mathematics

MDPI

Article

Element Aggregation for Estimation of High-Dimensional Covariance Matrices

Jingying Yang

School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China; jyyang1005@gmail.com

Abstract: This study addresses the challenge of estimating high-dimensional covariance matrices in financial markets, where traditional sparsity assumptions often fail due to the interdependence of stock returns across sectors. We present an innovative element-aggregation method that aggregates matrix entries to estimate covariance matrices. This method is designed to be applicable to both sparse and non-sparse matrices, transcending the limitations of sparsity-based approaches. The computational simplicity of the method's implementation ensures low complexity, making it a practical tool for real-world applications. Theoretical analysis then confirms the method's consistency and effectiveness with its convergence rate in specific scenarios. Additionally, numerical experiments validate the method's superior algorithmic performance compared to conventional methods, as well as the reduction in relative estimation errors. Furthermore, empirical studies in financial portfolio optimization demonstrate the method's significant risk management benefits, particularly its ability to effectively mitigate portfolio risk even with limited sample sizes.

Keywords: covariance matrix; factor models; high dimensionality; portfolio allocation; element aggregation

MSC: 62H20; 62J07

Citation: Yang, J. Element Aggregation for Estimation of High-Dimensional Covariance Matrices. *Mathematics* **2024**, *12*, 1045. https://doi.org/10.3390/math12071045

Academic Editors: Hong Wang, Liang Shen and Xuewei Cheng

Received: 29 February 2024
Revised: 26 March 2024
Accepted: 28 March 2024
Published: 30 March 2024

1. Introduction

Covariance matrix estimation is essential in statistics [1], econometrics [2], finance [3], genomic studies [4], and other related fields. This matrix quantifies the pairwise interdependencies between variables in a dataset, and each element signifies the covariance between two variables. The accuracy of this estimation is critical for statistical and data analyses. As the size of the matrix increases, the efficiency of the estimation process often decreases. Therefore, the development of efficient methods for this estimation remains a significant research challenge. Various approaches concentrate on assuming specific features in the matrix. Initial methods include principal component analysis (PCA) [5] and factor models [6]. Other prevalent techniques encompass the constant correlation approach [7], maximum likelihood estimation (MLE) [8], shrinkage methods [9,10], and so on [11]. Notably, shrinkage methods have demonstrated exceptional efficacy in financial portfolio allocation.

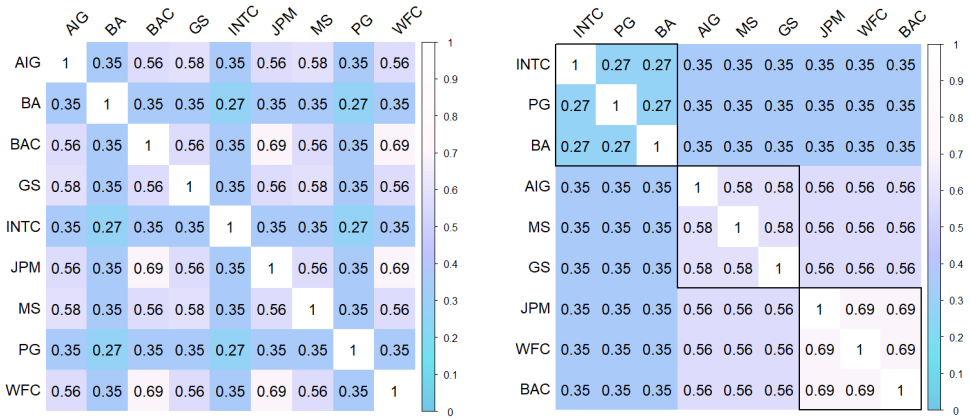
For the estimation of sparse or approximately sparse covariance matrices, various element-wise thresholding techniques for the sample covariance matrix have emerged. These include hard thresholding [12,13], soft thresholding with extensions [14], and adaptive thresholding [15,16]. Generally, these methods offer low computational demands and high consistency. However, the resulting matrix might not always be semi-positive definite. To address this, more sophisticated methods have been introduced to ensure the estimators are semi-positive definite [17,18].

While the sparsity condition is often assumed in [6,19], it is not universally applicable. For instance, in financial studies, variables such as stock returns often share common

factors, making the sparsity assumption less appropriate. Instead, a more common pattern in the covariance matrix of financial data is the clustering of entries. This occurs because stocks within the same industry sector are often similarly correlated with stocks from other sectors [20]. An illustrative example is provided by [21,22], who analyzed the daily returns of nine companies on the New York Stock Exchange (NYSE) market based on 2515 observations from 2000 to 2009. Their analysis revealed a matrix of correlation coefficients without zero entries, with many coefficients sharing identical values. Using statistical hypothesis testing, the correlation coefficients were grouped into five distinct values $(0.27, 0.35, 0.56, 0.58, 0.69)$, shown in Figure 1a. A similar observation was made by [23]. When stocks are arranged in a particular order, the correlation coefficient matrix manifests itself as a block-symmetric matrix, as depicted in Figure 1b, which can be written as

$$\begin{pmatrix} 1-0.27 & 0 & 0 \\ 0 & 1-0.58 & 0 \\ 0 & 0 & 1-0.69 \end{pmatrix} \otimes I_3 + \begin{pmatrix} 0.27 & 0.35 & 0.35 \\ 0.35 & 0.58 & 0.56 \\ 0.35 & 0.56 & 0.69 \end{pmatrix} \otimes \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix},$$

where $\otimes$ is the Kronecker product. We will demonstrate later that this uncomplicated framework is highly effective for estimating the covariance matrix in Corollary 3.



(**a**) Under the original order (**b**) Under the arranged order

Figure 1. Correlation coefficient matrix of the daily returns of 9 companies with stock symbols AIG, BA, BAC, GS, INTC, JPM, MS, PG, and WFC on the NYSE market.

The proposed estimation method in this paper derives from the correlation matrix structure mentioned earlier. This structure is widespread in numerous financial research situations, encompassing sparse covariance matrices, the formerly noted block-wise covariance matrices, and all correlation coefficients in a global constant [24]. However, many existing methods are unsuitable for estimating covariance matrices under this structure. To broaden our scope, we delve into the estimation of covariance matrices with clustered entries. To achieve this, we propose an element-aggregation method that is tailored towards covariance matrix estimation. Moreover, we examine the theoretical properties of our method and confirm its effectiveness through extensive numerical simulations and real-world data analysis.

The rest of this paper is organized as follows. In Section 2, we propose the element-aggregation method for covariance matrix estimation while also elucidating its implementation. Section 3 assesses the theoretical consistency of the estimation errors and exhibits corresponding convergence rates. All theoretical proofs are in Appendixes A–F. Numerical simulation analyses and real data analysis with portfolio allocation are presented in Sections 4 and 5, respectively. Conclusions are provided in Section 6.

2. Estimation and Implementation for Covariance Matrix

Suppose the multiple random variable $X = (\mathbf{x}_1, ..., \mathbf{x}_p)^\top$ has the covariance matrix $\Sigma = (\sigma_{ij})_{1\le i,j\le p}$, and samples $X_i = (x_{i1}, ..., x_{ip})^\top$, $i = 1, ..., n$ are generated from X. The sample covariance matrix is defined as $\hat{\Sigma} = (\hat{\sigma}_{ij})_{1\le i,j\le p}$. For $1 \le i, j \le p$,

$$\hat{\sigma}_{ij} = n^{-1} \sum_{\ell=1}^{n} \pi_{\ell,ij},$$

where $\pi_{\ell,ij} = (x_{\ell i} - \bar{x}_i)(x_{\ell j} - \bar{x}_j)$ and $\bar{x}_k = n^{-1} \sum_{\ell=1}^{n} x_{\ell k}$ for $k = 1, ..., p$.

Motivated by the structure of the correlation matrix discussed in the introduction and aiming to analyze the covariance matrices with clustered entries, we consider the following specification for the covariance matrix. Let $\Omega = \{\sigma_{ij} : i < j\}$ denote the collection of off-diagonal elements in the covariance matrix Σ, and let the set of distinct elements in Ω be represented by

$$\aleph = \{\varsigma_k \in \Omega : \varsigma_k \neq \varsigma_{k'}, \forall\, 1 \le k \neq k' \le K\},$$

where K is the cardinality of set $\aleph$, and K changes with p. In other words, $\sigma_{ij} \in \aleph$ for all $1 \le i < j \le p$. For $k = 1, ..., K$, the index set of covariance matrix elements that are equal to ς_k is defined as

$$\aleph_k = \{(i, j) : \sigma_{ij} = \varsigma_k, 1 \le i < j \le p\}.$$

In brief, $\aleph_k$ stands for the category of ς_k. Similarly, for $1 \le i < j \le p$, the index set of covariance matrix elements that are equal to σ_{ij} is defined as

$$\aleph_{ij} = \{(a, b) : \sigma_{ab} = \sigma_{ij}, 1 \le a < b \le p\}.$$

Then, for $(i, j) \in \aleph_k$, we have $\aleph_{ij} = \aleph_k$.

If the sets $\aleph_k, k = 1, ..., K$ are known, i.e., the indices for same elements in Σ are known, then we can estimate ς_k by averaging the sample covariance elements $\hat{\sigma}_{ij}$ as follows for all $(i, j) \in \aleph_k$,

$$\begin{aligned} \check{\varsigma}_k =& \frac{1}{\#\aleph_k} \sum_{(i,j)\in\aleph_k} \hat{\sigma}_{ij} \\ =& n^{-1} \frac{1}{\#\aleph_k} \sum_{\ell=1}^{n} \sum_{(i,j)\in\aleph_k} \pi_{\ell,ij}, \end{aligned} \tag{1}$$

where $\#\aleph_k$ is the cardinality of set $\aleph_k$. Thus, the covariance matrix Σ can be estimated by

$$\check{\Sigma} = (\check{\sigma}_{ij}), \text{ with } \check{\sigma}_{ij} = \check{\varsigma}_k, \text{ if } (i, j) \in \aleph_k. \tag{2}$$

Element-Aggregation Estimation Method (ELA)

We hereby introduce the element-aggregation (ELA) estimation method for estimation purposes. As $\aleph_k$ is unknown, we estimate $\aleph_{ij}$, i.e., the category of each σ_{ij}, as follows,

$$\tilde{\aleph}_{ij} = \{(a, b) : |\hat{\sigma}_{ij} - \hat{\sigma}_{ab}| < c\hat{\varrho}_{ij} \sqrt{\frac{\log(p \vee n)}{n}}, 1 \le a < b \le p\}, \tag{3}$$

where $\hat{\varrho}_{ij}$ is a sample estimate of $\varrho_{ij} = Var(\pi_{ij})$ and c is a tuning parameter. Regarding the choice of this tuning parameter c, Cai and Liu [15] show that a good choice of the tuning parameter c does not affect the rate of convergence but it does affect the numerical performance of the estimators. The tuning parameter c can be taken as fixed at $c = 2$, as suggested in [15], or it can be chosen empirically by cross-validation, such as via the five-fold cross-validation method used in [12,15]. The tuning parameter c operates akin to a significance level in statistical testing, and the 2-σ rule is widely endorsed for normally distributed data, wherein approximately 95% of observations are encapsulated within the interval of the mean plus or minus two standard deviations, denoted as $[\mu - 2\sigma, \mu + 2\sigma]$. In

our simulations, we also discover that employing $c = 2$ as the tuning parameter for the proposed ELA algorithm yields commendable performance in high-dimensional cases, and does not differ significantly from the parameter optimization obtained by cross-validation. Consequently, we adopt $c = 2$ as the tuning parameter for our subsequent analysis.

Analogous to (1), we estimate the covariance matrix element σ_{ij} by

$$\tilde{\sigma}_{ij} = n^{-1} \frac{1}{\#\tilde{\aleph}_{ij}} \sum_{\ell=1}^{n} \sum_{(s,t)\in\tilde{\aleph}_{ij}} \pi_{\ell,st}, \tag{4}$$

and Σ by $\tilde{\Sigma} = (\tilde{\sigma}_{ij})$. This is the element-aggregation (ELA) estimator of the covariance matrix.

In terms of computational complexity, the ELA process requires $O(p^2 \log(p))$ time, where $\log(p)$ accounts for the binary search computation of $\tilde{\aleph}_{ij}$ in $\{\hat{\sigma}_{ab} : 1 \le a < b \le p\}$. In comparison, element-wise thresholding has a computational complexity of $O(p^2)$. Therefore, the ELA estimation calculation is only slightly more complex than the element-wise thresholding approach.

3. Theoretical Properties

In this section, we provide the theoretical justification for the element-aggregation estimation under the assumption of a normal distribution of X. The results can be proven under more general conditions using more complicated techniques. Let $\|A\|$ denote the operator norm, i.e., $\|A\| = \{\lambda_{max}(AA^\top)\}^{1/2}$, where $\lambda_{max}(AA^\top)$ is the square root of the largest eigenvalue of $AA^\top$.

Lemma 1. *Let K be the number of different values of the off-diagonal elements in the covariance matrix $Cov(X) = \Sigma$. Let $n_{i,k}$ be the numbers of elements that are equal to ς_k in the i-th row of Σ, i.e., $n_{i,k} = \#\{j : \sigma_{ij} = \varsigma_k, 1 \le j \ne i \le p\}$. If $\max_{1\le i,j\le p} \sigma_{ij} \le M_0$ for some constant M_0 and*

$$Var\Big(\frac{1}{\sqrt{\#\aleph_k}} \sum_{(i,j)\in\aleph_k} \pi_{\ell,ij}\Big) = O(V_k^2),\ k = 1, ..., K,$$

where $V_k = \mathrm{Var}(\check{\varsigma}_k)$ and $\check{\varsigma}_k$ is the estimate in (1), then the covariance matrix estimator $\check{\Sigma}$ in (2) has the following estimation error

$$\|\check{\Sigma} - \Sigma\| = O_P\Big(\sqrt{\frac{\log(p \vee n)}{n}}\Big) \sum_{k=1}^{K} \frac{n_{j,k}}{\sqrt{\#\aleph_k}} V_k.$$

An important characteristic is that the effect of dimension on the estimation error is regulated by a slowly varying function, $\log(p)$. Lemma 1 could also be extended to the case that $\aleph_k$ is unknown. In such cases, we need to identify the corresponding category $\aleph_k$. To simplify the explanation, we will use $\aleph_{ij}$ here, since $\aleph_{ij} = \aleph_k$ if $(i, j) \in \aleph_k$.

Lemma 2. *Suppose X is normally distributed with $\max_{1\le i,j\le p} \sigma_{ij} \le M_0$ for some constant M_0. If*

$$\sqrt{n} \min\{|\varsigma_k - \varsigma_j|, k \ne j\} / (\log(p \vee n))^{1/2} \to \infty,$$

then the identification by (3) is consistent, i.e., for any $(i, j) \in \aleph_k$,

$$Pr\Big(\bigcup_{1\le i,j\le p} \{\tilde{\aleph}_{ij} \ne \aleph_{ij}\}\Big) \to 0.$$

Theorem 1. *With the above notation, if the conditions in Lemmas 1 and 2 are satisfied, then*

$$\|\tilde{\Sigma} - \Sigma\| = O_P\Big(\sqrt{\frac{\log(p \vee n)}{n}}\Big) \sum_{k=1}^{K} \frac{n_{j,k}}{\sqrt{\#\aleph_k}} V_k. \tag{5}$$

Below, we give some special cases of Theorem 1, and show the convergence rates in Corollaries 1–3.

Corollary 1. *Suppose* $\Sigma = (\sigma_{ij})_{1\le i,j\le p}$ *is sparse with a number of non-zero elements in each row of* $o(p)$*, i.e.,* $c(p)/p \to 0$*, where* $c(p) = \max_j \sum_{i=1}^{p} (\sigma_{ij} \neq 0)$*. If the conditions in Theorem 1 hold, then*

$$\|\tilde{\Sigma} - \Sigma\| = O_P\Big(c(p)\sqrt{\frac{\log(p \vee n)}{n}}\Big).$$

This result is a special case in [12]. It could be extended to more general cases.

Corollary 2. *Suppose* $X = (\mathbf{x}_1, ..., \mathbf{x}_p)^\top$ *follows normal distribution with* $\Sigma = (\varsigma_{|i-j|})_{1\le i,j\le p}$*. If* $|\varsigma_k| = O(c^k)$ *for some* $c < 1$*, then*

$$\|\tilde{\Sigma} - \Sigma\| = O_P\Big(\log(p)\sqrt{\frac{\log(p \vee n)}{n}}\Big).$$

Finally, we consider a simple block covariance matrix with the matrix in Figure 1b as a special case.

Corollary 3. *Suppose* $\Sigma = \Lambda_M \otimes I_m + \Phi_M \otimes \Xi_m$ *where* I_m *is a* $m \times m$ *identity matrix,* Λ_M *is a diagonal matrix,* Ξ_m *is a* $m \times m$ *matrix with all elements 1, and* Φ_M *is an* $M \times M$ *symmetric matrix. If* M *is fixed as* $p \to \infty$*, then we have*

$$||\tilde{\Sigma} - \Sigma|| = O_P(n^{-1/2}).$$

Remark 1. *Since the dimension* M *of the matrix* Φ_M *and the diagonal matrix* Λ_M *assumed in Corollary 3 are fixed (*$p \to \infty$*), the different values for their matrix elements are at most* $M(M+1)/2$ *and* M*, respectively, so* Σ *also has at most* $M(M+3)/2$ *different values, which is fixed at* $p \to \infty$*. By Lemma 2, the estimates* $\tilde{\Lambda}_M$ *and* $\tilde{\Phi}_M$ *satisfy the rate of* $O_P(n^{-1/2})$*, i.e., for any* $\delta > 0$*, a subset of the probability space* $S_n : Pr(S_n) > 1 - \delta$ *can be found, such that*

$$||\tilde{\Lambda}_M - \Lambda_M|| = O_P(n^{-1/2}), \quad ||\tilde{\Phi}_M - \Phi_M|| = O_P(n^{-1/2}).$$

where $\tilde{\Sigma} = \tilde{\Lambda}_M \otimes I_m + \tilde{\Phi}_M \otimes \Xi_m$*. Using the eigenvalue and Kronecker product properties, we show in the Appendixes A–F that the covariance matrix estimate* $\tilde{\Sigma}$ *also has the same rate* $O_P(n^{-1/2})$*. That is, since the number of selectable elements of the covariance matrix is determined by a fixed* M*, which is not dependent on* p*, the rate of convergence for the covariance matrix estimate* $\tilde{\Sigma}$ *is also independent of* p*.*

Our ELA method efficiently estimates the covariance matrix with the above structure, demonstrating both efficiency and dimensional independence. Mathematically, this block-wise structure is straightforward. Furthermore, we hypothesize that these conclusions can be extended to more general block-wise matrices, such as the Khatri–Rao product or the Tracy–Singh product [25].

4. Numerical Simulation

Our study presents the ELA method to estimate the covariance matrix using an element-aggregation idea. The ELA method is designed for simplicity and clarity, making it easy to implement. The steps for implementing the method were thoroughly detailed in the preceding section. This section focuses on comparing the algorithmic performance of the proposed method with several estimation methods for high-dimensional sparse and non-sparse covariance matrices. We compare our ELA estimation method with multiple established methods: (1) the adaptive thresholding method [15], (2) the POET$_k$ method [19] with $k = 0, 1, 2, \ldots$ factors, and (3) the Rothman method [18].

The numerical simulation focuses on the relative matrix errors when estimating the covariance matrix and the correlation coefficient matrix. To estimate the covariance matrix, first standardize each variable. Then, multiply the elements of the correlation coefficient matrix by the sample standard deviations of each random variable. This effectively transforms the estimation of the covariance matrix into the estimation of the corresponding correlation coefficient matrix. Thus, the ELA method's efficiency and accuracy in capturing the underlying data structure can be comprehensively assessed through the dual focus on both covariance and correlation coefficient matrix estimations.

The metric employed to compare the estimation efficiency between these estimation methods is the average matrix loss over 500 replications. The matrix losses are measured by the spectral norm of the correlation coefficient matrix and the covariance matrix, similar to that performed in [18]. To aid visualization, we define the relative estimation errors between the estimated matrix and the actual matrix for both the correlation coefficient matrix R and the covariance matrix Σ as follows:

$$e_{cor} = \frac{\|\hat{R} - R\|}{\|R\|} \times 100\%, \quad e_{cov} = \frac{\|\hat{\Sigma} - \Sigma\|}{\|\Sigma\|} \times 100\%.$$

where $\|A\| = \{\lambda_{max}(AA^\top)\}^{1/2}$ denote the operator norm. The next simulation compares the relative matrix estimation errors for various matrix types and algorithms, but also shows how the matrix estimation error changes as the matrix dimension p increases, in order to demonstrate more clearly the superiority of the proposed estimation algorithm.

The following three covariance matrices $\Sigma = (\sigma_{ij})_{1 \le i,j \le p}$ are considered.

Example 1. *$\sigma_{ij} = \sqrt{ij}\rho^{|i-j|}$ with $\rho = -0.9$.*

Example 2. *$\sigma_{ij} = (ij)^{1/4} r_{ij}$ where $r_{ii} = 1$, $r_{ij} = 0.6|i-j|^{-1}u_{ij}$ for $i \neq j$ and $u_{ij} = u_{ji}$ are IID uniform on [0, 1]. A similar example was used in [16].*

Example 3. *Constant block matrix*

$$\Sigma = \begin{pmatrix} \Sigma_1 & \Sigma_2 \\ \Sigma_2 & \Sigma_3 \end{pmatrix}$$

where $\Sigma_1 = 0.8\Xi_{p/2} + 0.2I_{p/2}, \Sigma_2 = -0.4\Xi_{p/2}, \Sigma_3 = 0.3\Xi_{p/2} + 0.7I_{p/2}$, $\Xi_{p/2}$ is a $(p/2) \times (p/2)$ matrix with all elements 1, and $I_{p/2}$ is the identity matrix of size $p/2$.

Remark 2. *Due to the assumption of sparse covariance matrices in the adaptive thresholding method, the POET$_0$ method, and the Rothman estimation method [6,15,18], the covariance matrix is estimated as a sparse matrix, where most of the matrix elements are estimated to be zero. This is significantly different from the non-sparse covariance matrix in Example 3. The relative estimation error of the covariance matrix estimated by these methods is relatively large compared to the sample covariance estimation method. Therefore, the adaptive thresholding method, the POET$_0$ method, and the Rothman method cannot be used directly for Example 3.*

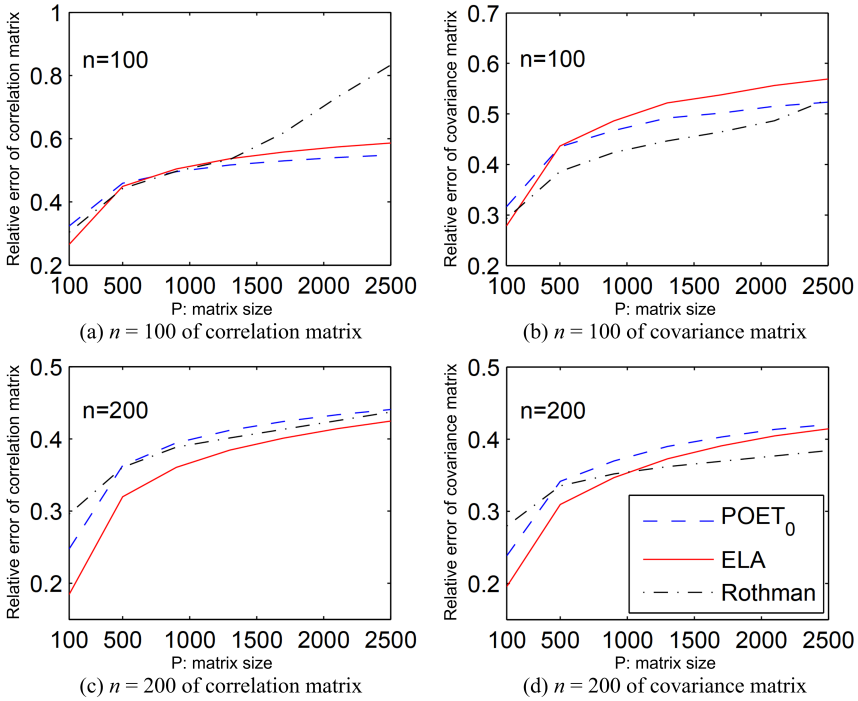
Note that Σ can be written as

$$\Sigma = \Sigma_0 + \lambda_1 \ell_1 \ell_1^\top + \lambda_2 \ell_2 \ell_2^\top,$$

where Σ_0 is a sparse matrix and λ_1 and λ_2 are the first two largest eigenvalues of Σ, with ℓ_1 and ℓ_2 as the corresponding eigenvectors, respectively. By this decomposition, the POET$_k$ method with $k = 2$ factors in [19] can be used, i.e., POET$_2$ is applicable to the estimation for Example 3.

Random samples are generated from $X \sim N(0, \Sigma)$ with sample sizes $n = (100, 200)$ and $p = (100, 500, 900, 1400, 1900, 2300, 2500)$. The averages of the relative estimation errors, based on 500 replications, are depicted in Figures 2–4. The ELA method is represented by a

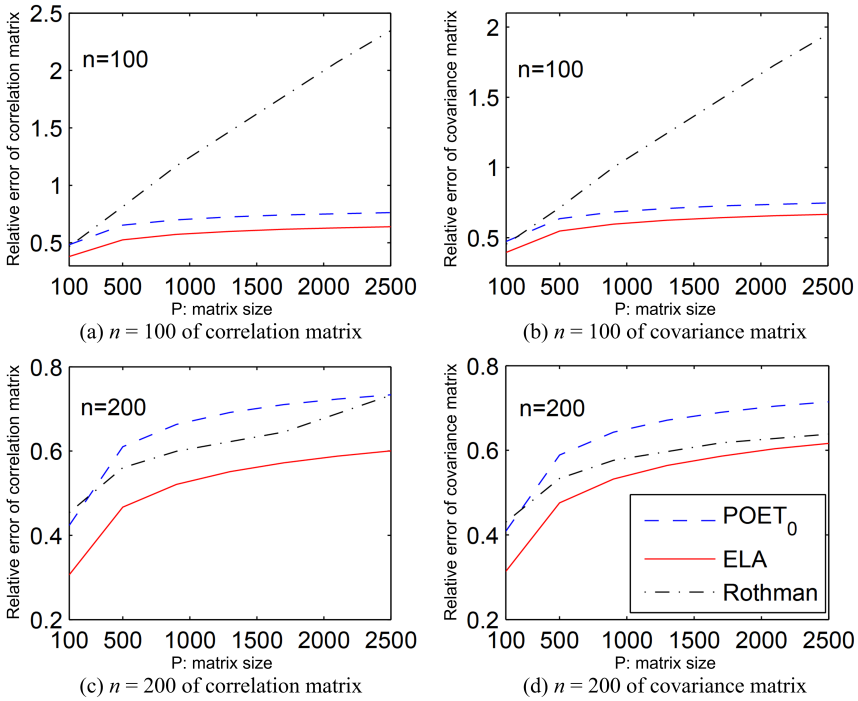
solid red line. Only methods with a performance comparable to the best one are displayed in each panel.



(a) $n = 100$ of correlation matrix (b) $n = 100$ of covariance matrix

(c) $n = 200$ of correlation matrix (d) $n = 200$ of covariance matrix

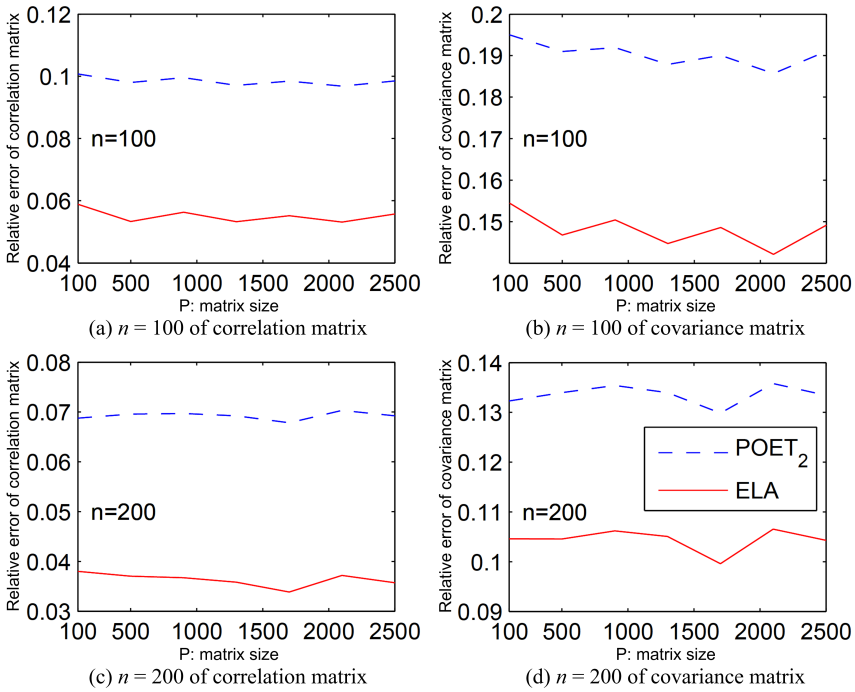
Figure 2. Relative estimation error for the correlation matrix (**left**) and the covariance matrix (**right**) in Example 1 with sample sizes $n = 100$ and 200; the estimation is based on 500 replications per data point.



(a) $n = 100$ of correlation matrix (b) $n = 100$ of covariance matrix

(c) $n = 200$ of correlation matrix (d) $n = 200$ of covariance matrix

Figure 3. Relative estimation error for the correlation matrix (**left**) and the covariance matrix (**right**) in Example 2 with sample sizes $n = 100$ and 200; the estimation is based on 500 replications per data point.



(a) $n = 100$ of correlation matrix (b) $n = 100$ of covariance matrix

(c) $n = 200$ of correlation matrix (d) $n = 200$ of covariance matrix

Figure 4. Relative estimation error for the correlation matrix (**left**) and the covariance matrix (**right**) in Example 3 with sample sizes $n = 100$ and 200; the estimation is based on 500 replications per data point.

In Figure 2 for Example 1 with $n = 100$, our ELA estimator for the correlation coefficient matrix R is slightly inferior to POET_0 but exceeds Rothman. For the covariance matrix Σ, our method slightly lags behind both POET_0 and Rothman. However, with $n = 200$, the ELA method outperforms them in both the correlation coefficient matrix and the covariance matrix with $p < 1000$, ranking first for the correlation coefficient matrix and second for the covariance matrix with $p > 1000$. In general, the ELA method is comparable to both POET and Rothman for Example 1.

In Figure 3 for Example 2, the ELA estimator consistently outperforms both POET and Rothman for both matrices when $n = (100, 200)$. Although Remark 2 suggests that POET_2 is suitable for Example 3, Figure 4 for Example 3 demonstrates the superior efficiency of the ELA method over POET_2 for both matrices with $n = (100, 200)$. In summary, the ELA method significantly outperforms both POET and Rothman for Examples 2 and 3.

Furthermore, we conducted a comprehensive analysis comparing the average computational time across 500 replications for our ELA method alongside the adaptive thresholding method [15], the POET method [19], and the Rothman method [18]. These comparisons are detailed in Table 1. The random samples are still generated from $X \sim N(0, \Sigma)$ with Σ following Example 1, sample sizes $n = (100, 200)$, and covariance matrix dimensions $p = (100, 500, 900, 1400, 1900, 2300, 2500)$. Notably, due to the POET_k methods exhibiting computational times that are broadly analogous for different values of k, we present only the results for POET_1.

Upon examination of Table 1, it is evident that, for lower dimensions of $p = (100, 500)$, the ELA method has a higher computational cost compared to the adaptive thresholding and Rothman methods but outperforms the POET method. For $p = (900, 1400)$, the ELA method is slightly slower than adaptive thresholding but remains superior to both the POET and Rothman methods. When dealing with higher dimensions of $p = (2300, 2500)$, the ELA method demonstrates a significant computational advantage over its counterparts. This suggests that the ELA method scales more efficiently with increasing dimensionality, which is a desirable attribute for high-dimensional data analysis.

Table 1. The average computational time over 500 replications of the ELA method, the adaptive thresholding method, the POET_1 method, and the Rothman method for sample sizes $n = (100, 200)$ and $p = (100, 500, 900, 1400, 1900, 2300, 2500)$ (in seconds).

Method	**Adaptive Thresholding**		**POET**		**Rothman**		**ELA**	
p	$n = 100$	$n = 200$	$n = 100$	$n = 200$	$n = 100$	$n = 200$	$n = 100$	$n = 200$
100	0.005	0.006	0.222	0.815	0.012	0.013	0.034	0.037
500	0.370	0.409	0.806	2.042	0.828	0.830	0.863	0.930
900	2.176	2.207	2.657	5.163	4.680	4.910	2.849	3.131
1400	7.967	7.467	9.248	12.711	18.444	18.442	7.743	8.087
2300	44.901	43.974	34.241	42.228	81.292	81.834	20.525	23.054
2500	60.117	60.876	43.077	52.826	108.323	110.301	25.317	25.570

Our simulation studies have yielded the following conclusions: The ELA method outperforms both POET and Rothman for Examples 2 and 3 in both the correlation coefficient matrix and the covariance matrix with $n = (100, 200)$ and $p = (100, \ldots, 2500)$. For Example 1, it performs comparably to them. These results emphasize the efficiency of the ELA method for estimating covariance matrices. Our proposed ELA technique is not just computationally efficient but also markedly lowers the relative estimation error, i.e., with minimal loss of spectral norm for the correlation coefficient and the covariance matrix.

5. Real Data Analysis and Portfolio Allocation

The component stocks of the SP500 index were analyzed using historical daily prices sourced from Yahoo Finance through the R package 'tidyquant'. We selected 430 stocks from SP500 with daily prices spanning over 3000 days, from 1 January 2008 to 31 December 2019, to ensure a sufficiently large sample size. This allowed for a larger window, $N = 500$, to be selected when we use the rolling window method later. Our emphasis is on the correlation coefficient matrix and covariance matrix for daily returns. Although the covariance matrix, in particular the individual stock variances, is known to fluctuate over time, the correlation coefficient matrix remains relatively stable. This stability underscores the significance of evaluating estimation methods based on the correlation coefficient matrix [26]. The constant conditional correlation multivariate GARCH model [26] is defined as $H_t = \Lambda_t R \Lambda_t$, where $\Lambda_t = diag(\sigma_{1t}, ..., \sigma_{pt})$, $R = (\rho_{ij})_{1 \le i,j \le p}$ is the constant correlation coefficient matrix, and σ_{it} is the conditional volatility modeled by GARH(1,1) for i-th stock based on its past daily returns. See [27] for more discussion about the performance of GARCH(1,1). Let $r_{k,t}$ be the return of the k-th stock on day t, where $k = 1, ..., 430$. We begin by standardizing the returns using

$$u_{k,t} = r_{k,t} / \sigma_{k,t}.$$

In order to evaluate the estimation performance of the covariance matrix, we assume that the covariance matrix of $u_t = (u_{1,t}, ..., u_{p,t})^\top$ remains constant over time or changes very slowly.

Rooted in modern portfolio theory (MPT), portfolio allocation advocates strategically distributing investments across diverse asset classes to optimize the trade-off between risk and return. The MPT theory highlights the significance of not only choosing individual stocks but also the proportional weighting of these stocks in a portfolio [28,29]. Therefore, we investigate the optimal allocation for minimum risk portfolios of stocks, utilizing the estimated covariance matrix of various methods.

To evaluate various methods, we apply their estimators to construct minimum risk portfolios. For any rolling window of time $(t - N, ..., t - 1)$, we utilize data from one window to estimate the covariance matrix via different methods, including (1) our ELA method, (2) the POET_k method [19] with $k = 0, 1, 2$ factors, (3) the shrinkage method [9], denoted by Shrink_{Market}, and (4) the simple sample covariance matrix, denoted by Sample. The simple sample estimate of the covariance matrix Σ for a given time t is defined as $\hat{\Sigma}_t^S = N^{-1} \sum_{s=t-N}^{t-1} (u_s - \bar{u})(u_s - \bar{u})^\top$ with $u_s = (u_{1,s}, ..., u_{p,s})^\top$.

For an estimated covariance matrix $\hat{\Sigma}_t$, the minimum risk portfolio $w_t = (w_{1,t}, ..., w_{p,t})^\top$ for time t is defined as

$$w_t^* = \arg\min_{w_t} w_t^\top \hat{\Sigma}_t w_t$$
$$\text{subject to } w_{i,t} \geq 0,\ i = 1, ..., p$$
$$w_{1,t} + ... + w_{p,t} = 1.$$

The portfolio for time t is then $w_t^\top u_t$. To determine the portfolio with the lowest risk, we calculate the standard deviation of the portfolio returns $\{w_t^\top u_t, t = N+1, ..., T\}$. Estimators yielding a smaller standard deviation are deemed superior in portfolio allocation.

To understand how risk varies with the number of stocks, we alphabetically order the stocks by their symbols on the New York Stock Exchange and analyze growing subsets of stocks with $p = 50, 70, 90, ..., 430$. The risks associated with portfolios based on different methods are illustrated in Figure 5, with the ELA method highlighted by a red solid line. As the sample size N in the rolling window increases from 100 to 500 across the panels, the performance gap between the different estimation methods narrows, illustrating the homogeneity of the correlation matrix.

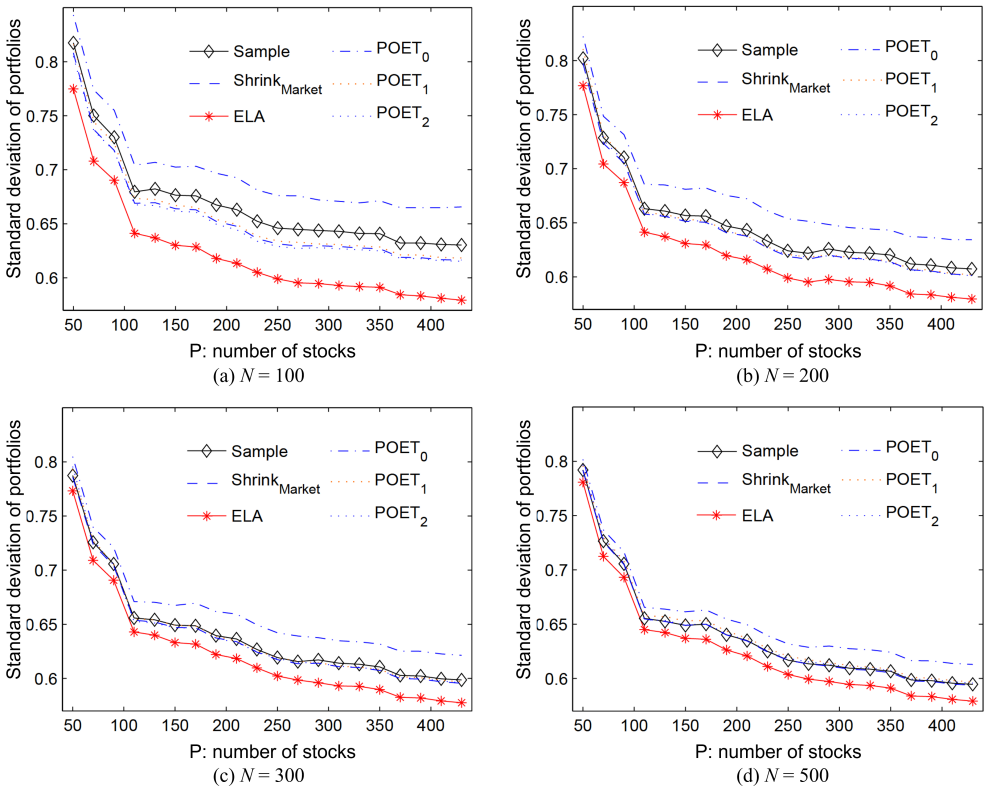


Figure 5. Portfolios of minimum risk based on different estimators of the covariance matrix.

Figure 5 shows that portfolios crafted using the ELA method consistently have the lowest risk compared to the POET$_k$ methods, the shrinkage method, and the sample covariance matrix method, especially with smaller sample sizes of $N = 100$ and $N = 200$. Therefore, the ELA method is superior in constructing portfolios that minimize risk, is an effective tool for reducing portfolio risk, and is suitable for a wide range of sample sizes.

6. Conclusions

In this paper, we introduce a novel method called element-aggregation (ELA) estimation for the estimation of covariance matrices. The ELA method stands out due to its simplicity, low computational complexity, and applicability to both sparse and non-sparse matrices. Our theoretical analysis shows that the ELA method offers strong consistency

while maintaining computational efficiency and dimensional independence, especially concerning block-wise covariance matrices.

In our numerical simulation study, we highlight the exceptional effectiveness of the ELA method in estimating correlation coefficient matrices and covariance matrices using diverse random samples. In comparison to established methods like POET and Rothman, the ELA method consistently either outperforms or equals them. The computational efficiency of our ELA method is complemented by a significant reduction in the relative estimation error for both correlation coefficients and covariance matrices.

In the real data analysis of the SP500 index stocks, the ELA method consistently generates portfolios with the lowest risk in comparison to other listed methods. This outcome is particularly pronounced in scenarios with smaller sample sizes, underscoring the potent ability of the ELA method to construct risk-minimized portfolios.

7. Future Work

This paper delves into the estimation of covariance matrices employing an elemental clustering approach, substantiating and evaluating the consistency and efficiency of the proposed methodology within the realm of high-dimensional data. The exponential growth in data volume and dimensionality due to advancements in computational and storage technologies has led to the emergence of ultra-high-dimensional and high-frequency datasets. Thus, the extensibility of the proposed method to these novel datasets, coupled with the demonstration of its consistency and efficacy, presents a significant avenue for future research. In addition, the exploration of computational strategies to increase efficiency under the constraints of ultra-high-dimensionality and to optimize the trade-off between computational resources and analytical accuracy is imperative. Furthermore, the proposed method has the potential to be integrated into diverse domains, such as psychology, social sciences, and genetic research. This also represents a promising direction for future research.

Funding: This research was funded by the Doctoral Foundation of Yunnan Normal University (Project No. 2020ZB014) and the Youth Project of Yunnan Basic Research Program (Project No. 202201AU070051).

Data Availability Statement: The author confirms that the data supporting the findings of this study are available within the article.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A. Proof of Lemma 1

Proof. First, recall that $\check{\varsigma}_k = \frac{1}{\#\aleph_k} \sum_{(i,j)\in\aleph_k} \hat{\sigma}_{ij}$ with $\sigma_{ij} = \varsigma_k$ for all $(i,j) \in \aleph_k$. We regard $\hat{\sigma}_{ij}$ as independent zero-mean random variables with $E(\hat{\sigma}_{ij}) = \varsigma_k$, for all $(i,j) \in \aleph_k$. By Bernstein inequalities we have

$$|\check{\varsigma}_k - \varsigma_k| = \frac{V_k}{\sqrt{\#\aleph_k}} O_P\Big(\sqrt{\frac{\log(p \vee n)}{n}}\Big), \quad k = 1, ..., K.$$

where $V_k = Var(\check{\varsigma}_k)$. For symmetric matrices, by [30], we have

$$\|\check{\Sigma} - \Sigma\| \leq \max_j \sum_{i=1}^{p} |\check{\sigma}_{ij} - \sigma_{ij}|. \tag{A1}$$

For the right hand side above,

$$\begin{aligned}\sum_{i=1}^{p}|\check{\sigma}_{ij}-\sigma_{ij}| &= \sum_{k=1}^{K}\sum_{i:\sigma_{ij}=\varsigma_k}|\check{\sigma}_{ij}-\varsigma_k| \\ &= \sum_{k=1}^{K} n_{j,k}|\check{\varsigma}_k-\varsigma_k| \\ &\le O_P\Big(\sqrt{\frac{\log(p\vee n)}{n}}\Big)\sum_{k=1}^{K}\frac{n_{j,k}}{\sqrt{\#\aleph_k}}V_k.\end{aligned}$$

We complete the proof. □

Appendix B. Proof of Lemma 2

Proof. Let $D_n = (M\log(p\vee n))^{1/2}\sqrt{\varrho_{ab}/n}$. For any (i,j) such that $\sigma_{ij}=\aleph_k$, recall that $\tilde{\aleph}_{ij}=\{(a,b):|\hat{\sigma}_{ij}-\hat{\sigma}_{ab}|<D_n\}$. Let $W_1=\{(a,b)\in\aleph_k:(a,b)\notin\tilde{\aleph}_{ij}\}$, and $W_2=\{(a,b)\in\tilde{\aleph}_{ij}:(a,b)\notin\aleph_k\}$. Obviously,

$$\{\tilde{\aleph}_{ij}\neq\aleph_k\}\subset W_1\cup W_2.$$

Thus, it is sufficient to prove

$$Pr(W_1)\to 0 \text{ and } Pr(W_2)\to 0.$$

For any $(a,b)\in W_1$, we have $(a,b)\in\aleph_k$, i.e., $\sigma_{ab}=\varsigma_k$, and

$$\{(a,b)\notin\tilde{\aleph}_{ij}\}=\{|\hat{\sigma}_{ij}-\hat{\sigma}_{ab}|\ge D_n\}\subset\{|\hat{\sigma}_{ij}-\varsigma_k|\ge\frac{1}{2}D_n\}\cup\{|\hat{\sigma}_{ab}-\varsigma_k|\ge\frac{1}{2}D_n\}.$$

For any $\sigma_{ab}=\varsigma_k$, if X follows normal distribution, by Bernstein inequality, we have

$$Pr\Big\{|\hat{\sigma}_{ab}-\varsigma_k|\ge\frac{1}{2}D_n\Big\}\le\exp(-\frac{1}{16}M\log(p\vee n))=(p\vee n)^{-M/16} \tag{A2}$$

and

$$Pr\Big\{|\hat{\sigma}_{ij}-\hat{\sigma}_{ab}|\ge D_n\Big\}\le Pr\Big\{|\varsigma_k-\hat{\sigma}_{st}|\ge\frac{1}{2}D_n\Big\}+Pr\Big\{|\varsigma_k-\hat{\sigma}_{ij}|\ge\frac{1}{2}D_n\Big\}\le 2(p\vee n)^{-M/16}.$$

Thus, as $n\to\infty$,

$$Pr(W_1)\le\sum_{(a,b)\in\aleph_k}Pr\{(a,b)\notin\tilde{\aleph}_{ij}\}\le 3(p\vee n)^{-M/16}p^2\to 0, \text{ if } M>32. \tag{A3}$$

Now, for any $(a,b)\in W_2$, we have $(a,b)\notin\aleph_k$, but instead $(a,b)\in\aleph_s$ for some $s\neq k$. By the assumption, we have $|\varsigma_k-\varsigma_s|>2D_n$ when $k\neq s$ and $\log(p\vee n)/n\to 0$. Thus,

$$\begin{aligned}\{|\hat{\sigma}_{ij}-\hat{\sigma}_{ab}|\le D_n\} &\subset\{|(\hat{\sigma}_{ij}-\varsigma_k)+(\varsigma_s-\hat{\sigma}_{ab})+(\varsigma_k-\varsigma_s)|\le D_n\} \\ &\subset\{(|\varsigma_k-\varsigma_s|-|\hat{\sigma}_{ij}-\varsigma_k|-|\varsigma_s-\hat{\sigma}_{ab}|)\le D_n\} \\ &=\{(|\hat{\sigma}_{ij}-\varsigma_k|+|\varsigma_s-\hat{\sigma}_{ab}|)\ge(|\varsigma_k-\varsigma_s|-D_n)\} \\ &\subset\{(|\hat{\sigma}_{ij}-\varsigma_k|+|\varsigma_s-\hat{\sigma}_{ab}|)\ge D_n\} \\ &\subset\{|\hat{\sigma}_{ij}-\varsigma_k|\ge\frac{1}{2}D_n\}\cup\{|\varsigma_s-\hat{\sigma}_{ab}|\ge\frac{1}{2}D_n\}.\end{aligned}$$

Similar to (A2) and (A3), we have

$$Pr(W_2)\le\sum_{(a,b)\in\aleph_k}Pr\{(a,b)\in\tilde{\aleph}_{ij}\}\le 3(p\vee n)^{-M/16}p^2\to 0, \text{ if } M>32. \tag{A4}$$

Thus, Lemma 2 follows from (A3) and (A4). □

Appendix C. Proof of Theorem 1

Proof. It follows immediately by applying Lemma 1 to $\bigcup_{1\le i,j\le p}\{\tilde{\aleph}_{ij} \neq \aleph_k\}$ in Lemma 2. □

Appendix D. Proof of Corollary 1

Proof. For simplicity, we only consider the case with $EX = 0$ and $\hat{\sigma}_{ij} = n^{-1}\sum_{k=1}^{n} \mathbf{x}_{ki}\mathbf{x}_{kj}$. For any $\ell = 1, ..., n$, define $\pi_{\ell,ij} = \mathbf{x}_{\ell i}\mathbf{x}_{\ell j}$ and $\pi_{ij} = \mathbf{x}_i\mathbf{x}_j$. Let $\aleph_1 = \{(i,j) : \sigma_{ij} = 0\}$. By the assumption, we have $\#\aleph_1 = p(p-1)(1-o(1))$ and $\#\aleph_k = O(c(p))$ for $k = 2, ..., K$. Note that

$$V_k^2 = Var(\sum_{(i,j)\in\aleph_k} \pi_{\ell,ij}/\sqrt{\#\aleph_k}) \le \#\aleph_k \max_{(i,j)\in\aleph_k} Var(\hat{\sigma}_{ij}) = O(1)\#\aleph_k. \tag{A5}$$

Obviously, if we can prove that for those $\sigma_{ij} = 0$, the covariance

$$V_1^2 = Var(\sum_{(i,j)\in\aleph_1} \pi_{ij}/\sqrt{\#\aleph_1}) = O(c(p)^2), \tag{A6}$$

then

$$\sum_{k=1}^{K} \frac{n_{j,k}}{\sqrt{\#\aleph_k}} V_k \le \frac{p}{\sqrt{\#\aleph_1}} V_1 + \sum_{k=2}^{K} V_k \le \frac{p}{\sqrt{\#\aleph_1}} O(c(p)) + \sum_{k=2}^{K} \#\aleph_k O(1) = O(c(p)),$$

thus the Corollary 1 follows.

By the properties of normal distribution, we have

$$\begin{aligned} Cov(\pi_{ij}, \pi_{ab}) &= E(\mathbf{x}_i\mathbf{x}_j\mathbf{x}_a\mathbf{x}_b) - \sigma_{ij}\sigma_{ab} \\ &= E(\mathbf{x}_i\mathbf{x}_j)E(\mathbf{x}_a\mathbf{x}_b) + E(\mathbf{x}_i\mathbf{x}_a)E(\mathbf{x}_j\mathbf{x}_b) + E(\mathbf{x}_i\mathbf{x}_b)E(\mathbf{x}_j\mathbf{x}_a) - \sigma_{ij}\sigma_{ab} \\ &= E(\mathbf{x}_i\mathbf{x}_a)E(\mathbf{x}_j\mathbf{x}_b) + E(\mathbf{x}_i\mathbf{x}_b)E(\mathbf{x}_j\mathbf{x}_a) \\ &= \sigma_{ia}\sigma_{jb} + \sigma_{ib}\sigma_{ja}. \end{aligned} \tag{A7}$$

Thus, by assumption that $\#\{\sigma_{ij} \neq 0, 1 \le i < j \le p\} = O(pc(p))$, it follows that

$$\#\{Cov(\mathbf{x}_{1i}\mathbf{x}_{1j}, \mathbf{x}_{1s}\mathbf{x}_{1t}) \neq 0, i < j, s < t\} = O(p^2c(p)^2).$$

Thus,

$$Var(\sum_{(i,j)\in\aleph_1} \pi_{ij}/\sqrt{\#\aleph_1}) = O(p^2c(p)^2/p^2) = O(c(p)^2).$$

This completes the proof of (A6). □

Appendix E. Proof of Corollary 2

Proof. By (A7) and the assumption that $\sigma_{ij} = \varsigma_{|i-j|}$, we have

$$|Cov(\pi_{ij}, \pi_{ab})| = O(\varsigma_{|i-a|}\varsigma_{|j-b|} + \varsigma_{|i-b|}\varsigma_{|j-a|}). \tag{A8}$$

For (i,j) and any $k > 0$, define

$$\mathcal{A}_k = \{(a,b) : |i-a| = k \text{ and } |j-b| \le k\} \cup \{(a,b) : |i-a| \le k \text{ and } |j-b| = k\}.$$

It is easy to see that $\min(|i-a| + |j-b|, |i-b| + |j-a|\} \ge k$ for any $(a,b) \in \mathcal{A}_k$ and $\#\mathcal{A}_k = 8k$. Thus,

$$\varsigma_{|i-a|}\varsigma_{|j-b|} + \varsigma_{|i-b|}\varsigma_{|j-a|} = O(c^k), \text{ for any } (a,b) \in \mathcal{A}_k,$$

and

$$\sum_{(a,b)} |Cov(\pi_{ij}, \pi_{ab})| = \sum_{k=1}^{p} \sum_{(a,b)\in\mathcal{A}_k} |Cov(\pi_{ij}, \pi_{ab})| = \sum_{k=1}^{p} 8kO(c^k) \le C_0$$

where C_0 is a constant. Moreover, for any set $\mathcal{S}$,

$$Var(\sum_{(a,b)\in\mathcal{S}} \pi_{ab}) \le C_0 \#\mathcal{S}. \tag{A9}$$

Let $\Pi_\ell = \mathbf{x}_1\mathbf{x}_{1+\ell} + ... + \mathbf{x}_{p-\ell}\mathbf{x}_p$. In a similar calculation, we can show that

$$Var(\Pi_\ell) = O(p-\ell).$$

Thus, we have

$$|\tilde{\varsigma}_\ell - \varsigma_\ell| = O\Big(\sqrt{\frac{\log(p\vee n)}{(p-\ell)n}}\Big).$$

Let $I = \{k : |\varsigma_k - \varsigma_\ell| > D_n \text{ for all } \ell \neq k\}$, i.e., I is the set of elements in the matrix that can be identified by D_n. It can also easily be seen that $\#I = O(\log(p))$, and thus

$$\sum_{k\in I} |\tilde{\varsigma}_k - \varsigma_k| = O(\#I\sqrt{\frac{\log(p\vee n)}{(p-\#I)n}}) = O(\sqrt{\frac{\log(p\vee n)}{n}}). \tag{A10}$$

Let $II = \{k : |\varsigma_k| \le D_n/p^2\}$, i.e., II are all the elements in the matrix that are very small by themselves. When $p\vee n$ is big enough, $I\cap II = \emptyset$ and $\#II = p - O(\log(p\vee n))$. For any $|\sigma_{ij}| \le D_n/p^2$, let

$$\tilde{\sigma}_{ij} = \frac{1}{\#\tilde{\aleph}_{ij}} \sum_{(s,t)\in\tilde{\aleph}_{ij}} \hat{\sigma}_{st}, \text{ and } \bar{\sigma}_{ij} = \frac{1}{\#\tilde{\aleph}_{ij}} \sum_{(a,b)\in\tilde{\aleph}_{ij}} \sigma_{ab},$$

where $\tilde{\aleph}_{ij} = \{(a,b) : |\hat{\sigma}_{ij} - \hat{\sigma}_{ab}| < D_n\}$. It is easy to see that $\#\tilde{\aleph}_{ij} = p(p - O(\log(p\vee n))$, and by (A9),

$$Var(\hat{\sigma}_{ij}) = O(1/(np^2)).$$

Thus,

$$\max_{(i,j)\in\bigcup_{k\in II}\aleph_k} |\tilde{\sigma}_{ij} - \bar{\sigma}_{ij}| = O_p(\sqrt{\frac{\log(p\vee n)}{p^2 n}}). \tag{A11}$$

On the other hand, we can see

$$\sum_{k\in II} |\varsigma_k| \le (D_n/p^2)\sum_{i=1}^{\infty} c^i = O(D_n/p^2),$$

and thus,

$$\max_{(i,j)\in\bigcup_{k\in II}\aleph_k} |\sigma_{ij} - \bar{\sigma}_{ij}| = O(D_n/p^2). \tag{A12}$$

It follows from (A11) and (A12) that

$$\max_{(i,j)\in\bigcup_{k\in II}\aleph_k} |\tilde{\sigma}_{ij} - \sigma_{ij}| = O_p(\sqrt{\frac{\log(p\vee n)}{p^2 n}}).$$

and

$$\sum_{k\in II} |\tilde{\varsigma}_k - \varsigma_k| = pO_p(\sqrt{\frac{\log(p\vee n)}{p^2 n}}) = O_p(\sqrt{\frac{\log(p\vee n)}{n}}). \tag{A13}$$

Finally, let $III = \Omega - I - II$. Because $\varsigma_k = O(c^k)$, if $(i,j) \notin I$ we must have $c^k - c^{k+1} < c_1 D_n$ for some $c_1 > 0$, where $\sigma_{ij} = \varsigma_k$. Thus, $\sigma_{ij} = O(D_n)$ for any $(i,j) \in III$. It is easy to see that $\#III = O(\log(p))$. Let $\bar{\sigma}_{ij}$ be defined similarly as above. Because $|\hat{\sigma}_{ij} - \sigma_{ij}| = O(D_n)$ and by definition

$$\sup_{(a,b)\in\aleph_{ij}} |\hat{\sigma}_{ij} - \hat{\sigma}_{ab}| \leq D_n$$

thus

$$|\bar{\sigma}_{ij} - \sigma_{ij}| < 4D_n.$$

On the other hand, we have

$$\max_{(i,j)\in\bigcup_{k\in III}\aleph_k} |\tilde{\sigma}_{ij} - \bar{\sigma}_{ij}| = O(\log(p))O_p(\sqrt{\frac{\log(p\vee n)}{(\log p)^2 n}}) = O_p(\sqrt{\frac{\log(p\vee n)}{n}}).$$

Similar to (A13), it follows from the above two equations that

$$\sum_{k\in III} |\tilde{\varsigma}_k - \varsigma_k| \leq O(\log(p)D_n). \tag{A14}$$

The lemma follows from (A1), (A10), (A13), and (A14). □

Appendix F. Proof of Corollary 3

Proof. It is easy to see that in Σ there are at most $M(M+3)/2$ different values, which is fixed as $p \to \infty$. By Lemma 2, these elements can be consistently identified and be estimated with root-n consistency, i.e., for any $\delta > 0$, we can find a subset of probability space $S_n : Pr(S_n) > 1 - \delta$ on which

$$||\tilde{\Lambda}_M - \Lambda_M|| = O_P(n^{-1/2}), \quad ||\tilde{\Phi}_M - \Phi_M|| = O_P(n^{-1/2}).$$

where $\tilde{\Lambda}_M = \text{diag}(\hat{\lambda}_1, ..., \hat{\lambda}_M)$ and that

$$\tilde{\Sigma} = \tilde{\Lambda}_M \otimes I_m + \tilde{\Phi}_M \otimes \Xi_m.$$

Let $\hat{\tau}_1, ..., \hat{\tau}_M$ be the M eigenvalues of $\tilde{\Phi}_M$, with corresponding eigenvectors $\hat{\mu}_1, ..., \hat{\mu}_M$, respectively. Then, it is easy to check that the eigenvalues of $\tilde{\Phi}_M \otimes \Xi_m$ are $m\hat{\tau}_1, ..., m\hat{\tau}_M$ and 0 of $(m-1) * M$ replications. Because I_m is the identity matrix with the same dimension as Ξ_m, the eigenvalues for $\tilde{\Lambda}_M \otimes I_m + \tilde{\Phi}_M \otimes \Xi_m$ are $\hat{\lambda}_1 + m\hat{\tau}_1, ..., \hat{\tau}_m + m\hat{\lambda}_M$ and $\hat{\lambda}_1$ of $(m-1)$ replications, ..., and $\hat{\lambda}_m$ of $(m-1)$ replications.

Let the M eigenvectors of Φ_M be $\mu_1, ..., \mu_M$, respectively. Note that the first eigenvectors of Ξ_m are $\nu_1 = (1, ..., 1)^\top / \sqrt{m}$. Thus, the first M eigenvectors of $\tilde{\Lambda}_M \otimes I_m + \tilde{\Phi}_M \otimes \Xi_m$ are $\hat{\ell}_M = \hat{\mu}_k \otimes \nu_1, k = 1, ..., M$ and that of $\Lambda_M \otimes I_m + \Phi_M \otimes \Xi_m$ are $\ell_k = \mu_k \otimes \nu_1, k = 1, ..., M$, respectively. It is easy to verify that

$$||\hat{\ell}_k - \ell_k|| = ||\hat{\mu}_k \otimes \nu_1 - \mu_k \otimes \nu_1|| = ||\hat{\mu}_k - \mu_k|| \times ||\nu_1|| = ||\hat{\mu}_k - \mu_k|| = O_P(n^{-1/2})$$

for $k = 1, ..., M$. □

References

1. Ledoit, O.; Wolf, M. The power of (non-) linear shrinking: A review and guide to covariance matrix estimation. *J. Financ. Econ.* **2022**, *20*, 187–218.
2. Zeileis, A. Econometric Computing with HC and HAC Covariance Matrix Estimators. *J. Stat. Softw.* **2004**, *11*, 1–17.
3. Ledoit, O.; Wolf, M. Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Rev. Financ. Stud.* **2017**, *30*, 4349–4388.
4. Schäfer, J.; Strimmer, K. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol.* **2005**, *4*, 1–32.
5. Tipping, M.E.; Bishop, C.M. Probabilistic principal component analysis. *J. R. Stat. Soc. Ser. B Stat. Method* **1999**, *61*, 611–622.
6. Fan, J.; Fan, Y.; Lv, J. High dimensional covariance matrix estimation using a factor model. *J. Econom.* **2008**, *147*, 186–197.

7. Driscoll, J.C.; Kraay, A.C. Consistent covariance matrix estimation with spatially dependent panel data. *Rev. Econ. Stat.* **1998**, *80*, 549–560.
8. Pourahmadi, M. Maximum likelihood estimation of generalised linear models for multivariate normal covariance matrix. *Biometrika* **2000**, *87*, 425–435.
9. Ledoit, O.; Wolf, M. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empir. Financ.* **2003**, *10*, 603–621.
10. Ledoit, O.; Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* **2004**, *88*, 365–411.
11. Pepler, P.T.; Uys, D.W.; Nel, D.G. Regularized covariance matrix estimation under the common principal components model. *Commun. Stat. Simul. Comput.* **2018**, *47*, 631–643.
12. Bickel, P.J.; Levina, E. Regularized estimation of large covariance matrices. *Ann. Stat.* **2008**, *36*, 199–227.
13. El Karoui, N. Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Stat.* **2008**, *36*, 2717–2756.
14. Rothman, A.J.; Levina, E.; Zhu, J. Generalized thresholding of large covariance matrices. *J. Am. Stat. Assoc.* **2009**, *104*, 177–186.
15. Cai, T.; Liu, W. Adaptive thresholding for sparse covariance matrix estimation. *J. Am. Stat. Assoc.* **2011**, *106*, 672–684.
16. Cai, T.T.; Yuan, M. Adaptive covariance matrix estimation through block thresholding. *Ann. Stat.* **2012**, *40*, 2014–2042.
17. Lam, C.; Fan, J. Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Stat.* **2009**, *37*, 4254–4278.
18. Rothman, A.J. Positive definite estimators of large covariance matrices. *Biometrika* **2012**, *99*, 733–740.
19. Fan, J.; Liao, Y.; Mincheva, M. Large covariance estimation by thresholding principal orthogonal complements. *J. R. Stat. Soc. Ser. B Stat. Method.* **2013**, *75*, 603–680.
20. Onnela, J.P.; Chakraborti, A.; Kaski, K.; Kertesz, J.; Kanto, A. Dynamics of market correlations: Taxonomy and portfolio analysis. *Phys. Rev. E* **2003**, *68*, 056110.
21. Matteson, D.; Tsay, R.S. *Multivariate Volatility Modeling: Brief Review and a New Approach*; Manuscript, Booth School of Business, University of Chicago: Chicago, IL, USA, 2011.
22. Qian, J. Shrinkage Estimation of Nonlinear Models and Covariance Matrix. Doctoral Thesis, National University of Singapore, Department of Statistics and Applied Probability, 2012. Available online: https://core.ac.uk/download/pdf/48656486.pdf (accessed on 27 March 2024).
23. Jiang, H.; Saart, P.W.; Xia, Y. Asymmetric conditional correlations in stock returns. *Ann. Appl. Stat.* **2016**, *10*, 989–1018.
24. Elton, E.J.; Gruber, M.J. Estimating the dependence structure of share prices–implications for portfolio selection. *J. Financ.* **1973**, *28*, 1203–1232.
25. Liu, S. Matrix results on the Khatri-Rao and Tracy-Singh products. *Linear Algebra Appl.* **1999**, *289*, 267–277.
26. Bollerslev, T. Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Rev. Econ. Stat.* **1990**, *72*, 498–505.
27. Hansen, P.R.; Lunde, A. A forecast comparison of volatility models: does anything beat a GARCH(1,1)? *J. Appl. Econ.* **2005**, *20*, 873–889.
28. Aghamohammadi, A.; Dadashi, H.; Sojoudi, M.; Sojoudi, M.; Tavoosi, M. Optimal portfolio selection using quantile and composite quantile regression models. *Commun. Stat. Simul. Comput.* **2022**, 1–11. https://doi.org/10.1080/03610918.2022.2094961.
29. Zhang, Z.; Yue, M.; Huang, L.; Wang, Q.; Yang, B. Large portfolio allocation based on high-dimensional regression and Kendall's Tau. *Commun. Stat. Simul. Comput.* **2023**, 1–13. https://doi.org/10.1080/03610918.2023.2245176.
30. Golub, G.H.; Van Loan, C.F. *Matrix Computations*; Johns Hopkins University Press: Baltimore, MD, USA, 2013.