

Mixed Membership Estimation in Bayesian Network Autoregression

ENDONG WANG[†] AND SIAO XU[‡]

[†]*Department of Economics, University of Mannheim, Mannheim, 68161, Germany.*

E-mail: `endong.wang@uni-mannheim.de`

[‡]*Department of Economics, University of Mannheim, Mannheim, 68161, Germany.*

E-mail: `siao.xu@students.uni-mannheim.de`

Summary We propose a generative network autoregressive model in which the network coefficient is a random matrix following a joint distribution specified by a Bayesian belief network. Rather than treating the coefficient as fixed and observed, we model it as the output of a latent generative process that simultaneously governs the network structure and the resulting panel data. A central feature is that units may exhibit mixed membership across latent groups rather than belonging to a single group. We develop an estimator that recovers the latent membership structure from panel observations, establish identification and consistency as both panel dimensions grow, and assess finite-sample performance through simulations. An application to the U.S. financial system documents substantial mixed exposure across latent groups, showing that mixed membership materially shapes the dependence structure among financial firms.

JEL Classification: C14, C32, C38, G10, G17.

Keywords: *Network Autoregression, Mixed Membership, Latent Group, Bayesian Belief Network, Cross-sectional Dependence.*

1. INTRODUCTION

Learning latent group structure from panel data is a central problem in econometrics. A large literature studies this problem through grouped panel models, where heterogeneity is represented by a finite number of latent groups and each unit belongs to one group only, see Bonhomme and Manresa (2015); Su et al. (2016); Chetverikov and Manresa (2022); Mugnier (2025). Related ideas have entered the analysis of cross-sectional dependence in dynamic systems since the development of network autoregression Zhu et al. (2017). In that literature, the network coefficient governing cross-sectional dynamic dependence is often assumed to exhibit latent group structure, and the econometric task is to recover group memberships from the panel induced by that matrix, see Zhu and Pan (2020); Chen et al. (2023); Guðmundsson and Brownlees (2021); Guðmundsson (2026). Whether in the group panel model or in the network autoregression with group structure mentioned above, a common maintained restriction is that the latent structure takes the form of pure membership. That is, one unit is 100% affiliated with one specific group and 0% affiliated with the others. For example, a U.S. voter exclusively belongs to either the Democrat or Republican. A firm exclusively belongs to one specific industry.

That restriction can be too strong in applications. For example, U.S. voters interchangeably vote for the Republican and Democrat. A fintech company may belong to both the financial and the high-tech industries. People may have mixed ancestry from different nations. The mixed membership structure exists everywhere. Furthermore, in many economic settings, units are linked to several latent groups at once, and their

dynamic dependence reflects that mixed membership structure. Financial institutions provide a natural example. Large firms often combine commercial banking, securities intermediation, asset management, insurance, payments, and real-estate-related activities within the same organization. More generally, economic units are often shaped by several latent dimensions of interaction at the same time. In such settings, a pure-membership representation may be systematically misleading. The issue is not simply that units differ. The issue is that the heterogeneity is organized through mixed affiliation rather than through pure affiliation.

This paper develops a generative mixed-membership network autoregression for panel data. The observed panel enters through a large-dimensional network autoregression, while the coefficient matrix, assumed to be a network governing cross-sectional dynamic dependence, is treated as a latent random object. Rather than taking that matrix as fixed and directly observed, we model it as generated by an underlying probabilistic mechanism encoded through a Bayesian belief network. This yields a joint description of latent memberships, coefficient formation, and panel dynamics. It also changes the econometric problem in a basic way. The matrix carrying the mixed-membership structure is not observed and must be recovered from the panel data. The matrix denotes the group structure of interest in the present model.

Our contribution is threefold. At the modeling level, we introduce mixed membership and generative structure into a dynamic environment in which the dependence matrix is latent. Existing grouped specifications in panel data and related large network autoregressive formulations are organized around hard assignment. By contrast, our framework allows each unit to load on several latent groups simultaneously. This extension is economically important in settings where interaction patterns reflect overlapping affiliations rather than a sharp partition.

At the econometric level, we develop a feasible estimator for this setting by combining ordinary least squares with an adaptation of SCORE Jin (2015); Jin et al. (2024). The procedure first estimates the network coefficient of a large network autoregression from panel data, then uses its spectral structure to recover mixed memberships. The theoretical analysis establishes identification of the latent membership structure and derives consistency and error bounds for the estimator. The argument links the leading eigenspace of the population coefficient matrix to the underlying mixed memberships and then controls the estimation error.

At the empirical level, the paper shows why the mixed-membership extension matters substantively. We apply the method to daily volatility data for U.S. financial firms and ask whether the cross section is better described by overlapping latent dependence groups than by an exclusive partition. This is a natural environment for the question, since large financial institutions rarely operate along a single business line and the dependence structure in firm-level volatility need not coincide with conventional sectoral boundaries. The estimates point to substantial overlap. In the application, many institutions lie well inside the simplex, the average largest membership weight is only 0.632, and an exclusive classification would mask an important part of the estimated dependence structure. The empirical finding is therefore not merely that firms are heterogeneous. It is that the heterogeneity takes the form of overlapping latent group exposure.

The paper is related to several strands of the literature. First, it contributes to the literature on latent group structure in panel data, including Brownlees et al. (2022); Su et al. (2016); Bonhomme and Manresa (2015); Zhang et al. (2019). Relative to that work, our framework replaces pure membership with mixed membership in a dynamic

setting with cross-sectional dependence. Second, it relates to the literature on mixed-membership models and their estimation in networks, including Airolidi et al. (2008); Mao et al. (2017, 2021); Jin et al. (2024); Zhang et al. (2020). The key difference is that those papers recover mixed memberships from an observed network or adjacency object, whereas in our setting the relevant dependence matrix is latent and must be inferred from panel dynamics. Third, it connects to the literature on network autoregressions and large VARs, including Zhu et al. (2017, 019a,b); Guðmundsson and Brownlees (2021). Our contribution there is to allow the latent group structure of the dependence matrix to be mixed rather than exclusive. Finally, the paper draws on probabilistic graphical models and Bayesian belief networks as a way to encode the latent generative structure, see Blei et al. (2003); Munro and Ng (2022); Koller and Friedman (2009).

The remainder of the paper is organized as follows. Section 2 introduces the model, identification, estimation, and asymptotic theory. Section 3 reports the finite-sample evidence. Section 4 presents the empirical application to the U.S. financial system. Section 5 concludes. Technical proofs are collected in the Appendix.

Notation. Let N denote the number of time series and T the number of observations. Of the N series, N_0 are pure-membership units and $N - N_0$ are mixed-membership units. Lowercase letters denote scalars, and uppercase letters denote matrices. For a matrix Y , Y_i , Y_j , and Y_{ij} denote its i -th row, j -th column, and (i, j) entry, respectively. For a square matrix Y , let $\lambda_i(Y)$ denote its i -th eigenvalue, ordered by magnitude, and let $\rho(Y)$ denote its spectral radius. For a symmetric matrix Y , $\lambda_{\max}(Y)$ and $\lambda_{\min}(Y)$ denote its largest and smallest eigenvalues. We write $\|Y\|$ for the spectral norm and $\|Y\|_F$ for the Frobenius norm. A directed weighted graph is denoted by $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{W})$, where $\mathcal{V} = \{1, \dots, N\}$, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, and $\mathcal{W}$ assigns edge weights. Its adjacency matrix is $A \in \mathbb{R}^{N \times N}$, with $A_{ij} \neq 0$ if there is an edge from node j to node i , and $A_{ij} = 0$ otherwise.

2. MODEL, IDENTIFICATION, AND ESTIMATION

2.1. The mixed membership network autoregressive model

We study a mixed membership network autoregressive model in which a latent weighted network determines cross-sectional dynamic dependence. The model separates three layers: mixed memberships determine link probabilities, realized links and their weights form a latent weighted network, and a degree-normalized version of this network governs the autoregressive dynamics.

Let K denote the number of latent groups. For each unit $i = 1, \dots, N$, let $\pi_i \in \mathbb{R}^K$ denote an unknown membership vector, with nonnegative entries summing to one, and let $\Pi \in \mathbb{R}^{N \times K}$ collect the row vectors $\pi'_1, \dots, \pi'_N$. Let $A \in \{0, 1\}^{N \times N}$ denote the adjacency matrix, $B \in \mathbb{R}^{K \times K}$ the group interaction matrix, $\Theta = \text{diag}(\theta_1, \dots, \theta_N)$ the diagonal matrix of activity parameters, and $\Phi \in \mathbb{R}^{N \times N}$ the latent weighted network. Let $D = \text{diag}(d_1, \dots, d_N)$, where $d_i = \sum_{j=1}^N \Phi_{ij} + \sum_{j=1}^N \Phi_{ji}$, and define the normalized network coefficient matrix

$$\mathbf{\Phi} = \rho D^{-1/2} \Phi D^{-1/2}, \quad 0 < \rho < 1. \quad (2.1)$$

For an isolated node with $d_i = 0$, we adopt the convention $D_{ii}^{-1/2} = 0$. Finally, $\mu \in \mathbb{R}^N$ and $\sigma > 0$ are the parameters in the conditional Gaussian law of the observed process.

The model is defined as follows.

DEFINITION 2.1. (MIXED MEMBERSHIP NETWORK AUTOREGRESSION) *Given parameters $\Pi, B, \Theta, (\underline{u}, \bar{u}), \rho, \mu$, and σ , where $0 < \underline{u} < \bar{u}$ and $0 < \rho < 1$, the latent weighted network $\Phi = (\Phi_{ij})_{i,j=1}^N$ is generated elementwise as*

$$\Phi_{ij} \mid A_{ij}, (\underline{u}, \bar{u}) \sim \begin{cases} \text{Uniform}(\underline{u}, \bar{u}), & \text{if } A_{ij} = 1, \\ 0, & \text{if } A_{ij} = 0, \end{cases} \quad (2.2)$$

$$A_{ij} \mid \Pi, B, \Theta \sim \text{Bernoulli}(\theta_i \pi_i' B \pi_j \theta_j), \quad (2.3)$$

for $i, j = 1, \dots, N$. Conditional on (Π, B, Θ) , the link indicators $\{A_{ij}\}_{i,j=1}^N$ are mutually independent. Conditional on A , the weights $\{\Phi_{ij}\}_{i,j=1}^N$ are mutually independent. Let $D = \text{diag}(d_1, \dots, d_N)$, with $d_i = \sum_{j=1}^N \Phi_{ij} + \sum_{j=1}^N \Phi_{ji}$, and define $\Phi = \rho D^{-1/2} \Phi D^{-1/2}$, with the convention $D_{ii}^{-1/2} = 0$ when $d_i = 0$. Conditional on Φ , the observed process follows

$$Y_t \mid Y_{t-1}, \mu, \sigma, \Phi \sim \mathcal{N}(\mu + \Phi Y_{t-1}, \sigma^2 I_N), \quad t = 1, \dots, T. \quad (2.4)$$

Definition 2.1 distinguishes the latent weighted network Φ from the coefficient matrix Φ governing the dynamics. The indicator A_{ij} determines whether the directed link from unit j to unit i is active, while, conditional on $A_{ij} = 1$, Φ_{ij} measures its raw strength. These nonzero weights need not shrink as N increases. The degree normalization maps the realized weighted network into $\Phi = \rho D^{-1/2} \Phi D^{-1/2}$, thereby controlling aggregate network propagation without requiring individual active links to vanish.

The membership vector π_i describes unit i 's exposure to the K latent groups. The matrix B governs interactions between these groups in link formation, while Θ captures unit-specific network activity. Hence the model separates membership structure, link formation and strength, and the propagation of the resulting network through the observed dynamics.

We maintain the restriction $\theta_i \pi_i' B \pi_j \theta_j \in [0, 1]$ for all i and j , so that the implied link probabilities are well defined. The degree normalization also provides a convenient sufficient condition for stability. For any unit vectors x, y , Cauchy–Schwarz gives $|x' D^{-1/2} \Phi D^{-1/2} y| \leq \left(\sum_{i,j} \Phi_{ij} \frac{x_i^2}{d_i} \right)^{1/2} \left(\sum_{i,j} \Phi_{ij} \frac{y_j^2}{d_j} \right)^{1/2} \leq 1$, where zero-degree terms are interpreted as zero. Hence $\|D^{-1/2} \Phi D^{-1/2}\| \leq 1$, and therefore $\|\Phi\| \leq \rho < 1$. The autoregressive process is therefore stable without requiring the nonzero entries of Φ to shrink with N : degree normalization controls aggregate network exposure, while ρ provides a strict stability margin. This normalization closely parallels that used in the SB-VAR(1) model of Guðmundsson and Brownlees (2021), where the normalized network is likewise scaled by a common autoregressive parameter.

Figure 1 summarizes the hierarchical structure of the model. At the top level, the latent objects Π, B , and Θ determine the link-indicator matrix A . The matrix A determines which entries of the latent weighted network Φ are active, while $(\underline{u}, \bar{u})$ determine the support of the corresponding nonzero weights. The realized network is then degree-normalized and scaled by ρ to form the autoregressive coefficient matrix Φ . Conditional on Φ, μ , and σ , the observed process follows a first-order large-dimensional vector autoregression.

Conditional on the initial value Y_0 , and holding $\Pi, \mu, \sigma, \underline{u}, \bar{u}, \rho, \Theta$, and B fixed, the

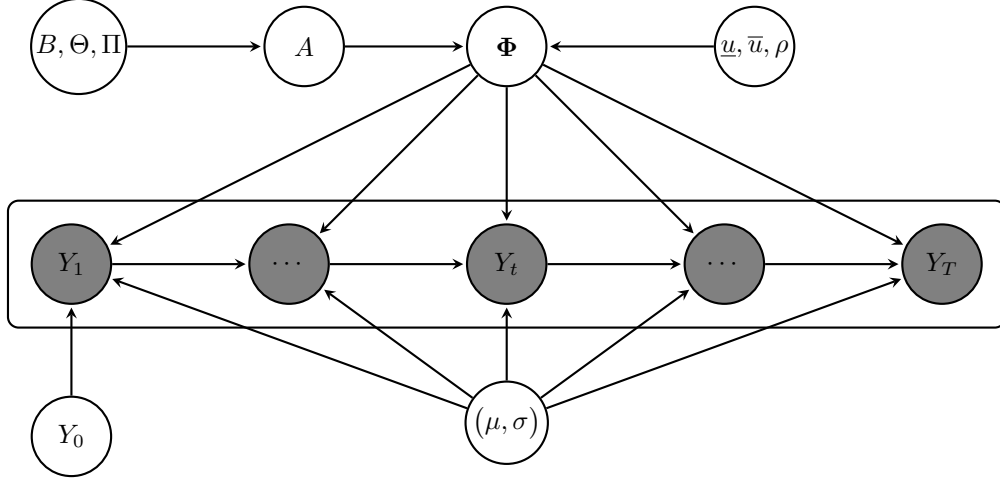


Figure 1. Graphical representation of the MMNAR model. The latent objects Π , B , and Θ determine the link-indicator matrix A , which determines the sparsity pattern of the latent weighted network Φ . The realized network is degree-normalized and scaled by ρ to form Φ , which governs the dynamics of the observed process. Shaded nodes are observed.

recursive structure of the model implies the factorization

$$\begin{aligned} \mathbb{P}(\{Y_t\}_{t=1}^T, \Phi, A) &= \left\{ \prod_{t=1}^T \mathbb{P}(Y_t \mid Y_{t-1}, \Phi, \mu, \sigma) \right\} \mathbb{P}(\Phi \mid A, \underline{u}, \bar{u}) \\ &\times \left\{ \prod_{i=1}^N \prod_{j=1}^N \mathbb{P}(A_{ij} \mid B, \pi_i, \pi_j, \Theta) \right\}. \end{aligned} \quad (2.5)$$

Integrating over link realizations and link weights yields a simple population representation of the latent weighted network.

THEOREM 2.1. *Under Definition 2.1,*

$$\mathbb{E}(\Phi) = \frac{\underline{u} + \bar{u}}{2} \Theta \Pi B \Pi' \Theta.$$

Theorem 2.1 links the latent membership structure to the population weighted network. Its first moment inherits the low-rank mixed-membership form $\Theta \Pi B \Pi' \Theta$, up to the common mean weight $(\underline{u} + \bar{u})/2$. This representation provides the population object from which the mixed-membership structure is identified. The proof in the Appendix follows directly by integrating over the latent adjacency matrix and the active link weights.

2.2. Identification

Our objective is to recover the mixed membership matrix Π . We first study structural identification from the mean latent weighted network. By Theorem 2.1, $\mathbb{E}(\Phi) = \frac{\underline{u} + \bar{u}}{2} \Theta \Pi B \Pi' \Theta$. Throughout this subsection, the objects Π , B , Θ , and $(\underline{u}, \bar{u})$ are treated as fixed unknown parameters.

The following conditions are sufficient for identification.

ASSUMPTION 2.1. *Under Definition 2.1,*

- 1 $B + B'$ is positive definite, with $B_{kk} = 1$ for $k = 1, \dots, K$ and $B_{ij} > 0$ for all i, j .
- 2 For each community $k = 1, \dots, K$, there exists at least one pure node, namely an i such that $\pi_{ik} = 1$.
- 3 The degree parameters satisfy $\theta_i > 0$ for all i .

The restrictions have distinct identifying roles. Positivity of B keeps the link-probability index nonnegative, while $B_{kk} = 1$ eliminates community-specific scale normalizations. Pure nodes anchor the K latent-community directions, and $\theta_i > 0$ ensures that these directions are preserved under node-specific activity scaling.

The next theorem states the identification result.

THEOREM 2.2. *Suppose that a MMNAR satisfying Definition 2.1 also satisfies Assumption 2.1. If two admissible parameter representations generate the same mean latent weighted network $\mathbb{E}(\Phi)$, with membership matrices Π and $\tilde{\Pi}$, respectively, then*

$$\tilde{\Pi} = \Pi P \quad (2.6)$$

for some $P \in \mathcal{P}_K$, where $\mathcal{P}_K$ denotes the set of $K \times K$ permutation matrices. Hence Π is identified up to a permutation of the community labels.

Theorem 2.2 establishes identification of the membership matrix from the mean latent weighted network. Equality of $\mathbb{E}(\Phi)$ also implies equality of its symmetric part, whose group-level interaction matrix is $(B + B')/2$. Positive definiteness preserves all K latent directions, pure nodes anchor these directions, and the unit-diagonal normalization of B removes community-specific rescaling. Finally, the unit-sum restriction on each membership vector recovers the membership shares. Thus Π is identified up to relabeling without requiring the remaining nuisance parameters to be separately identified. Related arguments appear in Latouche et al. (2011) and Zhang et al. (2020).

Theorem 2.2 concerns structural identification from the latent population network. For estimation, the relevant population object is its symmetrized degree-normalized counterpart. For estimation, it is sufficient that degree normalization preserves the mixed-membership geometry required for recovery of Π . Let $\bar{D} = \mathbb{E}(D) = \text{diag}(\bar{d}_1, \dots, \bar{d}_N)$ denote the population degree matrix, where $\bar{d}_i = \mathbb{E}(d_i) = \sum_{j=1}^N \mathbb{E}(\Phi_{ij}) + \sum_{j=1}^N \mathbb{E}(\Phi_{ji})$. Define the symmetrized population normalized network

$$\bar{\Phi}^S = \frac{\rho}{2} \bar{D}^{-1/2} \{ \mathbb{E}(\Phi) + \mathbb{E}(\Phi)' \} \bar{D}^{-1/2}. \quad (2.7)$$

By Theorem 2.1, $\bar{\Phi}^S = \rho(\underline{u} + \bar{u})/2 \bar{D}^{-1/2} \Theta \Pi \{ (B + B')/2 \} \Pi' \Theta \bar{D}^{-1/2}$. The next lemma links the latent membership structure to its leading eigenspace.

LEMMA 2.1. *Let the MMNAR in Definition 2.1 satisfy Assumption 2.1. Let $x_1, \dots, x_K$ denote the eigenvectors associated with the K nonzero eigenvalues of $\bar{\Phi}^S$, and define $X = [x_1, \dots, x_K]$. Then there exists a unique nonsingular matrix $P \in \mathbb{R}^{K \times K}$ such that*

$$X = \bar{D}^{-1/2} \Theta \Pi P. \quad (2.8)$$

Lemma 2.1 shows that symmetrization and degree normalization preserve the mixed-membership eigenspace structure. In particular, the population eigenspace spans the same column space as $\bar{D}^{-1/2}\Theta\Pi$. Since $\bar{D}^{-1/2}\Theta$ is diagonal, the SCORE ratio transformation removes this nuisance scale and preserves the simplex geometry generated by Π . Hence recovery of Π does not require inversion of the degree normalization or separate recovery of the latent scale parameters.

2.3. Estimation

We estimate Π in two stages. First, we estimate the normalized autoregressive coefficient matrix Φ by ordinary least squares. Second, we symmetrize this estimate and apply a Mixed-SCORE decomposition to recover the membership profiles. Throughout, K is known and fixed as $N, T \rightarrow \infty$.

To remove the intercept, the observations are demeaned over time. With a slight abuse of notation, continue to write Y_t for the demeaned observations, and let $\mathbf{Y} = [Y_2, \dots, Y_T]$ and $\mathbf{X} = [Y_1, \dots, Y_{T-1}]$. Provided $\mathbf{X}\mathbf{X}'$ is invertible,

$$\hat{\Phi} = \mathbf{Y}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}. \quad (2.9)$$

For the spectral analysis, we use $\hat{\Phi}^S = (\hat{\Phi} + \hat{\Phi}')/2$, which aligns the sample object with the symmetric population matrix Φ^S in (2.7).

Algorithm 1 MMNAR-Mixed-SCORE

Input: panel data $\{Y_t\}_{t=1}^T$, number of communities K .

- 1: Estimate the coefficient matrix $\hat{\Phi}$ by OLS.
- 2: Symmetrize the estimate, $\hat{\Phi}^S = (\hat{\Phi} + \hat{\Phi}')/2$.
- 3: Let $\hat{X} \in \mathbb{R}^{N \times K}$ collect the eigenvectors associated with the K largest eigenvalues of $\hat{\Phi}^S$.
- 4: Set the threshold $S = \log N$, and construct $\hat{R} \in \mathbb{R}^{N \times (K-1)}$ by
$$\hat{R}_{ik} = \text{sign}\left(\hat{X}_{i,k+1}/\hat{X}_{i1}\right) \min\left\{|\hat{X}_{i,k+1}/\hat{X}_{i1}|, S\right\}, \quad 1 \leq i \leq N, \quad 1 \leq k \leq K-1.$$
- 5: Find K corner points $\{\hat{v}_1, \dots, \hat{v}_K\}$ in $\hat{R}$ using a vertex-hunting routine, such as sketched vertex search.
- 6: For each $i = 1, \dots, N$, compute barycentric weights by solving

$$\hat{w}_i = \arg \min_{w \in \mathbb{R}^K} \left\| \hat{R}_i - \sum_{k=1}^K w_k \hat{v}_k \right\|^2 \quad \text{subject to} \quad w_k \geq 0, \quad \sum_{k=1}^K w_k = 1.$$

- 7: For $1 \leq i \leq N$ and $1 \leq k \leq K$, define

$$\hat{\pi}_{ik}^* = \max\left\{0, \hat{w}_{ik}/\hat{b}_{1k}\right\}, \quad \hat{b}_{1k} = \left[\hat{\lambda}_1 + \hat{v}_k' \text{diag}(\hat{\lambda}_2, \dots, \hat{\lambda}_K) \hat{v}_k\right]^{-1/2}.$$

- 8: Normalize,

$$\hat{\pi}_i = \hat{\pi}_i^* / \|\hat{\pi}_i^*\|_1, \quad i = 1, \dots, N.$$

Output: $(N \times K)$ matrix of estimated mixed memberships $\hat{\Pi}$.

By Lemma 2.1, the leading population eigenspace spans the same column space as $\bar{D}^{-1/2}\Theta\Pi$. Since $\bar{D}^{-1/2}\Theta$ is diagonal, the SCORE ratio transformation removes this nuisance scale and maps the rows into a simplex: pure units determine its vertices, while mixed units lie in its interior. Membership recovery therefore reduces to vertex estimation followed by barycentric reconstruction and normalization.

The estimator differs from standard applications of SCORE and Mixed-SCORE in one important respect. In those settings, the relevant network matrix is directly observed. Here, by contrast, the normalized network representation is not directly observed and must first be estimated from the time-series observations through the large VAR. The algorithm therefore combines time-series estimation with spectral recovery of the mixed-membership structure.

2.4. Large-sample theory

Throughout the large-sample analysis, $K, \rho \in (0, 1)$, $\underline{u}, \bar{u}$, and σ are fixed as $N, T \rightarrow \infty$. We consider the regime $N \log N/T \rightarrow 0$, under which the unrestricted VAR coefficient matrix can be estimated by ordinary least squares. The logarithmic strengthening relative to $N/T \rightarrow 0$ is used only for uniform row-wise control of the SCORE denominator; the average membership rate continues to depend on the time dimension through N/T .

We now establish consistency of the MMNAR-Mixed-SCORE estimator. The population target for spectral recovery is $\bar{\Phi}^S$ in (2.7), which retains the degree-corrected mixed-membership structure characterized in Lemma 2.1. The following conditions control its signal strength and SCORE geometry.

ASSUMPTION 2.2. *Under Definition 2.1, define*

$$G = \frac{K}{\|\Theta\|_F^2} \Pi' \Theta^2 \Pi. \quad (2.10)$$

There exists a constant $z > 1$, independent of N and T , such that:

- 1 $N^{-1}\|\Theta\|_F^2 \leq z$, $\Theta_{\max} \leq z\Theta_{\min}$, and $\log N/\|\Theta\|_F^2 \rightarrow 0$.
- 2 $z^{-1} \leq \lambda_{\min}((B+B')/2) \leq \lambda_{\max}((B+B')/2) \leq z$, $\|G\| \leq z$, and $\|G^{-1}\| \leq z$.
- 3 Let $\lambda_1^* \geq \dots \geq \lambda_K^* > 0$ denote the eigenvalues of $\frac{B+B'}{2}\Pi'\Theta\bar{D}^{-1}\Theta\Pi$, and let η_1 be the unit-norm right eigenvector associated with λ_1^* , normalized so that its entries are positive. Then $\lambda_1^* - \lambda_2^* \geq z^{-1}\lambda_1^*$ and $\max_{1 \leq k \leq K} \eta_1(k) \leq z \min_{1 \leq k \leq K} \eta_1(k)$.

Parts 1 and 2 of Assumption 2.2 control aggregate node activity and the conditioning of the latent-group structure. In particular, G is the activity-weighted Gram matrix of the membership profiles, so $\|G^{-1}\| \leq z$ rules out latent membership directions carrying asymptotically negligible θ_i^2 -weighted mass, while the eigenvalue bounds on $(B+B')/2$ separately keep group interactions nondegenerate. Part 3 separates and balances the leading spectral direction used as the SCORE denominator. Although the matrix in part 3 need not be symmetric, it is similar to the symmetric positive-definite matrix $(\Pi'\Theta\bar{D}^{-1}\Theta\Pi)^{1/2}(B+B')/2(\Pi'\Theta\bar{D}^{-1}\Theta\Pi)^{1/2}$, so its eigenvalues are real and positive. The Appendix shows that these conditions keep the nonzero eigenvalues of $\bar{\Phi}^S$ away from zero and imply $X_{i1} \asymp N^{-1/2}$ uniformly in i .

Write the MMNAR as

$$Y_t = \mu + \Phi Y_{t-1} + \varepsilon_t. \quad (2.11)$$

The time-series estimation argument is carried out conditional on the realized latent graph and the corresponding normalized coefficient matrix. Let $\mathcal{G}$ denote the realized latent graph. For this argument, we work with the conditionally centered process and, with a slight abuse of notation, continue to write Y_t for $Y_t - \mathbb{E}(Y_t \mid \mathcal{G}, \Phi)$.

ASSUMPTION 2.3. *Conditional on $(\mathcal{G}, \Phi)$, the process is initialized from its stationary distribution. As $N, T \rightarrow \infty$, $N \log N/T \rightarrow 0$.*

Together with the Gaussian innovations in Definition 2.1 and $\|\Phi\| \leq \rho < 1$, Assumption 2.3 yields a stationary Gaussian process with uniformly bounded covariance and geometrically decaying temporal dependence. These properties deliver the concentration bounds required for the OLS first stage; see the Appendix.

We next turn to the sample problem. The perturbation has two components. The first is the discrepancy between the realized normalized network Φ^S and its population counterpart $\bar{\Phi}^S$; this component includes both the realization error in the weighted network and the replacement of the realized degree matrix D by its population counterpart $\bar{D} = \mathbb{E}(D)$. The second is the time-series estimation error $\hat{\Phi}^S - \Phi^S$.

LEMMA 2.2. *Suppose that the MMNAR in Definition 2.1 satisfies Assumptions 2.1, 2.2, and 2.3. Then*

$$\|\hat{\Phi}^S - \bar{\Phi}^S\| = O_p \left(\sqrt{\frac{\log N}{\|\Theta\|_F^2}} + \sqrt{\frac{N}{T}} \right). \quad (2.12)$$

The first term in (2.12) is the network-realization error after degree normalization, while the second is the time-series estimation error. Under Assumption 2.2, the balance condition on Θ , the conditioning of G , and the normalization $B_{kk} = 1$ imply $\bar{d}_i \gtrsim \|\Theta\|_F^2$ uniformly in i . The Appendix controls the realized weighted network and the realized degree matrix jointly to obtain the first rate.

Lemma 2.2 provides the basic perturbation bound. The first term reflects network-realization uncertainty, including the random weighted network and the replacement of the realized degree matrix by its population counterpart. The second term is the time-series estimation error arising from recovery of the normalized coefficient matrix by OLS. Consistent recovery therefore requires both sufficient aggregate network activity and a sufficiently long time dimension.

The next result transfers this bound to the leading eigenspace.

LEMMA 2.3. *Suppose that the conditions of Lemma 2.2 hold. Let $\hat{X}$ collect the K leading eigenvectors of $\hat{\Phi}^S$, and let X be defined as in Lemma 2.1. Then there exists an orthonormal matrix $\mathcal{O} = \text{diag}(1, Q)$, with $Q \in \mathbb{R}^{(K-1) \times (K-1)}$, such that*

$$\|\hat{X} - X\mathcal{O}\| = O_p \left(\sqrt{\frac{\log N}{\|\Theta\|_F^2}} + \sqrt{\frac{N}{T}} \right). \quad (2.13)$$

Lemma 2.3 controls the estimated signal eigenspace in spectral norm and determines

the average error propagated through the SCORE step. For the ratio construction, the Appendix additionally establishes a row-wise refinement sufficient to keep the estimated leading eigenvector uniformly at its $N^{-1/2}$ population scale. This refinement is used only to control the SCORE denominator and does not affect the average membership rate.

ASSUMPTION 2.4. *There exists a constant $c_0 > 0$, independent of N and T , such that, for every $k = 1, \dots, K$, at least $c_0 N$ units satisfy $\pi_i = e_k$. The vertex-hunting routine is strongly efficient: there exists a constant C , independent of N and T , such that, for any orthogonal $Q \in \mathbb{R}^{(K-1) \times (K-1)}$, whenever $N^{-1} \sum_{i=1}^N \|\hat{R}_i - R_i Q\|^2 \leq \delta^2$,*

$$\min_{\tau} \max_{1 \leq k \leq K} \|\hat{v}_k - v_{\tau(k)} Q\|^2 \leq C \delta^2, \quad (2.14)$$

where the minimum is over permutations of the K vertices.

Under Assumption 2.2, $X_{i1} \asymp N^{-1/2}$ uniformly in i , and the population SCORE coordinates are uniformly bounded. The row-wise eigenspace refinement established in the Appendix shows that the sample leading eigenvector remains uniformly at the same $N^{-1/2}$ scale with probability approaching one. Hence $S = \log N$ is asymptotically non-binding, while the average SCORE error is governed by the global eigenspace bound in Lemma 2.3.

THEOREM 2.3. *Suppose that the MMNAR in Definition 2.1 satisfies Assumptions 2.1, 2.2, 2.3, and 2.4. Then*

$$\min_{P \in \mathcal{P}_K} \frac{1}{N} \|\hat{\Pi} - \Pi P\|_F^2 = O_p \left(\frac{\log N}{\|\Theta\|_F^2} + \frac{N}{T} \right). \quad (2.15)$$

In particular, $\min_{P \in \mathcal{P}_K} N^{-1} \|\hat{\Pi} - \Pi P\|_F^2 = o_p(1)$.

Theorem 2.3 gives the convergence rate directly for the estimated membership profiles. The first term is governed by the aggregate node-activity signal $\|\Theta\|_F^2$, while the second reflects estimation of the normalized coefficient matrix from the time series. Under the maintained assumptions, the MMNAR-Mixed-SCORE estimator is therefore consistent up to a permutation of the community labels. If node activity is uniformly bounded away from zero, then $\|\Theta\|_F^2 \asymp N$, and the rate reduces to $O_p(\log N/N + N/T)$.

3. FINITE-SAMPLE PERFORMANCE

3.1. Simulation design

This section studies the finite-sample performance of MMNAR-Mixed-SCORE under the data-generating process in Definition 2.1. The design varies three features that are central for recovery of the membership structure: the time dimension, the separation of the latent communities, and the number of pure nodes. Throughout, the number of latent communities is fixed at $K = 3$, and the cross-sectional dimension is fixed at $N = 120$.

The community interaction matrix B is taken to be symmetric. Its diagonal entries govern within-community interaction strength, whereas its off-diagonal entries govern between-community interaction strength. We consider three interaction designs, with $(b_{ii}, b_{ij}) \in \{(0.25, 0.01), (0.25, 0.05), (0.5, 0.01)\}$. These choices vary the sharpness of the

latent block structure. The design $(0.25, 0.05)$ yields the weakest separation between within- and between-community interactions, whereas $(0.5, 0.01)$ yields the strongest.

We also vary the number of pure nodes. Let N_0 denote the number of pure nodes, so that $N - N_0$ is the number of mixed-membership nodes. In the main design reported in Table 1, we consider $N_0 \in \{30, 50, 80\}$. For pure nodes, π_i is one of the canonical basis vectors $[1, 0, 0]$, $[0, 1, 0]$, or $[0, 0, 1]$. For mixed-membership nodes, we use the two profiles $[0.5, 0.5, 0]$ and $[0, 0.5, 0.5]$. This construction preserves a transparent geometry while allowing for nontrivial overlap across communities.

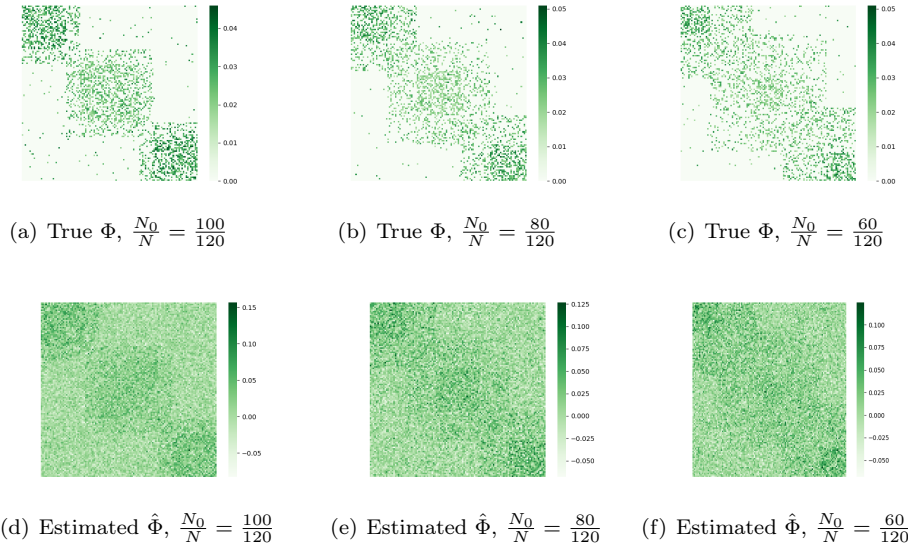


Figure 2. Heatmaps of the true and estimated coefficient matrices for three values of the pure-node fraction, holding the interaction design fixed. Moving from left to right, the share of pure nodes falls from $100/120$ to $60/120$. As the share of mixed-membership nodes rises, the block structure becomes less distinct, and the recovery of Φ becomes less accurate.

The remaining parameters are held fixed across designs. The support of active link weights is set to $[\underline{u}, \bar{u}] = [0.3, 1.0]$, and the heterogeneity parameters $\{\theta_i\}_{i=1}^N$ are evenly spaced over $[0.9, 1.0]$. For each specification, we generate data from the MMNAR model and repeat the experiment 500 times. Estimation accuracy is measured by the average squared ℓ_2 loss, $N^{-1} \sum_{i=1}^N \|\hat{\pi}_i - \pi_i\|^2$, and Table 1 reports the mean of this loss across replications. The figures reported below provide heatmap illustrations of the associated coefficient matrices.

3.2. Finite-sample results

Table 1 reports the main Monte Carlo results. The most stable pattern concerns the time dimension. For every interaction design and every value of N_0 , the estimation error declines as T increases. This pattern is consistent with the structure of the estimator.

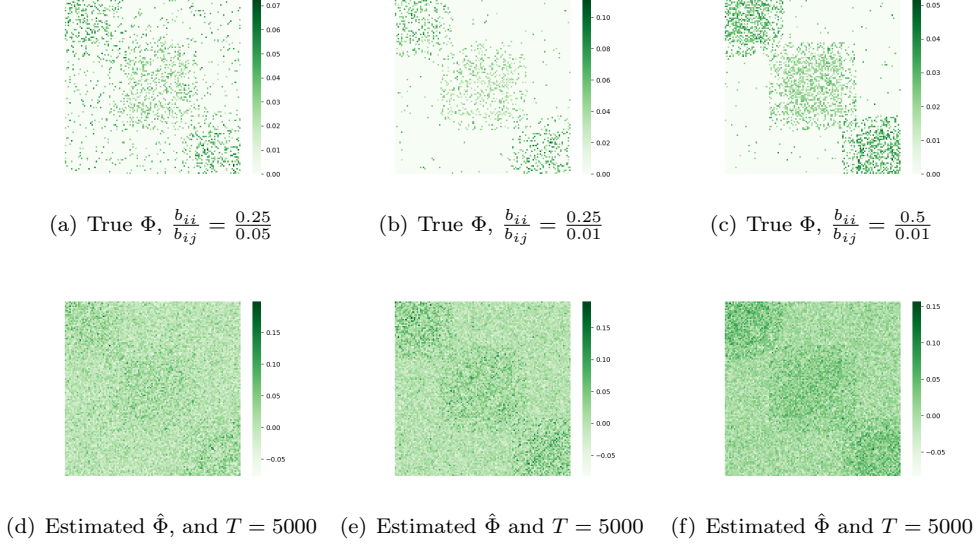


Figure 3. Heatmaps of the true and estimated coefficient matrices under alternative interaction designs and time dimensions. A larger contrast between within-community and between-community interaction yields a sharper block pattern in the population matrix. For a fixed interaction design, the estimated matrix tracks that pattern more closely as T increases.

A longer time series sharpens first-step recovery of the coefficient matrix Φ , and this improvement carries over to the spectral recovery of the membership simplex.

A second pattern concerns the separation of the latent block structure. Holding N_0 and T fixed, estimation error is generally smaller when within-community interaction is stronger relative to between-community interaction. In particular, the design $(0.5, 0.01)$ typically outperforms $(0.25, 0.05)$, with $(0.25, 0.01)$ lying between them. This ordering is natural. Stronger within-community interaction yields a clearer separation across latent groups, so the leading eigenspace of the coefficient matrix more sharply captures the underlying mixed-membership structure.

The role of pure nodes also emerges in the table, although more clearly once the time dimension is sufficiently large. For moderate and large values of T , designs with more pure nodes tend to produce lower estimation error. For example, under $(0.5, 0.01)$ and $T = 5000$, the loss falls from 0.064 at $N_0 = 30$ to 0.036 at $N_0 = 50$ and 0.031 at $N_0 = 80$. The ordering is less regular when T is small. This is consistent with the geometry of the problem. Pure nodes sharpen the simplex only to the extent that the first-step estimate of Φ is already precise enough for that geometry to be visible in the sample. When the time series is short, estimation noise in $\hat{\Phi}$ can obscure part of the gain from additional anchors.

Figure 3 complements Table 1 by illustrating the role of block separation and the time dimension. The top row reports the true coefficient matrices under the three interaction designs. As the gap between b_{ii} and b_{ij} widens, the latent block structure becomes more pronounced. The remaining rows report the corresponding estimates at $T = 2000$ and

<i>Mixed membership</i>					
$N = 120$					
	$T = 1000$	2000	3000	4000	5000
$N_0 = 30$					
0.25/0.01	0.14	0.12	0.109	0.099	0.092
0.50/0.01	0.13	0.11	0.094	0.076	0.064
0.25/0.05	0.146	0.142	0.138	0.135	0.130
$N_0 = 50$					
0.25/0.01	0.16	0.12	0.096	0.073	0.060
0.50/0.01	0.15	0.11	0.067	0.048	0.036
0.25/0.05	0.17	0.15	0.14	0.12	0.11
$N_0 = 80$					
0.25/0.01	0.17	0.12	0.085	0.052	0.040
0.50/0.01	0.16	0.10	0.066	0.040	0.031
0.25/0.05	0.19	0.16	0.13	0.114	0.094

Table 1. Estimation error of MMNAR-Mixed-SCORE.

Each entry reports the average over 500 replications of $N^{-1} \sum_{i=1}^N \|\hat{\pi}_i - \pi_i\|^2$, where $\hat{\pi}_i$ denotes the estimated membership vector and π_i denotes its population counterpart.

$T = 5000$. For a given interaction design, the estimated coefficient matrix tracks the population structure more closely as T increases.

Figure 2 isolates the role of pure nodes while holding the interaction design fixed. As the share of mixed-membership nodes rises, the block pattern becomes less distinct, and recovery of Φ becomes less accurate. This is the finite-sample manifestation of the pure-node condition underlying identification. Pure nodes determine the vertices of the simplex, whereas a larger mass of mixed nodes compresses the geometry of the membership space and makes recovery more demanding.

The simulation evidence points to a clear conclusion. MMNAR-Mixed-SCORE performs better when the time dimension is longer, the latent block structure is more sharply separated, and the design contains a larger set of pure nodes. These features strengthen the spectral signal associated with the mixed-membership structure and improve finite-sample recovery of Π .

4. EMPIRICAL APPLICATION: MIXED-MEMBERSHIP ESTIMATION IN THE U.S. FINANCIAL SYSTEM

We apply the MMNAR framework to daily volatility data for U.S. financial firms to assess whether cross-sectional dependence is better described by mixed exposure across latent groups than by an exclusive partition. The distinction is economically relevant: large financial institutions routinely combine commercial banking, securities intermediation, asset management, insurance, payments, and real-estate activities, so a pure-membership specification may impose a sharper classification than the data support.

The sample consists of U.S. financial firms with SIC codes between 6000 and 6799,

covering depository and non-depository institutions, broker-dealers, insurers, real-estate-related firms, and holding and investment offices. Daily prices are obtained from Yahoo Finance for the period from 3 January 2013 to 31 December 2024. After calendar alignment and removal of incomplete histories, the balanced panel contains $N = 98$ institutions and $T = 3018$ trading days. For firm i on day t , we use the Parkinson range estimator

$$\tilde{\sigma}_{i,t}^2 = 0.361(p_{i,t}^{high} - p_{i,t}^{low})^2, \quad (4.1)$$

where high and low prices are split-adjusted using the close-to-adjusted-close ratio. The empirical input is the panel of log volatilities $\log(\tilde{\sigma}_t^2)$.

We estimate the VAR

$$\log(\tilde{\sigma}_t^2) = \Phi_0 + \Phi_1 \log(\tilde{\sigma}_{t-1}^2) + \epsilon_t, \quad (4.2)$$

with lag order one, in line with Guðmundsson and Brownlees (2021). The estimated system is dynamically stable, with a largest eigenvalue 0.865. We fix $K = 3$; formal selection of the number of groups lies outside the scope of this application. We estimate Φ_1 by OLS, symmetrize, and apply MMNAR-Mixed-SCORE to obtain the 98×3 membership matrix $\hat{\Pi}$, whose rows are nonnegative and sum to one.

Table 2 reports the estimated memberships. The central finding is that mixed exposure, rather than exclusive assignment, dominates the cross section. The average largest membership weight is only 0.632: 83 of 98 firms have a largest weight below 0.8, 77 below 0.7, only 12 are estimated as exact pure members, and 16 load positively on all three groups. Many large institutions lie well inside the simplex. GS, JPM, MMC, NDAQ, SPGI, and FITB are close to an even split between groups 2 and 3, while BAC, BRK-B, ICE, MA, CME, TFC, NTRS, and CCI display nontrivial exposure to all three. Pure members are the exception, confined to firms such as AFL, AXP, BRO, IVZ, KKR, MAA (group 2), BX, V, HST (group 3), and FIS, SBAC, UDR (group 1).

The estimated structure is also markedly uneven across groups. Assigning each firm to its largest weight allocates only 6 firms to group 1, against 42 and 50 to groups 2 and 3, with average shares of 0.079, 0.453, and 0.469. Although the latent groups are unobserved, the estimated pattern admits a restrained reading: group 2 concentrates institutions associated with broad intermediation activities, including insurers and brokers (AFL, BRO) and asset managers (IVZ, KKR); group 3 is more heavily populated by market-based and asset-price-sensitive institutions (BX, V, IBKR, MET, STT, and several REITs); group 1 is a smaller, more specialized component (FIS, SBAC, UDR, with nontrivial weights for BK, PFG, CCI, JKHY, FDS). These descriptions are not literal industry labels, and the recovered groups do not reproduce SIC categories: firms from similar segments need not load on the same group, and firms from different segments may exhibit closely related membership vectors.

The application yields two conclusions. First, cross-sectional heterogeneity in the U.S. financial system is well captured by a small number of latent groups. Second, this heterogeneity does not take the form of an exclusive partition; the prevalence of interior membership vectors implies that hard classification would discard economically consequential variation that the mixed-membership specification preserves.

Firm	G1	G2	G3	Firm	G1	G2	G3	Firm	G1	G2	G3
AFL	0	1	0	GS	0	0.532	0.468	PSA	0.264	0.736	0
ALL	0	0.483	0.517	HIG	0	0.492	0.508	O	0	0.531	0.469
AXP	0	1	0	HBAN	0	0.414	0.586	REG	0.146	0.356	0.498
AIG	0	0.462	0.538	IBKR	0	0.207	0.793	SBAC	1	0	0
AMP	0	0.510	0.490	ICE	0.083	0.488	0.429	SPG	0	0.367	0.633
AON	0	0.511	0.489	IVZ	0	1	0	UDR	1	0	0
APO	0	0.691	0.309	JKHY	0.357	0.086	0.557	VTR	0	0.289	0.711
ACGL	0	0.539	0.461	JPM	0	0.533	0.467	WELL	0	0.149	0.851
AJG	0	0.695	0.305	KEY	0	0.361	0.639	WY	0	0.626	0.374
AIZ	0	0.406	0.594	KKR	0	1	0	WRB	0	0.532	0.468
BAC	0.135	0.237	0.628	L	0	0.482	0.518	WFC	0	0.327	0.673
BRK-B	0.289	0.327	0.383	MTB	0	0.220	0.780	WTW	0	0.433	0.567
BLK	0	0.559	0.441	MMC	0	0.500	0.500	ARE	0	0.431	0.569
BX	0	0	1	MA	0.067	0.509	0.424	AMT	0	0.514	0.486
BK	0.571	0.327	0.102	MET	0	0.238	0.762	AVB	0	0.598	0.402
BRO	0	1	0	MCO	0.087	0.533	0.381	BXP	0	0.577	0.423
COF	0	0.503	0.497	MS	0	0.574	0.426	CPT	0.071	0.463	0.466
CBOE	0	0.456	0.544	MSCI	0	0.514	0.486	CBRE	0	0.755	0.245
SCHW	0	0.428	0.572	NDAQ	0	0.487	0.513	CSGP	0.059	0.508	0.433
CB	0	0.003	0.997	NTRS	0.031	0.149	0.820	CCI	0.471	0.292	0.238
CINF	0	0.483	0.517	PNC	0	0.494	0.506	DLR	0	0.464	0.536
C	0.400	0	0.600	PFG	0.632	0.368	0	EQIX	0	0.451	0.549
CME	0.019	0.441	0.540	PGR	0	0.565	0.435	EQR	0	0.612	0.388
CPAY	0.090	0.300	0.611	PRU	0	0.391	0.609	ESS	0	0.434	0.566
ERIE	0	0.514	0.486	RJF	0	0.529	0.471	EXR	0	0.418	0.582
EG	0	0.484	0.516	RF	0	0.571	0.429	FRT	0	0.682	0.318
FDS	0.379	0	0.621	SPGI	0	0.493	0.507	DOC	0	0.611	0.389
FIS	1	0	0	STT	0	0.379	0.621	HST	0	0	1
FITB	0	0.496	0.504	TROW	0.307	0.390	0.303	IRM	0	0.571	0.429
FISV	0	0.471	0.529	TRV	0	0.454	0.546	KIM	0	0.414	0.586
BEN	0	0.509	0.491	TFC	0.270	0.319	0.411	MAA	0	1	0
GPN	0	0.334	0.663	USB	0	0.651	0.349	PLD	0	0.645	0.355
GL	0	0.470	0.530	V	0	0	1				

Table 2. Estimated three-group mixed memberships of U.S. financial firms. Each row reports the estimated membership vector for one institution. The columns G1, G2, and G3 denote the three latent groups discussed in the text. Entries are nonnegative and sum to one. Larger entries indicate stronger exposure to the corresponding group.

5. CONCLUSION

This paper studies a mixed membership network autoregressive model for panel data in which the network coefficient matrix is latent and generated by a Bayesian belief network. The main departure from existing grouped formulations is that units are not restricted to an exclusive partition. Instead, each unit may exhibit mixed exposure across latent groups. This structure allows the model to capture overlapping forms of cross-sectional heterogeneity in dynamic dependence.

We develop an econometric approach for recovering the latent membership structure from panel observations. We establish identification of the mixed-membership component and show that the proposed estimator is consistent as both dimensions of the panel grow. The empirical application to the U.S. financial system indicates that mixed membership is substantively important. Many institutions exhibit non-negligible exposure across more than one latent group, so a hard classification misses an important feature of the underlying dependence structure. More broadly, the results suggest that allowing

for mixed exposure across latent groups can improve empirical analysis in large network autoregressive systems when cross-sectional dependence is not well described by disjoint latent types.

ACKNOWLEDGEMENTS

We are very grateful to Carsten Trenkler, Jean-Marie Dufour, Matthew O. Jackson, Mengshan Xu, Evan Munro, Christopher Rothe, Frank Schorfheide, Weining Wang, Amilcar Velez, Bin Peng, for their very insightful suggestions and discussions. Seminar participants at Universität Mannheim have provided invaluable feedback. All faults are our own. This work was supported by the research funds provided by the University of Mannheim.

REFERENCES

- Airoldi, E. M., D. M. Blei, S. E. Fienberg, and E. P. Xing (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research* 9(65), 1981–2014.
- Blei, D. M., A. Y. Ng, and M. I. Jordan (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research* 3, 993–1022.
- Bonhomme, S. and E. Manresa (2015). Grouped patterns of heterogeneity in panel data. *Econometrica* 83(3), 1147–1184.
- Brownlees, C., G. S. Guðmundsson, and G. Lugosi (2022). Community detection in partial correlation network models. *Journal of Business & Economic Statistics* 40(1), 216–226.
- Cape, J., M. Tang, and C. E. Priebe (2019). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *The Annals of Statistics* 47(5), 2405–2439.
- Chen, E. Y., J. Fan, and X. Zhu (2023). Community network auto-regression for high-dimensional time series. *Journal of Econometrics* 235(2), 1239–1256.
- Chetverikov, D. and E. Manresa (2022). Spectral and post-spectral estimators for grouped panel data models. *arXiv e-prints*, arXiv–2212.
- Guðmundsson, G. S. (2026). Detecting giver and receiver spillover groups in large vector autoregressions. *Journal of Business & Economic Statistics* 44(1), 297–308.
- Guðmundsson, G. S. and C. Brownlees (2021). Detecting groups in large vector autoregressions. *Journal of Econometrics* 225(1), 2–26.
- Jin, J. (2015). Fast community detection by SCORE. *The Annals of Statistics* 43(1), 57–89.
- Jin, J., Z. T. Ke, and S. Luo (2024). Mixed membership estimation for social networks. *Journal of Econometrics* 239(2), 105369.
- Koller, D. and N. Friedman (2009). *Probabilistic graphical models: principles and techniques*. MIT press.
- Latouche, P., E. Birmelé, and C. Ambroise (2011). Overlapping stochastic block models with application to the French political blogosphere. *The Annals of Applied Statistics* 5(1), 309–336.
- Mao, X., P. Sarkar, and D. Chakrabarti (2017). On mixed memberships and symmetric nonnegative matrix factorizations. In *International Conference on Machine Learning*, pp. 2324–2333. PMLR.
- Mao, X., P. Sarkar, and D. Chakrabarti (2021). Estimating mixed memberships with sharp eigenvector deviations. *Journal of the American Statistical Association* 116(536), 1928–1940.

- Mugnier, M. (2025). A simple and computationally trivial estimator for grouped fixed effects models. *Journal of Econometrics* 250, 106011.
- Munro, E. and S. Ng (2022). Latent Dirichlet analysis of categorical survey responses. *Journal of Business & Economic Statistics* 40(1), 256–271.
- Su, L., Z. Shi, and P. C. B. Phillips (2016). Identifying latent structures in panel data. *Econometrica* 84(6), 2215–2264.
- Zhang, Y., E. Levina, and J. Zhu (2020). Detecting overlapping communities in networks using spectral methods. *SIAM Journal on Mathematics of Data Science* 2(2), 265–283.
- Zhang, Y., H. J. Wang, and Z. Zhu (2019). Quantile-regression-based clustering for panel data. *Journal of Econometrics* 213(1), 54–67.
- Zhu, X., X. Chang, R. Li, and H. Wang (2019b). Portal nodes screening for large scale social networks. *Journal of Econometrics* 209(2), 145–157.
- Zhu, X. and R. Pan (2020). Grouped network vector autoregression. *Statistica Sinica* 30(3), 1437–1462.
- Zhu, X., R. Pan, G. Li, Y. Liu, and H. Wang (2017). Network vector autoregression. *The Annals of Statistics* 45(3), 1096–1123.
- Zhu, X., W. Wang, H. Wang, and W. K. Härdle (2019a). Network quantile autoregression. *Journal of Econometrics* 212(1), 345–358.

APPENDIX A: PROOFS OF RESULTS

PROOF. (**PROOF OF THEOREM 2.1**) Treat Π as fixed throughout. By construction of the model, $\mathbb{P}(\Phi, A) = \mathbb{P}(\Phi \mid A, \underline{u}, \bar{u})\mathbb{P}(A \mid \Theta, B, \Pi)$. Hence, by iterated expectations,

$$\mathbb{E}(\Phi) = \mathbb{E}[\mathbb{E}(\Phi \mid A)]. \quad (\text{A.1})$$

Now fix (i, j) . If $A_{ij} = 0$, then $\Phi_{ij} = 0$. If $A_{ij} = 1$, then $\Phi_{ij} \mid A_{ij} = 1 \sim \text{Uniform}(\underline{u}, \bar{u})$, so $\mathbb{E}(\Phi_{ij} \mid A) = \frac{\underline{u} + \bar{u}}{2} A_{ij}$. Therefore, $\mathbb{E}(\Phi_{ij}) = \frac{\underline{u} + \bar{u}}{2} \mathbb{E}(A_{ij})$. Under Definition 2.1, $A_{ij} \sim \text{Bernoulli}(\theta_i \pi'_i B \pi_j \theta_j)$, and hence $\mathbb{E}(A_{ij}) = \theta_i \pi'_i B \pi_j \theta_j$. Collecting entries yields $\mathbb{E}(\Phi) = \frac{\underline{u} + \bar{u}}{2} \Theta \Pi B \Pi' \Theta$, which establishes the claim.

PROOF. (**PROOF OF THEOREM 2.2**) Let $c_r = (\underline{u}_r + \bar{u}_r)/2$ and $X_r = \Theta_r \Pi_r$ for the two admissible representations $r = 1, 2$. Equality of the mean latent weighted networks implies

$$c_1 X_1 B_1 X_1' = c_2 X_2 B_2 X_2'. \quad (\text{A.2})$$

By the pure-node condition and positivity of the degree parameters, each X_r has full column rank and contains a positive multiple of each canonical basis vector among its rows. Moreover, $B_r + B_r'$ is positive definite, so B_r is nonsingular. Consequently, both sides of (A.2) have rank K , and $\text{Col}(X_1) = \text{Col}(X_2)$. Hence $X_2 = X_1 Q$ for some nonsingular $K \times K$ matrix Q .

Because each X_r is nonnegative and contains a positive multiple of every canonical basis vector, the cone generated by its rows is the positive orthant. Thus Q maps the positive orthant onto itself. An invertible linear map with this property is a positive monomial matrix, so $Q = DP$, where D is positive diagonal and $P \in \mathcal{P}_K$.

Substituting $X_2 = X_1 DP$ into (A.2) and using the full column rank of X_1 gives $c_1 B_1 = c_2 D P B_2 P' D$. Since both B_1 and B_2 have unit diagonal, comparison of the diagonal entries implies $D = \sqrt{c_1/c_2} I_K$. Therefore $X_2 = \sqrt{c_1/c_2} X_1 P$.

Finally, because the rows of each Π_r sum to one, row normalization of $X_r = \Theta_r \Pi_r$ recovers Π_r . The common scalar $\sqrt{c_1/c_2}$ therefore cancels, yielding $\Pi_2 = \Pi_1 P$. Hence the membership matrix is identified up to a permutation of the community labels.

PROOF. (PROOF OF LEMMA 2.1) Let $c = (\underline{u} + \bar{u})/2$ and $H = (B + B')/2$. By (2.7), $\bar{\Phi}^S = \rho c \bar{D}^{-1/2} \Theta \Pi H \Pi' \Theta \bar{D}^{-1/2}$. By Assumption 2.1, H is positive definite. The pure-node condition implies that Π has rank K , while positivity of the degree parameters implies that $\bar{D}^{-1/2} \Theta \Pi$ also has rank K .

It follows that $\bar{\Phi}^S$ has rank K , and its column space coincides with that of $\bar{D}^{-1/2} \Theta \Pi$. Hence the eigenvectors associated with the K nonzero eigenvalues of $\bar{\Phi}^S$ form a basis for the same column space. Therefore there exists a nonsingular $P \in \mathbb{R}^{K \times K}$ such that $X = \bar{D}^{-1/2} \Theta \Pi P$. Since $\bar{D}^{-1/2} \Theta \Pi$ has full column rank, P is unique once X is fixed.

PROOF. (PROOF OF LEMMA 2.2) Write $\Phi^S = (\Phi + \Phi')/2$. By the triangle inequality,

$$\|\hat{\Phi}^S - \bar{\Phi}^S\| \leq \|\hat{\Phi}^S - \Phi^S\| + \|\Phi^S - \bar{\Phi}^S\|. \quad (\text{A.3})$$

We first control the network-realization term. Let $c = (\underline{u} + \bar{u})/2$, $H = (B + B')/2$, and $s = \Pi' \Theta \mathbf{1}_N$. By Theorem 2.1,

$$\bar{d}_i = 2c \theta_i \pi'_i H s. \quad (\text{A.4})$$

Since $G = K \Pi' \Theta^2 \Pi / \|\Theta\|_F^2$ and $\|G^{-1}\| \leq z$, $e'_k G e_k \geq z^{-1}$. Therefore $\sum_j \theta_j^2 \pi_{jk}^2 \geq \|\Theta\|_F^2 / (Kz)$. Since $\theta_j^2 \pi_{jk}^2 \leq \Theta_{\max} \theta_j \pi_{jk}$,

$$s_k \geq \frac{\|\Theta\|_F^2}{Kz\Theta_{\max}}, \quad k = 1, \dots, K. \quad (\text{A.5})$$

Because $H_{kk} = 1$, $H_{k\ell} > 0$, and $\Theta_{\max} \leq z\Theta_{\min}$,

$$\bar{d}_i \geq \frac{2c}{Kz^2} \|\Theta\|_F^2. \quad (\text{A.6})$$

For the upper bound, $\|H\| \leq z$, $\|s\| \leq \sqrt{N} \|\Theta\|_F$, and the balance condition implies $\theta_i \leq z \|\Theta\|_F / \sqrt{N}$. Hence

$$\bar{d}_i \leq 2cz^2 \|\Theta\|_F^2. \quad (\text{A.7})$$

Thus, uniformly in i ,

$$\bar{d}_i \asymp \|\Theta\|_F^2. \quad (\text{A.8})$$

Conditional independence of the links and weights in Definition 2.1, together with their bounded support, gives by Bernstein concentration $\|\Phi - \mathbb{E}(\Phi)\| = O_p(\sqrt{\|\Theta\|_F^2 \log N})$ and $\max_i |d_i - \bar{d}_i| = O_p(\sqrt{\|\Theta\|_F^2 \log N})$. Since $\log N / \|\Theta\|_F^2 \rightarrow 0$, (A.8) implies that, with probability approaching one, $\min_i d_i \asymp \min_i \bar{d}_i \asymp \|\Theta\|_F^2$. The mean-value theorem therefore gives $\|D^{-1/2} - \bar{D}^{-1/2}\| = O_p(\sqrt{\log N} / \|\Theta\|_F^2)$, while $\|D^{-1/2}\| + \|\bar{D}^{-1/2}\| = O_p(\|\Theta\|_F^{-1})$.

Using the definitions of Φ^S and $\bar{\Phi}^S$, write

$$\begin{aligned}\Phi^S - \bar{\Phi}^S &= \frac{\rho}{2} D^{-1/2} [\Phi + \Phi' - \mathbb{E}(\Phi) - \mathbb{E}(\Phi)'] D^{-1/2} \\ &\quad + \frac{\rho}{2} (D^{-1/2} - \bar{D}^{-1/2}) [\mathbb{E}(\Phi) + \mathbb{E}(\Phi)'] D^{-1/2} \\ &\quad + \frac{\rho}{2} \bar{D}^{-1/2} [\mathbb{E}(\Phi) + \mathbb{E}(\Phi)'] (D^{-1/2} - \bar{D}^{-1/2}).\end{aligned}\tag{A.9}$$

Moreover, $\|\mathbb{E}(\Phi) + \mathbb{E}(\Phi)'\| = O(\|\Theta\|_F^2)$ by (A.8). Substituting the preceding bounds into (A.9) shows that all three terms have order $O_p(\sqrt{\log N / \|\Theta\|_F^2})$. Hence

$$\|\Phi^S - \bar{\Phi}^S\| = O_p\left(\sqrt{\frac{\log N}{\|\Theta\|_F^2}}\right).\tag{A.10}$$

It remains to control the time-series estimation error. Conditional on the realized network, the process satisfies (2.11). The degree normalization gives $\|\Phi\| \leq \rho < 1$, while Assumption 2.3 supplies the mixing and tail conditions required for the OLS concentration argument. Applied to the demeaned regression, the usual score and sample-covariance bounds give $\|\hat{\Phi} - \Phi\| = O_p(\sqrt{N/T})$. Symmetrization cannot increase the spectral norm, and hence $\|\hat{\Phi}^S - \Phi^S\| = O_p(\sqrt{N/T})$.

Combining these bounds with (A.3) proves (2.12).

PROOF. (PROOF OF LEMMA 2.3) By Assumption 2.2 and the degree bounds used in the proof of Lemma 2.2, $\Pi' \Theta \bar{D}^{-1} \Theta \Pi$ is uniformly well conditioned. Since $(B + B')/2$ is also uniformly positive definite, the K nonzero eigenvalues of $\bar{\Phi}^S$ are uniformly bounded away from zero and infinity.

The third part of Assumption 2.2 additionally separates the leading eigenvalue from the remaining signal eigenvalues. Hence the first eigenvector is individually identified, while the remaining $K - 1$ signal directions are identified as an eigenspace. Weyl's inequality and Lemma 2.2 imply that the same separation holds for $\hat{\Phi}^S$ with probability approaching one.

Applying the Davis–Kahan perturbation theorem to the leading eigenvector and to the remaining signal eigenspace, and choosing the sign of the estimated leading eigenvector to agree with the population one, there exists an orthonormal matrix $\mathcal{O}$ satisfying $\mathcal{O}e_1 = e_1$ such that $\|\hat{X} - X\mathcal{O}\| = O_p\left(\sqrt{\frac{\log N}{\|\Theta\|_F^2}} + \sqrt{\frac{N}{T}}\right)$. This is (2.13).

PROOF. (PROOF OF THEOREM 2.3) Let $r_N = \sqrt{\log N / \|\Theta\|_F^2} + \sqrt{N/T}$. By Lemma 2.3 and fixed K ,

$$\|\hat{X} - X\mathcal{O}\|_F^2 = O_p(r_N^2),\tag{A.11}$$

where $\mathcal{O} = \text{diag}(1, Q)$.

We first establish the row-wise refinement needed for the SCORE denominator. Let $E = \hat{\Phi}^S - \bar{\Phi}^S = E_N + E_T$, where $E_N = \Phi^S - \bar{\Phi}^S$ and $E_T = \hat{\Phi}^S - \Phi^S$. Under Assumption 2.2, $\max_i \|X_i\|_2 = O(N^{-1/2})$, and the nonzero eigenvalues of $\bar{\Phi}^S$ are uniformly separated from zero.

Consider first the network component. The degree expansion used in the proof of Lemma 2.2, combined with Bernstein concentration after projection onto the fixed K -dimensional population eigenspace, gives $\max_i \|(E_N X)_i\|_2 = O_p(\sqrt{\log N / (N \|\Theta\|_F^2)})$.

The factor $N^{-1/2}$ relative to the spectral-norm bound comes from $\max_i \|X_{i\cdot}\|_2 = O(N^{-1/2})$; the random degree terms obey the same order by (A.8).

For the time-series component, conditional on the realized network, each coordinate of $E_T X$ is a fixed-dimensional projection of the OLS score. Uniform stability, $\|\Phi\| \leq \rho < 1$, together with Gaussianity and Assumption 2.3, gives $\max_i \|(E_T X)_{i\cdot}\|_2 = O_p(\sqrt{\log N/T})$ by the corresponding maximal concentration bound. Hence

$$\max_{1 \leq i \leq N} \|(EX)_{i\cdot}\|_2 = O_p \left(\sqrt{\frac{\log N}{N\|\Theta\|_F^2}} + \sqrt{\frac{\log N}{T}} \right). \quad (\text{A.12})$$

We now apply the two-to-infinity perturbation expansion of Cape et al. (2019); see also the row-wise eigenvector analysis underlying Mixed-SCORE in Jin et al. (2024). Since $\bar{\Phi}^S$ has fixed rank K , its signal eigenvalues are bounded away from zero, and $\|E\| = O_p(r_N) = o_p(1)$, the leading term in the expansion is the projected perturbation EX . The remaining terms contain either an additional factor r_N or the within-eigenspace Procrustes error. Using Lemma 2.3 and $\max_i \|X_{i\cdot}\|_2 = O(N^{-1/2})$, their combined order is $O_p[r_N\{\sqrt{\log N/(N\|\Theta\|_F^2)} + \sqrt{\log N/T}\} + N^{-1/2}r_N^2]$, which is of smaller order than the right-hand side of (A.12) under Assumptions 2.2 and 2.3. Consequently,

$$\max_{1 \leq i \leq N} \|\hat{X}_{i\cdot} - (X\mathcal{O})_{i\cdot}\|_2 = O_p \left(\sqrt{\frac{\log N}{N\|\Theta\|_F^2}} + \sqrt{\frac{\log N}{T}} \right). \quad (\text{A.13})$$

We next verify that this rate is sufficient for the SCORE ratio. Let $M = \Pi'\Theta\bar{D}^{-1}\Theta\Pi$. If η_1 is as in Assumption 2.2, then the unnormalized leading eigenvector of $\bar{\Phi}^S$ is $\bar{D}^{-1/2}\Theta\Pi\eta_1$. Therefore,

$$X_{i1} = \frac{\theta_i \bar{d}_i^{-1/2} \pi_i' \eta_1}{(\eta_1' M \eta_1)^{1/2}}.$$

By (A.8), the balance condition on Θ , the conditioning of G , and the balance condition on η_1 ,

$$X_{i1} \asymp N^{-1/2}$$

uniformly in i . The same conditions imply $\max_i \|R_i\| = O(1)$.

Moreover, Assumptions 2.2 and 2.3 imply

$$\sqrt{N} \max_{1 \leq i \leq N} \|\hat{X}_{i\cdot} - (X\mathcal{O})_{i\cdot}\|_2 = o_p(1),$$

because $\log N/\|\Theta\|_F^2 \rightarrow 0$ and $N \log N/T \rightarrow 0$. Hence

$$\max_i |\hat{X}_{i1} - X_{i1}| = o_p(N^{-1/2}),$$

so all sample SCORE denominators remain uniformly of order $N^{-1/2}$ with probability approaching one. Since the population SCORE coordinates are uniformly bounded, $S = \log N$ is asymptotically nonbinding for both the population and sample ratios.

The ratio expansion therefore gives

$$\frac{1}{N} \sum_{i=1}^N \|\hat{R}_i - R_i Q\|^2 \leq C \|\hat{X} - X\mathcal{O}\|_F^2 = O_p(r_N^2). \quad (\text{A.14})$$

Assumption 2.4 then implies, after a permutation of the community labels,

$$\max_{1 \leq k \leq K} \|\widehat{v}_k - v_k Q\|^2 = O_p(r_N^2). \quad (\text{A.15})$$

Let $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_K)$ and $\widehat{\Lambda} = \text{diag}(\widehat{\lambda}_1, \dots, \widehat{\lambda}_K)$. The eigenvalue equations, Lemmas 2.2 and 2.3, imply

$$\|\Lambda \mathcal{O} - \mathcal{O} \widehat{\Lambda}\| = O_p(r_N). \quad (\text{A.16})$$

Consequently, $\|Q \widehat{\Lambda}_{-1} Q' - \Lambda_{-1}\| = O_p(r_N)$, where $\Lambda_{-1} = \text{diag}(\lambda_2, \dots, \lambda_K)$. Together with (A.15), this yields $\max_k |\widehat{b}_{1k} - b_{1k}|^2 = O_p(r_N^2)$.

Since the group-level matrix in $\bar{\Phi}^S$ has common diagonal $\rho(\underline{u} + \bar{u})/2$, the formula for b_{1k} differs from the unit-diagonal Mixed-SCORE normalization only by the common factor $\{\rho(\underline{u} + \bar{u})/2\}^{-1/2}$, which cancels in the final unit-sum normalization of each membership vector.

Under Assumption 2.2, the population simplex is nondegenerate and b_{1k} is uniformly bounded away from zero and infinity. The barycentric recovery, eigenvalue rescaling, and unit-sum normalization are therefore locally Lipschitz. Using (A.14)–(A.16),

$$\min_{P \in \mathcal{P}_K} \frac{1}{N} \|\widehat{\Pi} - \Pi P\|_F^2 = O_p(r_N^2) = O_p\left(\frac{\log N}{\|\Theta\|_F^2} + \frac{N}{T}\right), \quad (\text{A.17})$$

which proves (2.15).