Group Network Multivariate GARCH*

Jian Chen[1] and Ganggang Xu[2]

[1]University of Sussex Business School, Brighton, BN1 9SL, United Kingdom
[2]Department of Management Science, University of Miami, Coral Gables, FL 33146, USA

January 20, 2024

Abstract

Traditional multivariate generalised autoregressive conditional heteroskedasticity (GARCH) models (e.g., BEKK, DCC model) often suffer from the curse of dimensionality. A group network multivariate GARCH model is proposed in which the transitions of past variance and return shocks among assets are subject to an adjacency matrix and a latent group structure. This approach significantly reduces the number of parameters in high dimensions, thus facilitating estimation and forecasting. The theoretical properties of an estimator are developed that uses an optimisation algorithm estimating parameters and group memberships simultaneously. Simulation results confirm our theoretical findings. An empirical analysis is conducted on the S&P 100 constituents from 2015 to 2022 and is shown that the model improves portfolio selection in out-of-sample forecasts compared to other models.

Keywords: Multivariate GARCH, group structure, financial networks, portfolio optimisation
JEL Classification codes: C53, C55, C58

*Jian Chen and Ganggang Xu are joint first authors, and Ganggang Xu is the corresponding author. Email: gangxu@bus.miami.edu

1 Introduction

Multivariate GARCH models, which are derived from the ARCH/GARCH family introduced by Engle (1982), have been widely used in risk management and asset allocation. The estimation of these models has also attracted considerable attention in the literature. Among them, two of the most popular models are the Baba, Engle, Kraft, and Kroner (BEKK) class model (Engle and Kroner, 1995) and the constant conditional correlation (CCC) model (Bollerslev, 1990), along with its extension, the dynamic conditional correlation (DCC) model, by Engle (2002). For detailed reviews on multivariate GARCH models, refer to the works of Bauwens et al. (2006), Silvennoinen and Teräsvirta (2009), Engle (2009) and Francq and Zakoian (2019).

The BEKK model is known to be over-parameterised with $N(5N + 1)/2$ parameters to be estimated simultaneously when applying a BEKK(1,1,1) model to N assets. This results in the curse of dimensionality with a heavy computational burden imposed on the estimation, which makes it seldom used when the number of assets exceeds 20. In contrast, the DCC model offers a more manageable two-stage estimation process: first fitting a univariate GARCH and then estimating the correlation matrix. This approach makes the DCC model more favourable in empirical analyses in higher-dimensional cases. However, it could also be substantially biased with larger datasets. See some relevant discussions by Scherrer and Ribarits (2007) and Engle et al. (2019).

In this paper, we introduce a group network BEKK model, which incorporates a latent group structure and a network structure into the coefficient matrices of a standard BEKK(1,1,1) framework. The entries of these coefficient matrices have the same values within and across asset groups. Moreover, we consider network effects by assigning zeros to coefficient matrix entries whose corresponding asset pairs are not connected. The model we propose significantly reduces the number of parameters and is computationally fast, which opens the door to estimating a BEKK model with high-dimensional datasets. Furthermore, the latent group and network structure enable the capture of the dynamics of return and variance transitions among assets. Moreover, it is a more generalised model that reduces to a basic BEKK(1,1,1) when setting the number of groups equal to the number of assets N and the adjacency matrix as an identity matrix of $N \times N$.

Our studies are relevant to a range of papers that have aimed to address the dimensionality issue of multivariate GARCH models. For example, Engle and Kelly (2012) propose a block equicorrelation structure that imposes equicorrelation within and between industries. Ledoit and Wolf (2012) propose a non-linear shrinkage method to improve the estimation of the correlation matrix in a DCC model. Pakel et al. (2021) examine a composite likelihood method that facilitates the estimation. Additionally, Engle and Mezrich (1996) propose a computationally less demanding variance-targeting approach, which is further studied by Francq et al. (2011) and Pedersen and Rahbek (2014). Our paper introduces a novel variant of the BEKK(1,1,1) model that retains the generality of the model and can be easily estimated.

The rest of the paper is organised as follows. Section 2 introduces our group network multivariate GARCH model. Section 3 describes an estimation algorithm. Section 4 presents the theoretical properties of the estimator. Section 5 presents our simulation studies. Section 6 introduces an empirical study on portfolio selection using S&P 100 constituents. Section 7 concludes the paper.

Notations. In the subsequent text, the subscript i is used to index the assets that range from 1 to N. N represents the total number of assets and the dimension of the covariance matrix. Similarly, the subscript t is used to index the dates that range from 1 to T, with T representing the sample size. Define $[G] = \{1, ..., G\}$ and $[G]^n = \{(g_1, ..., g_n)' : g_i \in [G]\}$. For an arbitrary matrix $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{m\times n}$, denote $\mathbf{A}_1 \circ \mathbf{A}_2 = (a_{1,ij}a_{2,ij}) \in \mathbb{R}^{m\times n}$ as the Hadamard product of two matrices. For an arbitrary vector $\mathbf{v} = (v_1, ..., v_n)^\top \in \mathbb{R}^n$, denote the L_2-norm as $\|\mathbf{v}\| = (\sum_{i=1}^n v_i^2)^{1/2}$.

2 A Group Network BEKK model

Denote the asset log-return process as $Y_{it} = log(P_{it}/P_{i(t-1)})$ where P_{it} represents the asset price and the vector form of returns $\mathbf{y_t} = (Y_{1t}, ..., Y_{Nt})'$ follows:

$$\mathbf{y_t} = \boldsymbol{\mu_t} + \boldsymbol{\epsilon_t} \tag{2.1}$$

where $\mu_t = \mathbb{E}(\mathbf{y_t}|\mathcal{F}_{t-1})$ and $\mathcal{F}_{t-1} = \boldsymbol{\sigma}(\mathbf{y_{t-1}}, \mathbf{y_{t-2}}, ...)$ is the sigma field generated by the past information. The noise process ϵ_t follows:

$$\boldsymbol{\epsilon_t} = H_t^{1/2} z_t \tag{2.2}$$

where H_t is some $\mathcal{F}_{t-1}$-measurable $(N \times N)$-dimensional symmetric matrix and z_t is an N-dimensional vector process that is independently and identically distributed with zero mean and unit variance, $z_t \sim iid(0, \mathbb{1}_N)$. Therefore, the variance-covariance matrix of $\mathbf{y_t}$ conditional on $\mathcal{F}_{t-1}$ equals H_t. Then, the BEKK(1,1,1) takes the following form:

$$H_t = CC' + A'\boldsymbol{\epsilon_{t-1}}\boldsymbol{\epsilon'_{t-1}}A + B'H_{t-1}B \tag{2.3}$$

where C, A and B are $N \times N$ matrices but C is upper triangular. This form is the simplest without additional lags of $\boldsymbol{\epsilon_t}$ and H_t. However, it is still over-parameterised and not applicable to larger datasets. One could impose a diagonal BEKK model where A and B are diagonal matrices, or a scalar BEKK model where A and B are two scalar times two matrices of ones. However, this may consequently reduce the generality of the model.

2.1 Latent Group Structure

In this paper, we impose a latent group structure on C, A and B. For a given number of group G, denote the group membership vector as $\mathbb{G} = (g_1, ..., g_N)^\top \in [G]^N$. Assuming N assets belong to G groups, we replace parameter matrices $\{C, A, B\}$ by $\{C_\mathbb{G}, A_\mathbb{G}, B_\mathbb{G}\}$ of the same size, which makes the model follows:

$$H_t = C_\mathbb{G} C'_\mathbb{G} + A'_\mathbb{G} \boldsymbol{\epsilon_{t-1}} \boldsymbol{\epsilon'_{t-1}} A_\mathbb{G} + B'_\mathbb{G} H_{t-1} B_\mathbb{G} \tag{2.4}$$

where $C_\mathbb{G}, A_\mathbb{G}$ and $B_\mathbb{G}$ are three $N \times N$ square matrices whose entries are labeled by group memberships:

$$C_\mathbb{G} = \begin{bmatrix} c_{(g_1,g_1)} & c_{(g_1,g_2)} & \cdots & c_{(g_1,g_N)} \\ 0 & c_{(g_2,g_2)} & \cdots & c_{(g_1,g_N)} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & c_{(g_N,g_N)} \end{bmatrix} + \begin{bmatrix} \tilde{c}_{(g_1)} - c_{(g_1,g_1)} & 0 & \cdots & 0 \\ 0 & \tilde{c}_{(g_2)} - c_{(g_2,g_2)} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{c}_{(g_N)} - c_{(g_N,g_N)} \end{bmatrix}$$

$$A_\mathbb{G} = \begin{bmatrix} a_{(g_1,g_1)} & a_{(g_1,g_2)} & \cdots & a_{(g_1,g_N)} \\ a_{(g_2,g_1)} & a_{(g_2,g_2)} & \cdots & a_{(g_2,g_N)} \\ \vdots & & \ddots & \vdots \\ a_{(g_N,g_1)} & a_{(g_N,g_2)} & \cdots & a_{(g_N,g_N)} \end{bmatrix} + \begin{bmatrix} \tilde{a}_{(g_1)} - a_{(g_1,g_1)} & 0 & \cdots & 0 \\ 0 & \tilde{a}_{(g_2)} - a_{(g_2,g_2)} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{a}_{(g_N)} - a_{(g_N,g_N)} \end{bmatrix}$$

$$B_{\mathbb{G}} = \begin{bmatrix} b_{(g_1,g_1)} & b_{(g_1,g_2)} & \cdots & b_{(g_1,g_N)} \\ b_{(g_2,g_1)} & b_{(g_2,g_2)} & \cdots & b_{(g_2,g_N)} \\ \vdots & & \ddots & \vdots \\ b_{(g_N,g_1)} & b_{(g_N,g_2)} & \cdots & b_{(g_N,g_N)} \end{bmatrix} + \begin{bmatrix} \tilde{b}_{(g_1)} - b_{(g_1,g_1)} & 0 & \cdots & 0 \\ 0 & \tilde{b}_{(g_2)} - b_{(g_2,g_2)} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{b}_{(g_N)} - b_{(g_N,g_N)} \end{bmatrix}$$

where $\tilde{c}, \tilde{a}$ and $\tilde{b}$ are compensators for diagonal elements. With a suitable permutation, $C_{\mathbb{G}}, A_{\mathbb{G}}$ and $B_{\mathbb{G}}$ can be expressed as the following block structures:

$$C_{\mathbb{G}} = \begin{bmatrix} c_{(1,1)}\mathbf{I}^*_{\{1,1\}} & c_{(1,2)}\mathbf{I}^*_{\{1,2\}} & \cdots & c_{(1,G)}\mathbf{I}^*_{\{1,G\}} \\ 0 & c_{(2,2)}\mathbf{I}^*_{\{2,2\}} & & c_{(2,G)}\mathbf{I}^*_{\{2,G\}} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & c_{(G,G)}\mathbf{I}^*_{\{G,G\}} \end{bmatrix} + \begin{bmatrix} (\tilde{c}_{(1)} - c_{(1,1)})\mathbf{I}^d_{\{1,1\}} & 0 & \cdots & 0 \\ 0 & (\tilde{c}_{(2)} - c_{(2,2)})\mathbf{I}^d_{\{2,2\}} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & (\tilde{c}_{(G)} - c_{(G,G)})\mathbf{I}^d_{\{G,G\}} \end{bmatrix}$$

$$A_{\mathbb{G}} = \begin{bmatrix} a_{(1,1)}\mathbf{I}_{\{1,1\}} & a_{(1,2)}\mathbf{I}_{\{1,2\}} & \cdots & a_{(1,G)}\mathbf{I}_{\{1,G\}} \\ a_{(2,1)}\mathbf{I}_{\{2,1\}} & a_{(2,2)}\mathbf{I}_{\{2,2\}} & & a_{(2,G)}\mathbf{I}_{\{2,G\}} \\ \vdots & & \ddots & \vdots \\ a_{(G,1)}\mathbf{I}_{\{G,1\}} & a_{(G,2)}\mathbf{I}_{\{G,2\}} & \cdots & a_{(G,G)}\mathbf{I}_{\{G,G\}} \end{bmatrix} + \begin{bmatrix} (\tilde{a}_{(1)} - a_{(1,1)})\mathbf{I}^d_{\{1,1\}} & 0 & \cdots & 0 \\ 0 & (\tilde{a}_{(2)} - a_{(2,2)})\mathbf{I}^d_{\{2,2\}} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & (\tilde{a}_{(G)} - a_{(G,G)})\mathbf{I}^d_{\{G,G\}} \end{bmatrix}$$

$$B_{\mathbb{G}} = \begin{bmatrix} b_{(1,1)}\mathbf{I}_{\{1,1\}} & b_{(1,2)}\mathbf{I}_{\{1,2\}} & \cdots & b_{(1,G)}\mathbf{I}_{\{1,G\}} \\ b_{(2,1)}\mathbf{I}_{\{2,1\}} & b_{(2,2)}\mathbf{I}_{\{2,2\}} & & b_{(2,G)}\mathbf{I}_{\{2,G\}} \\ \vdots & & \ddots & \vdots \\ b_{(G,1)}\mathbf{I}_{\{G,1\}} & b_{(G,2)}\mathbf{I}_{\{G,2\}} & \cdots & b_{(G,G)}\mathbf{I}_{\{G,G\}} \end{bmatrix} + \begin{bmatrix} (\tilde{b}_{(1)} - b_{(1,1)})\mathbf{I}^d_{\{1,1\}} & 0 & \cdots & 0 \\ 0 & (\tilde{b}_{(2)} - b_{(2,2)})\mathbf{I}^d_{\{2,2\}} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & (\tilde{b}_{(G)} - b_{(G,G)})\mathbf{I}^d_{\{G,G\}} \end{bmatrix}$$

where $c_{(i,j)}$ of $C_{\mathbb{G}}$ represents the parameter value at the block (i, j), and $\tilde{c}_{(i)}$ represents the parameter value of $\tilde{c}$ for group i. Same for $A_{\mathbb{G}}$ and $B_{\mathbb{G}}$. $\mathbf{I}_{\{i,j\}}$ denotes the $\tilde{i} \times \tilde{j}$-dimensional identity matrix, and $\tilde{i}$ is the number of assets in the group i. $\mathbf{I}^*_{\{i,j\}}$ denotes the upper triangular elements of $\mathbf{I}_{\{i,j\}}$. $\mathbf{I}^d_{\{i,j\}}$ denotes the $\tilde{i} \times \tilde{j}$-dimensional diagonal identity matrix. For a specific example of 3 groups for 8 assets, the matrix $A_{\mathbb{G}}$ can have the following format:

$$A_{\mathbb{G}} = \begin{bmatrix} 0.7 & 0.08 & 0.06 & 0.06 & 0.06 & -0.11 & -0.11 & -0.11 \\ -0.03 & 0.7 & 0.06 & 0.06 & 0.06 & -0.11 & -0.11 & -0.11 \\ 0.01 & 0.01 & 0.6 & -0.1 & -0.1 & 0.09 & 0.09 & 0.09 \\ 0.01 & 0.01 & 0.04 & 0.6 & -0.1 & 0.09 & 0.09 & 0.09 \\ 0.01 & 0.01 & 0.04 & 0.04 & 0.6 & 0.09 & 0.09 & 0.09 \\ 0.13 & 0.13 & -0.07 & -0.07 & -0.07 & -0.8 & -0.03 & -0.03 \\ 0.13 & 0.13 & -0.07 & -0.07 & -0.07 & 0.06 & -0.8 & -0.03 \\ 0.13 & 0.13 & -0.07 & -0.07 & -0.07 & 0.06 & 0.06 & -0.8 \end{bmatrix}$$

In this manner, we constrain the transition from past return shocks and covariance to the future covariance matrix, shifting from pair-wise transition to group-wise. This approach reduces the number of parameters being estimated from $N(5N+1)/2$ to $5G(G+1)/2$ and eases off the computational burden largely.

2.2 with Network Effects

In addition, we further impose a network structure on $A_{\mathbb{G}}$ and $B_{\mathbb{G}}$. For a given network with N nodes, where the relationships among these nodes are systematically recorded through an adjacency matrix, denoted as $\mathbf{M} = (m_{ij}) \in \{0,1\}^{N\times N}$. $m_{ij} = 1$ if the ith node and the jth node are connected and 0 otherwise. By convention, we set $m_{ii} = 1$ for all $i = 1, ..., N$. The group network BEKK model can be expressed as:

$$H_t = C_{\mathbb{G}}C'_{\mathbb{G}} + A^{*'}_{\mathbb{G}}\boldsymbol{\epsilon_{t-1}}\boldsymbol{\epsilon'_{t-1}}A^{*}_{\mathbb{G}} \circ K + B^{*'}_{\mathbb{G}}H_{t-1}B^{*}_{\mathbb{G}} \circ K \tag{2.5}$$

where $A^*_{\mathbb{G}} = A_{\mathbb{G}} \circ M^*$, $B^*_{\mathbb{G}} = B_{\mathbb{G}} \circ M^*$ and $M^* = M + \mathbf{I}^d_{N\times N}$. K is the element-wise reciprocal of an $N \times N$ matrix K^* and $K^* = M'\mathbf{I}_{N\times N}M$. The reason for introducing K is to average the network effect since $A^*_{\mathbb{G}}$ and $B^*_{\mathbb{G}}$ become sparse matrices after multiplying the adjacency matrix. This can potentially lead to some entries of H_t becoming very large or very small, especially when rows of $A^*_{\mathbb{G}}$ and $B^*_{\mathbb{G}}$ are sparse.

The model we proposed offers greater flexibility in capturing the dynamics of the covariance matrix, while also maintaining its generality. With a proper setting of the number of groups G and the adjacency matrix $\mathbf{M}$, one can model a high-dimensional H_t with a relatively small number of parameters. This model reduces to a basic BEKK(1,1,1) configuration when setting $G = N$ and $\mathbf{M} = \mathbf{I}_{N\times N}$. The constraint of the model is that an available network relationship, i.e. adjacency matrix $\mathbf{M}$, is required.

3 Model Estimation

For a given number of groups G, we stack non-zero elements in the $C^*_{\mathbb{G}}$, $A^*_{\mathbb{G}}$ and $B^*_{\mathbb{G}}$ structure to a parameter vector $\boldsymbol{\theta} \in \mathbb{R}^{5G(G+1)/2}$:

$$\boldsymbol{\theta} = vec\left(\bigcup_{i=1}^{G}\bigcup_{j=i}^{G}\{c_{(i,j)}\} \;\cup\; \bigcup_{i=1}^{G}\bigcup_{j=1}^{G}\{a_{(i,j)}, b_{(i,j)}\} \;\cup\; \bigcup_{i=1}^{G}\{\tilde{c}_{(i)}, \tilde{a}_{(i)}, \tilde{b}_{(i)}\}\right) \tag{3.1}$$

Define the true number of groups as G_0 and true parameter vector as $\boldsymbol{\theta}^0 \in \mathbb{R}^{5G_0(G_0+1)/2}$. The parameter vector $\boldsymbol{\theta}$ and the membership vector $\mathbb{G}$ can be estimated by maximising the following log-likelihood function:

$$L(\boldsymbol{\theta}, \mathbb{G}) = -\frac{1}{2}\sum_{t=1}^{T} log|(H_t|\boldsymbol{\theta}, \mathbb{G}, \mathbf{M})| - \frac{1}{2}\sum_{t=1}^{T}(\mathbf{y_t} - \boldsymbol{\mu_t})^\top (H_t|\boldsymbol{\theta}, \mathbb{G}, \mathbf{M})^{-1}(\mathbf{y_t} - \boldsymbol{\mu_t}) \tag{3.2}$$

which is commonly adopted in MGARCH literature. We realised that the inversion of the conditional covariance matrix H_t in the Gaussian (quasi-) likelihood estimation poses significant computational challenges, with the difficulty increasing rapidly as N grows.

There is a stream of literature dedicated to exploring computationally less demanding approximation methods for the likelihood function (see e.g., Hafner and Reznikova, 2012; Pakel et al., 2021). While these methods could be easily applied in our context, they are not the main focus of this paper. We opt for the most commonly used log-likelihood function.

3.1 An Optimisation Algorithm

In this section, we introduce an iterative algorithm to maximise the log-likelihood function in Eq.(3.2) with respect to $\boldsymbol{\theta}$ and $\mathbb{G}$. For a given G, it contains the following steps:

(a) The initial value of group memberships are uniformly sampled from $[G]$, i.e., $\widehat{\mathbb{G}}^{(0)} = (\widehat{g}_1^{(0)}, ..., \widehat{g}_N^{(0)})^\top$, $\widehat{g}_i^{(0)}$ is uniformly distributed over set $[G]$ with $P(\widehat{g}_i^{(0)} = j) = \frac{1}{G}$ for $j = 1, 2, ..., G$. The initial parameters $\widehat{\boldsymbol{\theta}}^{(0)}$ can be found by maximising the log-likelihood function:

$$\widehat{\boldsymbol{\theta}}^{(0)} = \underset{\boldsymbol{\theta} \in \mathbb{R}^{5G(G+1)/2}}{\arg\max} \; L(\boldsymbol{\theta}, \widehat{\mathbb{G}}^{(0)}) \tag{3.3}$$

(b) **Update group memberships**: in the $(k+1)$th iteration, sequentially update entries of the group membership estimator in the kth step $\widehat{\mathbb{G}}^{(k)}$. The group membership of node i is updated by:

$$\widehat{g}_i^{(k+1)} = \underset{g \in [G]}{\arg\max} \; L(\widehat{\boldsymbol{\theta}}^{(k)}, \widehat{\mathbb{G}}_{-i}^{(k)}(g)) \tag{3.4}$$

where $\widehat{\mathbb{G}}_{-i}^{(k)}(g) = \left(\widehat{g}_1^{(k+1)}, ..., \widehat{g}_{i-1}^{(k+1)}, g, \widehat{g}_{i+1}^{(k+1)}, ..., \widehat{g}_N^{(k+1)}\right)^\top$, for $i = 1, ..., N$.

(c) **Update the parameter estimates**: given the group membership $\widehat{\mathbb{G}}^{(k+1)}$, update the parameter estimates $\widehat{\boldsymbol{\theta}}^{(k+1)}$:

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \underset{\boldsymbol{\theta} \in \mathbb{R}^{5G(G+1)/2}}{\arg\max} \; L(\boldsymbol{\theta}, \widehat{\mathbb{G}}^{(k+1)}) \tag{3.5}$$

(d) Repeat (b)-(c) until the convergence criterion is met.

This optimization algorithm is of the k-means type. Although this algorithm framework is nowhere found in multivariate GARCH-related literature, it is widely adopted in recent studies on group panel data models (see e.g., Zhang et al., 2019; Liu et al., 2020; Zhu et al., 2023).

3.2 Consistent Selection of G_0

...

With a slight abuse of notation, denote $\widehat{\boldsymbol{\theta}}^{(G)}$ as the parameter set estimated with G groups, we consider the following likelihood-based criterion:

$$\text{LIC}(G) = -L(\widehat{\boldsymbol{\theta}}^{(G)}, G) + \lambda(N, T, G) \tag{3.6}$$

where $L(\widehat{\boldsymbol{\theta}}^{(G)}, G)$ is the log-likelihood function defined in Eq.(3.2) and $\lambda(N, T, G)$ is a penalty function.

4 Simulation Studies

We conduct a number of simulation studies with different parameter settings to illustrate the finite sample properties of the estimator. We use the group network BEKK model, as outlined in Eq.(2.5), to model the covariance dynamics in the data generating process. $H_t^{1/2}$ of Eq.(2.2) is in its Cholesky form and z_t is a standard normal vector. Additionally, we set $\boldsymbol{\mu_t} \equiv 0$ in Eq.(2.1) for simplicity. We experiment with different numbers of groups $G_0 = \{3, 4, 5, 6, 7, 10\}$, and independently sample the true group memberships of each asset uniformly from $[G_0]$, with each group having a probability of $\frac{1}{G_0}$. For the network structure, we consider a stochastic block model. If two assets are in the same group, we let them be connected with a probability $P(m_{ij} = 1) = N^{-0.3}$ or with a probability $P(m_{ij} = 1) = N^{-0.6}$ otherwise. We experiment with varying $N = \{30, 50, 100\}$, $T = \{126, 252, 504, 1260\}$, $G = \{3, 5, 10\}$ and two sets of true parameters shown in the below table:

Table 1: Parameter settings of simulation studies

	$\tilde{c}_{i,i}$	$c_{i,j}$	$\tilde{a}_{i,i}$	$a_{i,j}$	$\tilde{b}_{i,i}$	$b_{i,j}$
Scenario 1	[0, 0.3]	[-0.3, 0]	[0.1, 0.3]	[0.1, 0.3]	[0.4, 0.6]	[0.2, 0.4]
Scenario 2	[0, 0.3]	[-0.2, 0.2]	[0.1, 0.4]	[-2, 2]	[0.1, 0.5]	[-4, 4]

Notes: $\tilde{c}_{i,i}$ and $c_{i,j}$ denotes the diagonal and off-diagonal elements of the matrix $C^*_{\mathbb{G}}$ in Eq. 2.5, which are equally-spaced sampled from the intervals listed in the table. For a specific example, when $G = 4$, $\tilde{b}_{i,i}$ is sampled at $\{0.4, 0.467, 0.533, 0.6\}$ in Scenario 1. The same applies to $A^*_{\mathbb{G}}$ and $B^*_{\mathbb{G}}$.

4.1 Estimation Accuracy when $G = G_0$

We first evaluate the parameter estimation accuracy when setting $G = G_0$ by the root mean squared error (RMSE) of the parameter estimates after some suitable permutation on group memberships. We access three coefficient matrices $\{C_{\mathbb{G}}, A^*_{\mathbb{G}}, B^*_{\mathbb{G}}\}$ individually. For kth simulation, denote the coefficient matrix estimates as $\widehat{C}^{(k)}_{\mathbb{G}}$, $\widehat{A}^{*(k)}_{\mathbb{G}}$ and $\widehat{B}^{*(k)}_{\mathbb{G}}$. The RMSE of $C_{\mathbb{G}}$, for example, is defined as $\text{RMSE}(C_{\mathbb{G}}) = K^{-1}\sum_{k=1}^{K}\|\widehat{C}^{(k)}_{\mathbb{G}} - C^0_{\mathbb{G}}\|$. The group membership estimation accuracy rate (GAR) is defined as $\text{GAR} = K^{-1}\sum_{k=1}^{K}\{N^{-1}\sum_{i=1}^{N} I(\widehat{g}_i^{(k)} = g_i^0)\}$. As a comparison, we also obtain the estimates of an "Oracle" estimator for which true group memberships of all assets are known and fixed when finding the maximum log-likelihood. The simulation results based on $K = 500$ simulations are presented in Table 2.

As shown in the table, the parameter estimation accuracy substantially improves and approaches that of the "Oracle" estimator as T increases. Additionally, the GAR gradually approaches 100% as T increases. However, no clear trends are observed in the parameter estimation accuracy when N increases. Moreover, we find that the performance is better with a smaller number of groups G. Regarding the two scenarios of parameter settings, there are no clear differences between their performances. Overall, these simulation results support the estimation consistency of our model given a correctly specified G.

Table 2: RMSE and GAR of Simulation Parameter Estimates

		$N=30$								$N=50$								$N=100$							
		Scenario 1				Scenario 2				Scenario 1				Scenario 2				Scenario 1				Scenario 2			
T	G	$C_{\mathbb{G}}$	$A^*_{\mathbb{G}}$	$B^*_{\mathbb{G}}$	GAR	$C_{\mathbb{G}}$	$A^*_{\mathbb{G}}$	$B^*_{\mathbb{G}}$	GAR	$C_{\mathbb{G}}$	$A^*_{\mathbb{G}}$	$B^*_{\mathbb{G}}$	GAR	$C_{\mathbb{G}}$	$A^*_{\mathbb{G}}$	$B^*_{\mathbb{G}}$	GAR	$C_{\mathbb{G}}$	$A^*_{\mathbb{G}}$	$B^*_{\mathbb{G}}$	GAR	$C_{\mathbb{G}}$	$A^*_{\mathbb{G}}$	$B^*_{\mathbb{G}}$	GAR
252*0.5	3	0.074	0.151	0.151	86.5%	0.071	0.123	0.128	85.3%	0.073	0.12	0.146	87.3%	0.074	0.133	0.121	85.3%	0.067	0.118	0.133	89.5%	0.06	0.117	0.135	90.3%
		(0.03)	(0.088)	(0.123)		(0.032)	(0.084)	(0.113)		(0.032)	(0.074)	(0.11)		(0.034)	(0.071)	(0.107)		(0.032)	(0.064)	(0.103)		(0.028)	(0.06)	(0.101)	
	5	0.063	0.152	0.145	85.7%	0.062	0.147	0.155	84.9%	0.057	0.135	0.143	84.8%	0.058	0.136	0.139	82.0%	0.066	0.126	0.136	80.8%	0.066	0.129	0.137	84.7%
		(0.033)	(0.083)	(0.11)		(0.029)	(0.082)	(0.112)		(0.023)	(0.076)	(0.11)		(0.024)	(0.074)	(0.105)		(0.023)	(0.064)	(0.108)		(0.024)	(0.07)	(0.112)	
	10	0.06	0.175	0.165	79.2%	0.06	0.196	0.163	78.5%	0.051	0.159	0.167	84.9%	0.054	0.159	0.162	83.4%	0.065	0.156	0.17	80.6%	0.064	0.151	0.167	79.7%
		(0.029)	(0.1)	(0.128)		(0.027)	(0.102)	(0.12)		(0.04)	(0.099)	(0.141)		(0.031)	(0.091)	(0.132)		(0.026)	(0.094)	(0.129)		(0.027)	(0.084)	(0.135)	
252	3	0.051	0.083	0.11	93.1%	0.041	0.086	0.105	93.8%	0.055	0.072	0.092	95.2%	0.047	0.077	0.089	93.6%	0.047	0.083	0.099	95.1%	0.052	0.085	0.097	93.2%
		(0.027)	(0.049)	(0.098)		(0.027)	(0.059)	(0.097)		(0.031)	(0.062)	(0.092)		(0.031)	(0.064)	(0.091)		(0.031)	(0.051)	(0.091)		(0.031)	(0.045)	(0.092)	
	5	0.051	0.098	0.1	88.8%	0.051	0.099	0.097	90.2%	0.045	0.073	0.1	91.9%	0.042	0.081	0.096	91.7%	0.055	0.089	0.124	91.7%	0.044	0.094	0.125	92.3%
		(0.035)	(0.077)	(0.074)		(0.037)	(0.08)	(0.069)		(0.022)	(0.053)	(0.066)		(0.02)	(0.054)	(0.066)		(0.019)	(0.061)	(0.088)		(0.019)	(0.059)	(0.087)	
	10	0.043	0.11	0.137	88.7%	0.043	0.107	0.138	89.8%	0.047	0.09	0.118	89.1%	0.047	0.09	0.113	88.7%	0.045	0.092	0.126	90.6%	0.045	0.095	0.128	89.5%
		(0.03)	(0.084)	(0.11)		(0.026)	(0.08)	(0.108)		(0.053)	(0.064)	(0.111)		(0.039)	(0.063)	(0.116)		(0.022)	(0.058)	(0.09)		(0.027)	(0.06)	(0.092)	
252*2	3	0.041	0.056	0.064	96.8%	0.036	0.053	0.062	97.2%	0.047	0.066	0.063	95.8%	0.043	0.056	0.063	97.3%	0.034	0.058	0.069	97.8%	0.035	0.056	0.067	97.7%
		(0.024)	(0.035)	(0.052)		(0.024)	(0.036)	(0.05)		(0.026)	(0.039)	(0.055)		(0.026)	(0.038)	(0.053)		(0.02)	(0.034)	(0.063)		(0.02)	(0.035)	(0.063)	
	5	0.033	0.056	0.072	96.0%	0.034	0.062	0.072	95.9%	0.032	0.064	0.073	95.7%	0.035	0.062	0.07	95.6%	0.035	0.057	0.063	96.7%	0.033	0.061	0.064	96.1%
		(0.025)	(0.042)	(0.063)		(0.023)	(0.043)	(0.061)		(0.019)	(0.041)	(0.064)		(0.017)	(0.039)	(0.061)		(0.023)	(0.043)	(0.049)		(0.023)	(0.041)	(0.049)	
	10	0.033	0.071	0.079	94.3%	0.033	0.074	0.079	94.2%	0.032	0.083	0.09	94.5%	0.033	0.079	0.086	93.9%	0.034	0.074	0.083	93.9%	0.034	0.072	0.083	94.7%
		(0.038)	(0.052)	(0.068)		(0.028)	(0.055)	(0.069)		(0.025)	(0.062)	(0.07)		(0.025)	(0.054)	(0.069)		(0.02)	(0.045)	(0.062)		(0.019)	(0.046)	(0.061)	
252*5	3	0.029	0.045	0.063	99.1%	0.033	0.045	0.067	99.1%	0.027	0.038	0.057	99.4%	0.026	0.039	0.057	99.5%	0.031	0.044	0.065	99.4%	0.028	0.043	0.065	99.5%
		(0.024)	(0.034)	(0.068)		(0.023)	(0.032)	(0.067)		(0.022)	(0.027)	(0.046)		(0.021)	(0.031)	(0.046)		(0.024)	(0.033)	(0.063)		(0.024)	(0.03)	(0.063)	
	5	0.028	0.058	0.069	98.6%	0.03	0.055	0.066	98.6%	0.028	0.041	0.053	99.1%	0.028	0.039	0.055	99.0%	0.028	0.055	0.064	98.8%	0.028	0.052	0.064	99.0%
		(0.023)	(0.042)	(0.065)		(0.023)	(0.043)	(0.064)		(0.016)	(0.035)	(0.039)		(0.017)	(0.034)	(0.038)		(0.019)	(0.037)	(0.045)		(0.019)	(0.036)	(0.044)	
	10	0.028	0.057	0.065	98.3%	0.028	0.056	0.07	98.5%	0.031	0.05	0.067	98.5%	0.033	0.05	0.066	98.4%	0.028	0.064	0.073	98.6%	0.026	0.062	0.073	98.6%
		(0.024)	(0.052)	(0.06)		(0.024)	(0.051)	(0.061)		(0.031)	(0.043)	(0.067)		(0.028)	(0.045)	(0.067)		(0.017)	(0.042)	(0.056)		(0.017)	(0.042)	(0.056)	

Notes: The table presents the RMSE of estimated parameters in simulations and the corresponding oracle estimator in the bracket. T denotes the length of time series, which are 0.5, 1, 2 and 5 years. G denotes the number of groups. GAR denotes the group accuracy rate.

4.2 Coverage Probability when $G = G_0$

We now evaluate the performance of statistical inference derived from the theorem. We construct 95% confidence intervals for model parameters. For example, the confidence interval of $c_{(i,j)}$ in $C_{\mathbb{G}}$ is defined as $\mathrm{CI}^{(k)}_{c_{(i,j)}} = (\widehat{c}^{(k)}_{(i,j)} - 1.96\widehat{\sigma}^{(k)}_{c_{(i,j)}}, \widehat{c}^{(k)}_{(i,j)} + 1.96\widehat{\sigma}^{(k)}_{c_{(i,j)}})$, where $\widehat{\sigma}^{(k)}_{c_{(i,j)}}$ denotes the estimated asymptotic standard error in the kth simulation. The same applies to entries in $A^*_{\mathbb{G}}$ and $B^*_{\mathbb{G}}$. We evaluate the coverage probability by an average error (AE). For example, the average error of coverage probability of parameters in $C_{\mathbb{G}}$ is computed as $\mathrm{AE}_{\mathrm{cp},C_{\mathbb{G}}} = ((1+G)G/2)^{-1} \sum_{i=1}^{G} \sum_{j=i}^{G} |K^{-1} \sum_{k=1}^{K} I(c^0_{(i,j)} \in \mathrm{CI}^{(k)}_{c_{(i,j)}}) - 0.95|$. The average errors for $A^*_{\mathbb{G}}$ and $B^*_{\mathbb{G}}$ are similarly defined. Additionally, we compute the average error of the oracle estimators as a comparison. The results are reported in Table 3.

While $A^*_{\mathbb{G}}$ and $B^*_{\mathbb{G}}$ show smaller empirical average errors than $C_{\mathbb{G}}$, they all diminish as T increases, which is in line with our theoretical findings. Additionally, similar to the estimation accuracy of parameters, there are no distinct differences between the performance of two settings of parameters.

4.3 Estimation Accuracy with Mis-specified Gs

We also evaluate the parameter estimation given a mis-specified G. We set the true number of groups at $G_0 = 5$ and estimate the model with a range of G. In the evaluation, we first recover $C_{\mathbb{G}}$, $A^*_{\mathbb{G}}$ and $B^*_{\mathbb{G}}$ using estimates with different G, then calculate estimation accuracy similarly as $\mathrm{RMSE}(C_{\mathbb{G}}) = K^{-1} \sum_{k=1}^{K} \|\widehat{C}^{(k)}_{\mathbb{G}} - C^0_{\mathbb{G}}\|$. The results are summarised in Table 4. We can see RMSE diminishes with GAR approaching 100% with T increases for $G \geq 5$. Therefore, our consistency results still hold when G is over-specified, i.e., $G > G_0$.

Table 3: Average error of coverage probability (%)

N	T	G	Scenario 1			Oracle (Scenario 1)			Scenario 2			Oracle (Scenario 2)		
			$C_{\mathbb{G}}$	$A^*_{\mathbb{G}}$	$B^*_{\mathbb{G}}$	$C_{\mathbb{G}}$	$A^*_{\mathbb{G}}$	$B^*_{\mathbb{G}}$	$C_{\mathbb{G}}$	$A^*_{\mathbb{G}}$	$B^*_{\mathbb{G}}$	$C_{\mathbb{G}}$	$A^*_{\mathbb{G}}$	$B^*_{\mathbb{G}}$
30	252*0.5	3	10.6	8.5	6.3	8.5	4.1	3.8	12.1	7.6	6.9	4.3	3.7	3.7
		5	10.0	7.7	6.0	7.1	2.6	5.5	11.4	7.2	5.8	6.7	2.1	5.5
		10	10.1	7.5	5.2	7.9	4.4	4.7	9.6	7.1	5.4	10.5	4.9	4.5
	252	3	7.7	5.8	4.0	7.6	2.8	2.5	6.7	5.2	4.0	7.4	2.1	2.6
		5	7.1	4.0	4.3	3.5	1.5	2.0	7.1	4.8	4.2	2.7	1.3	2.1
		10	6.7	3.9	3.1	5.9	1.5	2.3	6.7	3.9	3.1	5.7	1.6	2.3
	252*2	3	5.9	3.8	4.3	3.7	0.9	3.0	4.0	4.0	4.0	3.2	1.1	2.7
		5	4.8	3.6	2.4	3.0	1.5	1.5	4.8	3.9	2.5	3.0	1.4	1.6
		10	3.8	1.9	3.1	2.0	1.2	1.6	3.9	1.6	2.9	2.0	1.1	1.4
	252*5	3	4.0	2.0	1.3	4.0	1.3	0.8	3.3	1.9	1.2	3.9	1.4	0.8
		5	2.2	1.5	2.1	3.5	0.9	1.4	2.5	1.3	2.1	3.6	0.6	1.4
		10	2.7	2.4	1.6	3.7	0.8	1.1	2.7	2.9	1.6	3.9	1.1	1.2
50	252*0.5	3	12.6	7.2	6.1	6.1	3.2	3.8	11.5	7.8	6.2	6.3	4.1	3.7
		5	11.1	6.7	5.6	6.6	5.1	3.4	9.8	6.3	5.7	6.6	5.0	3.2
		10	11.0	6.1	5.9	7.4	3.9	4.5	11.4	5.9	5.7	10.0	4.2	4.6
	252	3	8.3	7.1	4.1	8.1	1.9	3.2	9.0	6.0	4.1	8.4	1.9	3.2
		5	6.7	4.6	3.1	5.4	3.8	1.4	6.7	5.0	3.0	5.7	4.2	1.3
		10	7.9	5.1	3.9	2.3	1.7	1.9	8.5	4.9	3.8	2.7	1.5	1.9
	252*2	3	6.3	3.1	2.3	3.8	1.6	1.3	6.9	3.4	2.2	4.1	1.6	1.2
		5	5.3	2.4	2.5	6.1	2.1	1.8	4.9	2.9	2.4	5.6	2.1	1.8
		10	6.0	1.8	1.4	3.8	1.2	0.7	5.4	1.9	1.2	3.4	0.9	0.6
	252*5	3	4.3	2.8	1.3	1.9	0.9	0.7	4.2	3.0	1.3	1.5	0.9	0.7
		5	2.4	1.6	2.7	3.8	1.7	2.4	2.6	1.8	2.7	4.0	1.5	2.4
		10	1.3	2.5	1.4	5.2	1.3	1.1	2.7	2.6	1.4	4.9	1.3	1.1
100	252*0.5	3	10.6	7.3	6.4	8.2	2.6	4.4	12.0	7.6	7.2	8.4	2.3	4.2
		5	10.9	9.1	6.4	7.9	4.6	4.4	11.6	9.0	6.6	8.4	4.6	4.4
		10	10.7	9.0	5.9	7.7	5.0	3.5	10.3	8.2	5.7	7.1	4.5	3.5
	252	3	8.4	6.0	4.3	8.9	3.0	4.2	8.2	5.8	5.3	9.2	2.3	4.2
		5	7.2	5.4	3.2	5.6	2.0	1.8	8.0	5.6	3.3	5.4	2.5	1.8
		10	8.3	4.3	4.2	4.0	2.4	2.0	8.6	4.1	4.2	3.9	2.5	2.0
	252*2	3	6.3	4.1	2.9	2.0	1.0	1.5	6.9	4.5	2.9	2.4	1.2	1.5
		5	6.6	2.8	3.0	5.0	2.2	2.0	4.6	3.1	3.1	5.0	2.1	2.0
		10	5.6	2.1	2.1	5.2	1.5	1.3	5.7	2.3	2.1	5.8	1.3	1.3
	252*5	3	4.1	3.0	1.4	3.6	0.7	0.8	3.1	2.5	1.4	3.8	0.6	0.7
		5	3.9	1.8	2.2	1.8	0.8	1.1	3.3	1.6	2.1	1.7	0.7	1.1
		10	2.3	2.3	1.3	5.0	2.0	1.1	5.5	2.2	1.3	5.3	2.2	1.2

Notes:

Table 4: RMSE and GAR of Simulation Parameter Estimates with Mis-specified G

		$N=30$								$N=50$								$N=100$							
		Scenario 1				Scenario 2				Scenario 1				Scenario 2				Scenario 1				Scenario 2			
T	G	C^*	A^*	G^*	GAR	C^*	A^*	G^*	GAR	C^*	A^*	G^*	GAR	C^*	A^*	G^*	GAR	C^*	A^*	G^*	GAR	C^*	A^*	G^*	GAR
	3	0.082	0.146	0.180	60.6%	0.087	0.138	0.211	60.3%	0.089	0.162	0.205	59.7%	0.080	0.180	0.214	59.1%	0.071	0.138	0.165	57.1%	0.094	0.140	0.178	59.6%
	4	0.061	0.139	0.147	77.0%	0.060	0.144	0.166	76.4%	0.056	0.155	0.164	76.6%	0.055	0.159	0.161	73.1%	0.048	0.117	0.149	77.2%	0.060	0.136	0.145	79.6%
	5	0.063	0.152	0.145	85.6%	0.062	0.147	0.155	84.8%	0.057	0.135	0.143	84.8%	0.058	0.136	0.139	82.0%	0.066	0.126	0.136	80.8%	0.066	0.129	0.137	84.6%
252*0.5	6	0.055	0.152	0.140	83.2%	0.059	0.158	0.172	80.0%	0.055	0.172	0.156	84.9%	0.047	0.164	0.158	80.9%	0.040	0.122	0.143	79.0%	0.049	0.142	0.163	80.7%
	7	0.058	0.152	0.153	85.5%	0.065	0.154	0.188	78.1%	0.063	0.167	0.140	82.2%	0.056	0.176	0.156	81.6%	0.047	0.129	0.135	78.5%	0.063	0.157	0.157	79.3%
	10	0.076	0.161	0.155	79.7%	0.083	0.157	0.192	73.8%	0.087	0.173	0.176	55.7%	0.072	0.175	0.171	59.4%	0.074	0.126	0.151	50.9%	0.103	0.140	0.155	49.7%
	oracle	0.033	0.083	0.110		0.029	0.082	0.112		0.023	0.076	0.110		0.024	0.074	0.105		0.023	0.064	0.108		0.024	0.070	0.112	
	3	0.093	0.076	0.166	56.3%	0.084	0.078	0.170	58.9%	0.084	0.083	0.174	57.6%	0.081	0.092	0.176	56.4%	0.126	0.079	0.188	56.6%	0.077	0.080	0.210	58.6%
	4	0.076	0.102	0.141	72.3%	0.068	0.096	0.144	75.9%	0.064	0.092	0.165	75.3%	0.060	0.101	0.164	77.5%	0.094	0.098	0.153	73.8%	0.055	0.090	0.169	79.1%
	5	0.051	0.098	0.100	88.7%	0.051	0.099	0.097	90.1%	0.045	0.073	0.100	91.9%	0.042	0.081	0.096	91.6%	0.055	0.089	0.124	91.7%	0.044	0.094	0.125	92.2%
252	6	0.057	0.107	0.110	85.3%	0.050	0.099	0.113	86.0%	0.049	0.078	0.115	84.7%	0.042	0.097	0.121	84.4%	0.074	0.082	0.115	83.7%	0.039	0.077	0.128	89.0%
	7	0.057	0.102	0.103	81.7%	0.051	0.096	0.103	89.1%	0.049	0.086	0.109	85.9%	0.046	0.096	0.119	83.8%	0.071	0.094	0.116	87.3%	0.045	0.090	0.129	87.7%
	10	0.070	0.114	0.109	74.6%	0.061	0.112	0.120	81.5%	0.066	0.083	0.114	75.3%	0.058	0.091	0.120	78.4%	0.134	0.120	0.140	57.2%	0.087	0.111	0.146	55.5%
	oracle	0.035	0.077	0.074		0.037	0.080	0.069		0.022	0.053	0.066		0.020	0.054	0.066		0.019	0.061	0.088		0.019	0.059	0.087	
	3	0.083	0.061	0.159	62.6%	0.100	0.072	0.155	61.9%	0.094	0.115	0.201	61.8%	0.086	0.105	0.175	65.1%	0.096	0.103	0.164	58.7%	0.093	0.108	0.177	57.7%
	4	0.056	0.057	0.124	77.5%	0.066	0.071	0.128	80.5%	0.054	0.102	0.154	85.2%	0.060	0.090	0.151	82.1%	0.066	0.094	0.129	78.7%	0.067	0.101	0.134	76.1%
	5	0.033	0.056	0.072	96.0%	0.034	0.062	0.072	95.9%	0.032	0.064	0.073	95.7%	0.035	0.062	0.070	95.6%	0.035	0.057	0.063	96.7%	0.033	0.061	0.064	96.1%
252*2	6	0.038	0.059	0.094	92.2%	0.041	0.068	0.089	92.4%	0.036	0.084	0.111	97.7%	0.041	0.080	0.116	97.0%	0.042	0.074	0.088	90.1%	0.040	0.082	0.101	90.1%
	7	0.038	0.070	0.093	94.5%	0.043	0.073	0.088	94.7%	0.047	0.099	0.118	94.2%	0.041	0.083	0.102	98.0%	0.039	0.073	0.090	93.1%	0.039	0.081	0.103	92.9%
	10	0.047	0.073	0.093	90.8%	0.056	0.084	0.092	89.9%	0.052	0.122	0.132	95.5%	0.043	0.096	0.115	97.0%	0.060	0.087	0.098	84.2%	0.060	0.089	0.102	84.4%
	oracle	0.025	0.042	0.063		0.023	0.043	0.061		0.019	0.041	0.064		0.017	0.039	0.061		0.023	0.043	0.049		0.023	0.041	0.049	
	3	0.116	0.164	0.234	58.9%	0.169	0.157	0.263	59.7%	0.161	0.125	0.243	58.1%	0.207	0.164	0.303	60.1%	0.134	0.198	0.230	57.5%	0.119	0.172	0.309	58.8%
	4	0.073	0.117	0.201	79.4%	0.106	0.118	0.232	80.0%	0.099	0.114	0.170	75.5%	0.128	0.148	0.211	77.2%	0.087	0.195	0.177	76.8%	0.078	0.172	0.248	75.7%
	5	0.028	0.058	0.069	98.5%	0.030	0.050	0.066	98.5%	0.028	0.041	0.053	99.0%	0.028	0.039	0.055	99.0%	0.028	0.055	0.064	98.7%	0.028	0.052	0.064	98.9%
252*5	6	0.032	0.072	0.083	98.8%	0.048	0.069	0.094	99.7%	0.039	0.046	0.056	98.5%	0.050	0.064	0.074	99.4%	0.043	0.089	0.077	99.4%	0.038	0.081	0.103	98.8%
	7	0.034	0.066	0.062	94.6%	0.053	0.068	0.073	95.2%	0.048	0.060	0.068	97.2%	0.065	0.083	0.088	97.7%	0.039	0.087	0.062	99.5%	0.032	0.074	0.087	98.8%
	10	0.037	0.079	0.076	99.9%	0.058	0.075	0.091	98.1%	0.051	0.076	0.074	99.1%	0.067	0.092	0.087	98.1%	0.047	0.117	0.072	94.4%	0.039	0.105	0.095	96.2%
	oracle	0.023	0.042	0.065		0.023	0.043	0.064		0.016	0.035	0.039		0.017	0.034	0.038		0.019	0.037	0.045		0.019	0.036	0.044	

Notes: The table presents the RMSE of estimated parameters in simulations with different G while $G_0 = 5$. T denotes the length of time series, which are 0.5, 1, 2 and 5 years. GAR denotes the group accuracy rate.

Furthermore, we select the number of groups using the LIC defined in Eq.(3.6) and set the penalty term as $\lambda(N, T, G) = 5log(NT)(N+P)/min(10, n_{0.9})$, where $P = 5G(G+1)/2$ denotes the number of parameters and $n_{0.9}$ is the 90% quantile of nodal in/out-degrees. We compute the selection accuracy rate (SR) as $\mathrm{SR}(G) = K^{-1}\sum_{k=1}^{K} I(\widehat{G}^{(k)} = G_0)$, where $\widehat{G}^{(k)}$ is the number of groups selected by LIC in kth simulation. We experiment with different $G_0 = \{3, 5, 7\}$ and results are presented in Table 5. As is shown in the table, the LIC can correctly select the number of groups (G) with SR gradually approaching 1 with T increases, which supports our theorem.

Table 5: Selection Accuracy Rate with Different G_0 (%)

		$N = 30$						$N = 50$						$N = 100$					
		Scenario 1			Scenario 2			Scenario 1			Scenario 2			Scenario 1			Scenario 2		
T	G	$G_0 = 3$	$G_0 = 5$	$G_0 = 7$	$G_0 = 3$	$G_0 = 5$	$G_0 = 7$	$G_0 = 3$	$G_0 = 5$	$G_0 = 7$	$G_0 = 3$	$G_0 = 5$	$G_0 = 7$	$G_0 = 3$	$G_0 = 5$	$G_0 = 7$	$G_0 = 3$	$G_0 = 5$	$G_0 = 7$
252*0.5	3	34	35	0	22	38	0	20	30	0	49	48	0	62	0	0	18	2	0
	4	10	40	10	16	28	12	34	43	0	46	33	0	24	88	0	22	92	0
	5	38	15	38	34	21	43	22	21	40	3	13	45	12	9	87	60	6	84
	6	14	6	29	26	5	19	24	3	44	2	6	43	2	3	13	0	0	12
	7	4	4	23	2	8	24	0	3	16	0	0	12	0	0	0	0	0	4
	10	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
252	3	82	0	0	40	0	0	48	0	0	58	0	0	78	0	0	75	0	0
	4	15	17	6	35	8	12	33	0	0	36	4	0	16	0	0	18	0	0
	5	3	64	18	16	73	33	13	94	18	4	86	42	5	100	17	4	96	12
	6	0	13	27	9	15	29	4	2	31	1	6	24	1	0	33	3	4	51
	7	0	6	49	0	4	26	2	4	51	1	4	34	0	0	50	0	0	37
	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
252*2	3	100	0	0	80	0	0	72	0	0	76	0	0	90	0	0	89	0	0
	4	0	4	0	17	0	0	11	8	0	6	7	0	5	0	0	4	0	0
	5	0	84	6	2	88	3	9	86	0	8	90	0	3	100	0	3	100	0
	6	0	4	6	1	7	9	6	6	10	10	3	0	2	0	2	4	0	1
	7	0	8	88	0	5	88	2	0	90	0	0	100	0	0	98	0	0	99
	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
252*5	3	100	0	0	100	0	0	99	0	0	95	0	0	92	0	0	100	0	0
	4	0	0	0	0	0	0	1	0	0	4	0	0	8	0	0	0	0	0
	5	0	95	2	0	94	0	0	100	0	1	96	0	0	100	0	0	100	0
	6	0	2	3	0	6	3	0	0	0	0	4	0	0	0	0	0	0	0
	7	0	3	94	0	0	90	0	0	100	0	0	100	0	0	100	0	0	100
	10	0	0	1	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0

Notes:

5 Empirical Applications

In this section, we examine the performance of our model based on Markowitz's portfolio theory. We collected daily stock data of S&P 100 constituents spanning from 2015 to 2022, excluding stocks listed after 2015, from Center for Research in Security Prices (CRSP). This results in a dataset comprising 97 stocks over an eight-year period. We set the first five years (2015-2019) as the in-sample period and the remaining three years as the out-of-sample period. For the network structure, we construct the adjacency matrix based on the common mutual fund ownership. Anton and Polk (2014) found volatility transition among stocks that are connected via common mutual fund ownership. In particular, we let the stocks be connected if they are invested in by at least 5% of active common shareholders for at least four quarters from 2015 to 2019.

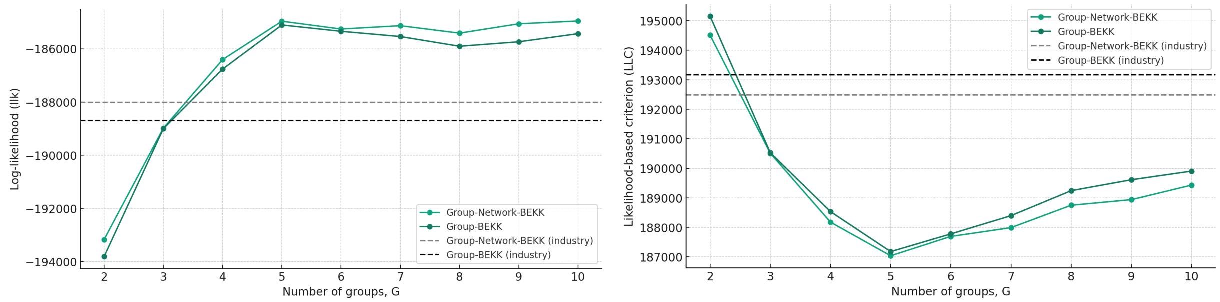
5.1 In-sample Performance

5.1.1 Group Memberships

We fit the model with S&P 100 constituents data using $G = \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$. We compare the model with one setting the group membership as the 10-industry classification from the Kenneth French library. The log-likelihood and LIC of the model with different G and the industry classification are plotted in Figure 1.

As illustrated in the figure, the model incorporating network specification exhibits marginally superior performance compared to the model without network structure. We also highlight that there could be 'more-informed' network specifications other than common mutual fund ownership, which could potentially yield better performance. However, it is not our main focus to explore other specifications in this paper. Furthermore, when

Figure 1: Log-likelihood and LLC of the model with different G and the industry classification



setting $\lambda(N, T, G) = 5log(NT)(N + P)/min(10, n_{0.9})$, the LIC tends to favor $G = 5$.

When setting $G = 5$, we investigate stock industries across various groups. The specifics of the Kenneth French 10-portfolio industry classification are detailed in Table 6. Additionally, Figure 2 presents a cumulative bar graph depicting the distribution of stock industries among these groups. As shown in this figure, stocks in three industries (namely, industries 3, 6, and 9) are categorised within the same group, whereas stocks from the remaining seven industries are dispersed across multiple groups.

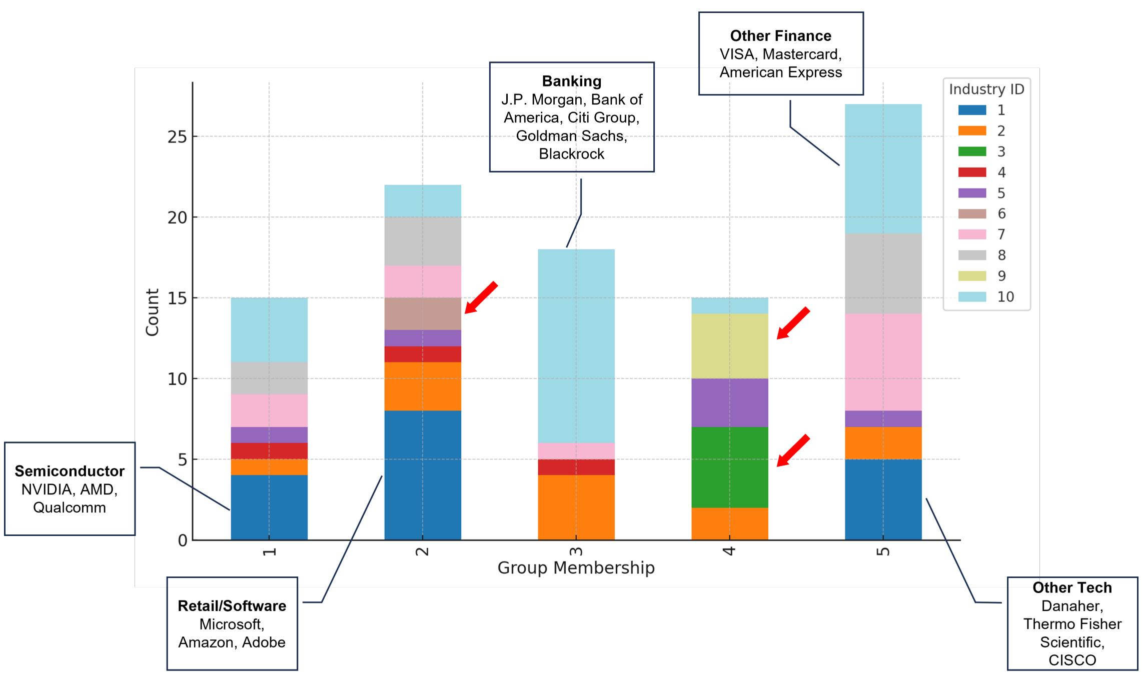
Table 6: Kenneth French 10 portfolio industry classification

1	HiTec	Business Equipment - Computers, Software, and Electronic Equipment
2	Manuf	Manufacturing - Machinery, Trucks, Planes, Chemicals, Off Furn, Paper, Com Printing
3	Durbl	Consumer Nondurables - Food, Tobacco, Textiles, Apparel, Leather, Toys
4	Utils	Utilities
5	Telcm	Telephone and Television Transmission
6	NoDur	Consumer Durables - Cars, TVs, Furniture, Household Appliances
7	Hlth	Healthcare, Medical Equipment, and Drugs
8	Shops	Wholesale, Retail, and Some Services (Laundries, Repair Shops)
9	Enrgy	Oil, Gas, and Coal Extraction and Products
10	Other	Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment, Finance

Additionally, our findings suggest that the algorithm tends to classify stocks with similar characteristics into the same group. For instance, within Industry 1 (represented by the dark blue bar), there are 17 stocks. In this subset, semiconductor companies such as

NVIDIA, AMD, and Qualcomm are classified into Group One; retail and software companies including Microsoft, Amazon, and Adobe into Group Two; and other technology firms like Danaher, Thermo Fisher Scientific, and CISCO into Group Five. Similarly, Although Industry 10 (indicated by the light blue bar) is a generic classification for miscellaneous stocks, the algorithm consistently places all banking firms, including J.P. Morgan, Bank of America, Citi Group, Goldman Sachs, and Blackrock, into Group Three. Meanwhile, other financial entities like VISA, Mastercard, and American Express are assigned to Group Five. These patterns suggest that industry classification may not be the most effective method for capturing volatility transitions among diverse stocks.

Figure 2: The number of stock industries in each group



5.1.2 Estimated Parameters

The parameter matrices when $G = 5$ are as follows:

As indicated in the matrices, all entries within the baseline variance and covariance matrix

$$C_{\mathbb{G}} = \begin{pmatrix} 0.199 & 0.094 & 0.027 & 0.008 & 0.030 \\ 0 & 0.040 & 0.019 & 0.016 & 0.042 \\ 0 & 0 & 0.194 & 0.015 & 0.017 \\ 0 & 0 & 0 & 0.149 & 0.022 \\ 0 & 0 & 0 & 0 & 0.025 \end{pmatrix} \quad A^*_{\mathbb{G}} = \begin{pmatrix} -0.579 & -1.301 & -1.522 & -1.473 & -0.619 \\ 1.596 & 1.704 & 1.201 & 2.993 & 0.487 \\ -2.345 & -2.297 & -2.207 & -1.355 & -1.720 \\ -2.022 & -3.015 & -3.084 & -0.988 & -2.307 \\ 1.583 & 2.143 & 2.761 & -0.015 & 2.487 \end{pmatrix} \quad B^*_{\mathbb{G}} = \begin{pmatrix} -3.019 & -3.767 & -3.410 & -1.810 & -2.234 \\ 0.402 & 1.064 & 0.880 & -0.607 & -0.650 \\ 0.160 & -0.022 & 0.102 & -2.748 & -0.546 \\ -0.954 & -1.363 & -1.399 & -3.786 & -1.898 \\ 4.586 & 5.069 & 4.914 & 5.841 & 5.796 \end{pmatrix}$$

$$diag(C_{\mathbb{G}}) = \begin{pmatrix} 0.017 & 0.512 & 0.507 & 0.059 & 0.369 \end{pmatrix} \quad diag(A^*_{\mathbb{G}}) = \begin{pmatrix} 0.396 & 0.217 & 0.181 & 0.276 & 0.262 \end{pmatrix} \quad diag(B^*_{\mathbb{G}}) = \begin{pmatrix} 0.804 & 0.689 & 0.461 & 0.693 & 0.688 \end{pmatrix}$$

$C_{\mathbb{G}}$ are positive. This outcome aligns with our expectations given that the dataset comprises solely large-cap stocks from the S&P 100 Index. Furthermore, the baseline variances for stocks within groups 2, 3, and 5 are notably higher compared to those in the remaining groups. In terms of the coefficient matrix for past return shocks $A^*_{\mathbb{G}}$ and the variances $B^*_{\mathbb{G}}$, the estimated parameters suggest a more even distribution of the impact of past return shocks across all five groups. However, the covariance matrix appears to be predominantly affected by the variances and covariances of stocks in groups 1 and 5.

5.2 Out-of-sample Forecasting

In the out-of-sample period, we assess the variance and covariance matrix forecasts by Morkowitz (1952)'s portfolio theory. A superior model will report a lower portfolio variance compared to others. For N stocks with an expected return vector $\boldsymbol{\mu}$ and a covariance matrix H_t. The weight of global minimum variance (GMV) portfolio is the solution to the following problem:

$$\min_{\omega} \omega' H_t \omega, \quad \text{subject to } \omega' \iota = 1$$

Therefore,

$$\omega_{GMV} = \frac{1}{A} H_t^{-1} \iota$$

where ι is a vector of one and $A = \iota' H_t^{-1} \iota$. The MV portfolio imposes an additional minimum return constraint $\omega' \boldsymbol{\mu} \geq q$. Then the optimal weight can be expressed as:

$$\omega_{MV} = \frac{C - qB}{AC - B^2} H_t^{-1} \iota + \frac{qA - B}{AC - B^2} H_t^{-1} \boldsymbol{\mu}$$

where $B = \iota' H_t^{-1} \boldsymbol{\mu}$ and $C = \boldsymbol{\mu}' H_t^{-1} \boldsymbol{\mu}$.

We employ the parameters estimated during the in-sample period to forecast H_t in the out-of-sample period, subsequently constructing the portfolio. Following this, we calculate the variance of the portfolio's return. We compare the model with and without the network structure. In addition, we also consider a model whose group memberships are defined by the industry classification. Moreover, We also fit the model with the DCC model and the dynamic equicorrelation (DECO) model by Engle and Kelly (2012). The results are reported in Table 7.

As illustrated in the table, portfolio variances do not make that much difference when $G \geq 5$ and the portfolio when $G = 7$ reports the best performance. In addition, both the model incorporating network structures and the one without them outperform the model that uses industry classification, as well as DCC and DECO models in terms of performance.

6 Concluding Remarks

In this paper, we propose a group network multivariate GARCH model based on BEKK(1,1,1). The integration of group and network structures significantly reduces the number of parameters, thereby facilitating model estimation. Model parameters and group membership are simultaneously estimated through an optimisation algorithm, and the theoretical properties of the estimator are investigated. Additionally, an empirical analysis is

Table 7: Global minimum variance and minimum variance portfolio comparison

Global Minimum Variance (GMV) Portfolio									
Number of Groups	2	3	4	5	6	7	8	9	10
Group Network BEKK	1.551	1.268	0.802	0.757	0.775	**0.751**	0.773	0.768	0.755
Group BEKK	1.554	1.290	0.887	0.824	0.824	**0.817**	0.835	0.818	0.833
Group Network BEKK (Industry)									0.895
Group BEKK (Industry)									0.926
DECO (Industry)									1.062
DCC									1.268
Minimum Variance (MV) Portfolio									
Number of Groups	2	3	4	5	6	7	8	9	10
Group Network BEKK	1.545	1.262	0.802	0.756	0.769	**0.749**	0.768	0.762	0.751
Group BEKK	1.551	1.288	0.885	0.821	0.822	0.817	0.827	**0.808**	0.822
Group Network BEKK (Industry)									0.875
Group BEKK (Industry)									0.917
DECO (Industry)									0.993
DCC									1.158

conducted on the S&P 100 constituents to demonstrate the model's usefulness. The out-of-sample forecasting performance indicates that our model can construct hedge portfolios with substantially lower variance compared to competitor models.

References

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

Bauwens, L., S. Laurent, and J. V. Rombouts (2006). Multivariate garch models: a survey. *Journal of Applied Econometrics 21*(1), 79–109.

Bollerslev, T. (1990). Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The Review of Economics and Statistics*, 498–505.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

Engle, R. (2009). *Anticipating correlations: a new paradigm for risk management*. Princeton University Press.

Engle, R. and B. Kelly (2012). Dynamic equicorrelation. *Journal of Business & Economic Statistics 30*(2), 212–228.

Engle, R. and J. Mezrich (1996). Garch for groups. *Risk 9*, 36–40.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the econometric society*, 987–1007.

Engle, R. F. and K. F. Kroner (1995). Multivariate simultaneous generalized arch. *Econometric Theory 11*(1), 122–150.

Engle, R. F., O. Ledoit, and M. Wolf (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics 37*(2), 363–375.

Francq, C., L. Horvath, and J.-M. Zakoïan (2011). Merits and drawbacks of variance targeting in garch models. *Journal of Financial Econometrics 9*(4), 619–656.

Francq, C. and J.-M. Zakoian (2019). *GARCH models: structure, statistical inference and financial applications*. John Wiley & Sons.

Hafner, C. M. and O. Reznikova (2012). On the estimation of dynamic conditional correlation models. *Computational Statistics & Data Analysis 56*(11), 3533–3545.

Ledoit, O. and M. Wolf (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics 40*(2), 1024–1060.

Liu, R., Z. Shang, Y. Zhang, and Q. Zhou (2020). Identification and estimation in panel models with overspecified number of groups. *Journal of econometrics 215*(2), 574–590.

Morkowitz, H. (1952). Portfolio selection. *Journal of Finance 7*(1), 77–91.

Pakel, C., N. Shephard, K. Sheppard, and R. F. Engle (2021). Fitting vast dimensional time-varying covariance models. *Journal of Business & Economic Statistics 39*(3), 652–668.

Pedersen, R. S. and A. Rahbek (2014). Multivariate variance targeting in the bekk–garch model. *The Econometrics Journal 17*(1), 24–55.

Scherrer, W. and E. Ribarits (2007). On the parametrization of multivariate garch models. *Econometric Theory 23*(3), 464–484.

Silvennoinen, A. and T. Teräsvirta (2009). Multivariate garch models. In *Handbook of financial time series*, pp. 201–229. Springer.

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., G. Xu, and J. Fan (2023). Simultaneous estimation and group identification for network vector autoregressive model with heterogeneous nodes. *Journal of Econometrics*, 105564.