.98(0.03)	0.98(0.03)	1.10(0.04)	1.10(0.04)	1.10(0.04)	1.03(0.05)
	500	B	0.27(0.02)	0.48(0.03)	1.50(0.15)	0.48(0.03)	0.48(0.03)	0.47(0.03)	0.45(0.02)
		Σ_Y	1.28(0.03)	1.59(0.04)	3.25(0.24)	6.34(0.04)	1.60(0.04)	6.34(0.04)	1.55(0.04)
		Σ_e	0.76(0.04)	0.76(0.04)	0.85(0.07)	6.37(0.04)	0.78(0.05)	6.37(0.04)	0.76(0.04)
		F	0.97(0.04)	0.98(0.04)	0.98(0.04)	1.10(0.04)	1.10(0.04)	1.10(0.04)	1.00(0.04)

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

Table 2: Simulation results of Example I, with $(p, T) \in \{50, 100, 150\} \times \{50, 100\}$. Each cell shows the mean$\times$10 (standard error $\times$10).

p	T	ME	NAAF	PML	TSM	PCA	POET	PC-L	ML-nL
50	50	B	2.43(0.32)	4.84(0.49)	3.48(0.38)	4.89(0.49)	4.89(0.49)	4.48(0.46)	4.34(0.43)
		Σ_Y	3.82(0.27)	5.49(0.34)	4.44(0.34)	9.98(0.35)	5.50(0.32)	9.78(0.34)	4.15(0.28)
		Σ_e	2.21(0.19)	2.32(0.18)	2.30(0.28)	9.33(0.31)	2.52(0.18)	9.30(0.32)	2.34(0.18)
		F	3.60(0.41)	3.72(0.42)	3.65(0.40)	3.81(0.46)	3.81(0.46)	3.81(0.46)	6.47(0.76)
	100	B	1.49(0.20)	2.51(0.26)	3.32(0.38)	2.52(0.27)	2.52(0.27)	2.36(0.26)	2.19(0.25)
		Σ_Y	2.81(0.16)	3.63(0.20)	4.18(0.31)	7.11(0.20)	3.67(0.20)	7.04(0.19)	3.11(0.17)
		Σ_e	1.61(0.13)	1.63(0.13)	1.79(0.18)	7.06(0.16)	1.90(0.15)	7.05(0.16)	1.65(0.13)
		F	3.54(0.24)	3.62(0.24)	3.60(0.23)	3.76(0.24)	3.76(0.24)	3.76(0.24)	4.49(0.40)
100	50	B	1.99(0.17)	4.66(0.36)	2.39(0.22)	4.72(0.37)	4.72(0.37)	4.45(0.35)	4.25(0.33)
		Σ_Y	3.87(0.18)	6.18(0.31)	4.34(0.23)	14.06(0.34)	6.19(0.31)	13.90(0.33)	4.73(0.26)
		Σ_e	2.22(0.16)	2.31(0.16)	2.25(0.20)	13.19(0.32)	2.37(0.15)	13.17(0.32)	2.32(0.16)
		F	2.07(0.26)	2.21(0.30)	2.17(0.28)	2.46(0.32)	2.46(0.32)	2.46(0.32)	4.29(0.64)
	100	B	1.26(0.12)	2.37(0.18)	2.31(0.21)	2.39(0.18)	2.39(0.18)	2.27(0.17)	2.15(0.15)
		Σ_Y	2.83(0.14)	3.89(0.18)	4.09(0.25)	10.02(0.18)	3.90(0.17)	9.96(0.18)	3.38(0.15)
		Σ_e	1.58(0.12)	1.61(0.11)	1.70(0.15)	9.71(0.16)	1.69(0.11)	9.70(0.16)	1.61(0.11)
		F	2.03(0.16)	2.10(0.17)	2.09(0.17)	2.36(0.18)	2.36(0.18)	2.36(0.18)	2.70(0.21)
150	50	B	1.68(0.13)	4.57(0.27)	1.69(0.16)	4.62(0.28)	4.62(0.28)	4.34(0.26)	4.20(0.25)
		Σ_Y	3.75(0.15)	6.80(0.27)	3.95(0.24)	17.25(0.28)	6.82(0.27)	17.05(0.27)	5.03(0.22)
		Σ_e	2.15(0.13)	2.26(0.12)	2.16(0.14)	16.25(0.27)	2.30(0.12)	16.23(0.27)	2.26(0.12)
		F	1.22(0.12)	1.29(0.13)	1.27(0.13)	1.46(0.16)	1.46(0.16)	1.47(0.16)	3.46(0.37)
	100	B	1.01(0.09)	2.30(0.15)	1.64(0.15)	2.32(0.15)	2.32(0.15)	2.23(0.15)	2.14(0.13)
		Σ_Y	2.68(0.09)	4.10(0.15)	3.66(0.25)	12.23(0.17)	4.12(0.15)	12.18(0.17)	3.56(0.14)
		Σ_e	1.53(0.09)	1.58(0.09)	1.60(0.12)	11.84(0.16)	1.61(0.09)	11.83(0.16)	1.58(0.09)
		F	1.24(0.09)	1.27(0.09)	1.27(0.10)	1.46(0.11)	1.46(0.11)	1.46(0.11)	1.80(0.16)

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

Table 3: Simulation results of Example II. Each cell shows the mean×10 (standard error ×10).

		$p = 100$				$p = 150$			
T	c	ME_B	ME_{Σ_Y}	ME_{Σ_e}	ME_F	ME_B	ME_{Σ_Y}	ME_{Σ_e}	ME_F
300	PML	0.80(0.06)	2.03(0.07)	0.96(0.06)	2.08(0.08)	0.78(0.04)	2.06(0.06)	0.91(0.05)	1.25(0.05)
	9	0.64(0.05)	1.85(0.06)	0.96(0.06)	2.07(0.08)	0.59(0.04)	1.82(0.05)	0.91(0.05)	1.25(0.05)
	8	0.61(0.05)	1.81(0.06)	0.96(0.06)	2.07(0.09)	0.55(0.04)	1.76(0.06)	0.91(0.05)	1.25(0.05)
	7	0.58(0.04)	1.78(0.06)	0.96(0.06)	2.07(0.08)	0.48(0.04)	1.67(0.06)	0.91(0.05)	1.25(0.05)
	6	0.55(0.04)	1.74(0.06)	0.95(0.06)	2.07(0.08)	0.44(0.03)	1.61(0.05)	0.91(0.05)	1.24(0.05)
500	PML	0.50(0.04)	1.55(0.06)	0.76(0.05)	2.06(0.07)	0.47(0.03)	1.54(0.04)	0.72(0.04)	1.24(0.04)
	9	0.43(0.04)	1.46(0.05)	0.76(0.05)	2.06(0.07)	0.37(0.02)	1.40(0.04)	0.72(0.04)	1.24(0.04)
	8	0.41(0.04)	1.43(0.05)	0.76(0.05)	2.06(0.07)	0.36(0.02)	1.38(0.04)	0.72(0.04)	1.24(0.04)
	7	0.39(0.03)	1.40(0.05)	0.76(0.05)	2.06(0.07)	0.32(0.02)	1.33(0.04)	0.72(0.04)	1.24(0.04)
	6	0.36(0.03)	1.37(0.05)	0.76(0.05)	2.06(0.07)	0.30(0.02)	1.30(0.04)	0.72(0.04)	1.24(0.04)

same fund (Anton and Polk, 2014).

To provide an intuitive measure of similarity between the networks and the factor loadings, we introduce the following metric $\mathcal{R}$. Specifically, we first estimate $\bar{\Gamma} = J_p\Gamma$ based solely on model (2.2) in the paper, denoted as $\widehat{\bar{\Gamma}}$. Using the principal component method, we obtain an initial estimate of B_0, denoted as $\widehat{B}_{PCA}$. After decentralizing $\widehat{B}_{PCA}$, we project it onto $\widehat{\bar{\Gamma}}$ and calculate the proportion of $\widehat{B}_{PCA}$ that is explained by $\widehat{\bar{\Gamma}}$. That is

$$\mathcal{R} = ||P_{\widehat{\bar{\Gamma}}} J_p \widehat{B}_{PCA}||_F^2 / ||J_p \widehat{B}_{PCA}||_F^2,$$

where $P_{\widehat{\bar{\Gamma}}} = \widehat{\bar{\Gamma}}(\widehat{\bar{\Gamma}}^\top \widehat{\bar{\Gamma}})^{-1}\widehat{\bar{\Gamma}}$. A larger value of $\mathcal{R}$ indicates a higher explanatory

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

power of the network for the factor loadings. The values of $\mathcal{R}$ are 0.30 for the fund-based network and 0.18 for the industry-based network. Both networks capture information about the factor loadings, with the fund-based network demonstrating superior performance.

We apply the NAAF to analyze the stock returns. We denote NAAF(fund) for the fund-based network and NAAF(industry) for the industry-based network. The number of factors is selected as $r = 4$ based on the based on the information criteria proposed by Bai and Ng (2002). The methods mentioned in the simulation, including POET, PML, and ML-nL (denoted as ML-nL(fund) for the fund-based network and ML-nL(industry) for the industry-based network) are used for comparison. Additionally, we consider a regression-based method incorporating industry classification as a covariate into the traditional factor model (denoted as "covar-based"). For more details on the comparison methods, please refer to the Section S.9 in the supplementary materials.

To evaluate the models' performance, we first consider the prediction error. Denote $\{t_i\}_{i=1}^{28}$ as the first trade day of i-th month and $t_{29} = 563 + 1$. Define $Y_{,n_1:n_2}$ as the submatrix consisting of the columns from n_1 to n_2 of the matrix Y. For $i \in \{1, ..., 25\}$, we analyze the submatrix $Y_{,t_i:(t_{i+3}-1)}$ using the proposed methods and comparison methods. Denote the estimators of

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

factor loading matrix as $\widehat{B}_i$ for $i = 1, \cdots, 25$. Denote the prediction error as $(1/25)\sum_{i=1}^{25} \|(I_p - \widehat{B}_i(\widehat{B}_i^\top \widehat{B}_i)^{-1}\widehat{B}_i^\top)(Y_{,t_{i+3}:(t_{i+4}-1)} - \bar{Y}_{,t_{i+3}:(t_{i+4}-1)}\mathbf{1}_{t_{i+4}-t_{i+3}}^\top)\|_F$ for all methods except "covar-based", where $\bar{Y}_{,t_{i+3}:(t_{i+4}-1)}$ is the row mean of $Y_{,t_{i+3}:(t_{i+4}-1)}$. For the covar-based, the prediction error is defined as $(1/25)\sum_{i=1}^{25} \|(I_p - \widehat{B}_i(\widehat{B}_i^\top \widehat{B}_i)^{-1}\widehat{B}_i^\top)\{Y_{,t_{i+3}:(t_{i+4}-1)} - (\bar{Y}_{,t_{i+3}:(t_{i+4}-1)} + X_c\widehat{\beta}_i)\mathbf{1}_{t_{i+4}-t_{i+3}}^\top\}\|_F$, where X_c is the design matrix for the industry, and $\widehat{\beta}_i$ is the coefficient vector. The results of the prediction error is provided in Table 4. Methods that incorporate networks perform better than those that do not. Methods that incorporate networks perform better than those that do not. The NAAF(fund) outperforms the alternatives. The NAAF(fund) performs better than NAAF(industry), because the fund-based network more effectively captures the co-movement of stocks. We also conduct pairwise Wilcoxon tests with the null hypothesis that NAAF(funds) does not have smaller prediction errors than the compared methods. The resulting p-values are all smaller than 10^{-3}. These tests support the conclusion that NAAF(fund) has smaller prediction errors. The introduction of appropriate networks can enhance the performance of factor models.

Next, we construct portfolios based on the estimated covariance matrix to evaluate the performance of the proposed models. Denote the estimated covariance matrix of of $Y_{,t_i:(t_{i+3}-1)}$ as $\widehat{\Sigma}_{Y,i}$. We adopt the minimum-variance

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

portfolio, a widely recognized and established method in portfolio optimization (Xue et al., 2012; Zou et al., 2017). Specifically, we solve the following optimization problem to determine the portfolio weights:

$$\widehat{\omega}_i = \text{argmin}_{\omega^\top \mathbf{1}_p=1, \omega \geq 0} \, \omega^\top \widehat{\Sigma}_{Y,i} \omega.$$

Then, we calculate the cumulative return of this portfolio within this month and obtain a monthly return series. We calculate the three commonly used measures of return series for different methods: (a) mean, which represents the geometric average of the returns on investment portfolios. (b) SD, which represents the standard deviation of returns on investment portfolios. (c) Sharpe ratio, which measures the earnings under the same risk. The results are provided in Table 4. The NAAF(fund) outperforms the alternatives in terms of mean, SD, and Sharpe ratio. The assistance of an appropriate network improves the effectiveness of portfolio construction.

Table 4: The real data results.

	NAAF(fund)	NAAF(industry)	PML	ML-nL(fund)	ML-nL(industry)	POET	covar-based
prediction error	1.4244	1.4293	1.4294	1.4280	1.4268	1.4297	1.4422
mean($\times 10^3$)	2.4563	1.7547	1.6192	1.8236	2.0053	1.4499	1.4686
SD($\times 10^2$)	2.2458	2.3524	2.2796	2.2775	2.2658	2.4336	2.3129
Sharpe ratio	0.1201	0.0859	0.0820	0.0910	0.0994	0.0712	0.0745

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

6. Conclusion

Accurately estimating the loadings and factors are crucial in factor models. In this paper, we take the network information into account and propose the NAAF model, which significantly mitigates the necessity of a large sample size for consistent estimation. We propose a joint quasi-maximum likelihood estimator of the factor loadings and the covariance matrix of idiosyncratic errors. Under mild conditions, the NAAF yields a significant improvement in the rates of convergence than the regular methods. An efficient gradient descent algorithm is developed to optimize the objective function. The asymptotic properties of estimators are established. Lots of numerical studies demonstrate the practical use of NAAF.

To broaden the applications of NAAF, we identify avenues for future research. The first is considering multiple networks between individuals. The second is to study the theoretical properties and algorithm of the model extension when the dimension of latent space exceeds the number of factors. Third, a nonparametric relationship between factor loadings and latent locations can be considered. Finally, the impact of combining factor and network models on node community detection or link prediction can be explored. We believe that these efforts would further increase the applicability of the NAAF model.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

Supplementary Material

The Supplementary Material consists of ten sections (S.1–S.10). Section S.1 provides a more general form of Theorem 3. Section S.2 introduces some useful notations and lemmas that are used to prove the theoretical properties in Section 3. Sections S.3–S.7 present the proofs of Theorems 1, 2, S.1 and 3, 4, and Proposition 1, respectively. Section S.8 provides additional algorithmic details. Section S.9 details the comparison methods. Section S.10 presents additional simulation results.

Acknowledgments

This work is supported by National natural Science Foundation of China (72271232, 71873137, 12201626), the MOE Project of Key Research Institute of Humanities and Social Sciences (22JJD110001), and Public Computing Cloud of Renmin University of China.

References

Anton, M. and C. Polk (2014). Connected stocks. *The Journal of Finance 69*(3), 1099–1127.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

REFERENCES

Bai, J. and K. Li (2012). Statistical analysis of factor models of high dimension. *The Annals of Statistics 40*(1), 436–465.

Bai, J. and K. Li (2016). Maximum likelihood estimation and inference for approximate factor models of high dimension. *Review of Economics and Statistics 98*(2), 298–309.

Bai, J. and Y. Liao (2016). Efficient estimation of approximate factor models via penalized maximum likelihood. *Journal of Econometrics 191*(1), 1–18.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bien, J. and R. J. Tibshirani (2011). Sparse estimation of a covariance matrix. *Biometrika 98*(4), 807–820.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys 2*, 107–144.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica: Journal of the Econometric Society*, 1281–1304.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *The Journal of Finance 47*(2), 427–465.

Fan, J., Y. Liao, and M. Mincheva (2011). High dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics 39*(6), 3320–3356.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

REFERENCES

Fan, J., Y. Liao, and M. Mincheva (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 75*(4), 603–680.

Fan, J., Y. Liao, and W. Wang (2016). Projected principal component analysis in factor models. *The Annals of Statistics 44*(1), 219–254.

Fountalis, I., A. Bracco, and C. Dovrolis (2014). Spatio-temporal network analysis for studying climate patterns. *Climate Dynamics 42*, 879–899.

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.

Huang, J. and L. Yang (2010). Correlation matrix with block structure and efficient sampling methods. *Journal of Computational Finance 14*(1), 81–94.

Krivitsky, P. N., M. S. Handcock, A. E. Raftery, and P. D. Hoff (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks 31*(3), 204–213.

Lin, F. and Z. Qiu (2024). Pairs trading strategy and connected stocks: Evidence from china. *Available at SSRN 4790701*.

Linton, O. and G. Connor (2000). Semiparametric estimation of a characteristic-based factor model of stock returns. Technical report, Financial Markets Group.

Mayrink, V. D. and J. E. Lucas (2013). Sparse latent factor models with interactions: Analysis

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

of gene expression data. *The Annals of Applied Statistics*, 799–822.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology 27*(1), 415–444.

Merlevède, F., M. Peligrad, and E. Rio (2011). A bernstein type inequality and moderate deviations for weakly dependent sequences. *Probability Theory and Related Fields 151*(3), 435–474.

Rubin-Delanchy, P., J. Cape, M. Tang, and C. E. Priebe (2022). A statistical interpretation of spectral embedding: the generalised random dot product graph. *Journal of the Royal Statistical Society Series B: Statistical Methodology 84*(4), 1446–1473.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological) 58*(1), 267–288.

Von Ferber, C., T. Holovatch, Y. Holovatch, and V. Palchykov (2009). Public transport networks: empirical analysis and modeling. *The European Physical Journal B 68*, 261–275.

Wong, S. L., L. V. Zhang, A. H. Tong, Z. Li, D. S. Goldberg, O. D. King, G. Lesage, M. Vidal, B. Andrews, H. Bussey, et al. (2004). Combining biological networks to predict genetic interactions. *Proceedings of the National Academy of Sciences 101*(44), 15682–15687.

Xie, W.-J., M.-X. Li, Z.-Q. Jiang, Q.-Z. Tan, B. Podobnik, W.-X. Zhou, and H. E. Stanley (2016). Skill complementarity enhances heterophily in collaboration networks. *Scientific reports 6*(1), 18727.

Statistica Sinica: Preprint
doi:10.5705/ss.202024.0170

REFERENCES

Xue, L., S. Ma, and H. Zou (2012). Positive-definite ℓ1-penalized estimation of large covariance matrices. *Journal of the American Statistical Association 107*(500), 1480–1491.

Yi, H., Q. Zhang, C. Lin, and S. Ma (2022). Information-incorporated gaussian graphical model for gene expression data. *Biometrics 78*(2), 512–523.

Yu, L., Y. He, X. Zhang, and J. Zhu (2020). Network-assisted estimation for large-dimensional factor model with guaranteed convergence rate improvement. *arXiv preprint arXiv:2001.10955*.

Zhang, X., G. Xu, and J. Zhu (2022). Joint latent space models for network data with high-dimensional node variables. *Biometrika 109*(3), 707–720.

Zhang, X., S. Xue, and J. Zhu (2020). A flexible latent space model for multilayer networks. In *International Conference on Machine Learning*, pp. 11288–11297. PMLR.

Zou, T., W. Lan, H. Wang, and C.-L. Tsai (2017). Covariance regression analysis. *Journal of the American Statistical Association 112*(517), 266–281.

Center for Applied Statistics and School of Statistics, Renmin University of China

E-mail: 2021103748@ruc.edu.cn

Center for Applied Statistics and School of Statistics, Renmin University of China

E-mail: 1031820039@qq.com

Center for Applied Statistics and School of Statistics, Renmin University of China

E-mail: mabzhang@ruc.edu.cn