ABSTRACT

Title of Dissertation: NEW STATISTICAL METHODS FOR HIGH-DIMENSIONAL INTERCONNECTED DATA WITH UNIFORM BLOCKS

Yifan Yang
Doctor of Philosophy, 2023

Dissertation Directed by: Professor Shuo Chen
Department of Mathematics

Empirical analyses of high-dimensional biomedical data, including genomics, proteomics, microbiome, and neuroimaging data, consistently reveal the presence of strong modularity in the dependence patterns. In these analyses, highly correlated features often form a few distinct communities or modules, which can be interconnected with each other. While the interconnected community structure has been extensively studied in biomedical research (e.g., gene co-expression networks), its potential to assist in statistical modeling and inference remains largely unexplored. To address this research gap, we propose novel statistical models and methods that capitalize on the prevalent community structures observed in large covariance and precision matrices derived from high-dimensional biomedical interconnected data.

The first objective of this dissertation is to delve into the algebraic properties of the proposed interconnected community structures at the population level. Specifically, this pattern partitions the population covariance matrix into uniform (i.e., equal variances and covariances)

blocks. To accomplish this objective, we introduce a block Hadamard product representation in Chapter 2, which relies on two lower-dimensional "coordinate" matrices and a pre-specific vector. This representation enables the explicit expressions of the square or power, determinant, inverse, eigendecomposition, canonical form, and the other matrix functions of the original larger-dimensional matrix on the basis of these lower-dimensional "coordinate" matrices.

Estimating a covariance matrix is central to high-dimensional data analysis. Our second objective is to consistently estimate a large covariance or precision matrix having an interconnected community structure with uniform blocks. In Chapter 3, we derive the best-unbiased estimators for covariance and precision matrices in closed forms and provide theoretical results on their asymptotic properties. Our proposed method improves the accuracy of covariance and precision matrix estimation and demonstrates superior performance compared to the competing methods in both simulations and real data analyses.

In Chapter 4, our goal is to investigate the effects of alcohol intake (as an exposure) on metabolomics outcome features. However, similar to other omics data, metabolomic outcomes often consist of numerous features that exhibit a structured dependence pattern, such as a co-expression network with interconnected modules. Effectively addressing this dependence structure is crucial for accurate statistical inferences and the identification of alcohol intake-related metabolomic outcomes. Nevertheless, incorporating the structured dependence patterns into multivariate outcome regression models remains difficulties in accurate estimation and inference. To bridge this gap, we propose a novel multivariate regression model that accounts for the correlations among outcome features using a network structure composed of interconnected modules. Additionally, we derive closed-form estimators of regression parameters and provide inference tools. Extensive simulation analysis demonstrates that our approach yields much-improved sen-

sitivity with a well-controlled discovery rate when benchmarking against existing multivariate regression models.

Confirmatory factor analysis (CFA) models play a crucial role in revealing underlying latent common factors within sets of correlated variables. However, their implementation often relies on a strong prior theory to categorize variables into distinct classes, which is frequently unavailable (e.g., in omics data analysis scenarios). To address this limitation, in Chapter 5, we propose a novel strategy based on network analysis that allows data-driven discovery to substitute for the lacking prior theory. By leveraging the detected interconnected community structure, our approach offers an elegant statistical interpretation and yields closed-form uniformly minimum variance unbiased estimators for all unknown matrices. To evaluate the effectiveness of our proposed estimation procedure, we compare it to conventional numerical methods and thoroughly validate it through extensive Monte Carlo simulations and real-world applications.

NEW STATISTICAL METHODS FOR HIGH-DIMENSIONAL INTERCONNECTED DATA WITH UNIFORM BLOCKS

by

Yifan Yang

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Professor Shuo Chen, Chair/Advisor
Professor Benjamin Kedem, Co-Chair
Professor Vince Lyzinski
Professor Tianzhou Ma
Professor Xin He

Acknowledgments

I would like to express my heartfelt gratitude to Dr. Shuo Chen, my doctoral research advisor, for his invaluable guidance throughout this journey at University of Maryland, College Park. He is always available for academic advice whenever the students need it. Dr. Chen not only taught me the foundations of research, academic writing, and data analysis in real-world applications but also nurtured my collaborative skills. His insightful suggestions and unwavering support have been instrumental in shaping my future career.

I extend my sincere appreciation to all the professors and staff in the Department of Mathematics at University of Maryland, College Park. In particular, I am grateful to Dr. Benjamin Kedem for his exceptional expertise in mathematical statistics, time series analysis, Bayesian statistics, and various other fields. Dr. Paul Smith's teachings on linear models, along with his guidance during numerous fruitful RITs, have been immensely valuable. I would also like to thank Dr. Eric Slud for his mentorship in computational statistics and categorical data analysis. Dr. Vince Lyzinski, an expert in graph theory and a member of my dissertation committee, deserves special thanks. Dr. Takumi Saegusa's instruction on survival analysis and other remarkable statistical methods in semiparametric models has been invaluable. I am indebted to Cristina Garcia for her constant kindness and careful planning for the students, as well as to Stephanie Padgett for assisting us in reserving rooms. I am grateful to Matthew Griffin, the instructor of STAT 100, for his collaboration and for imparting valuable teaching techniques. Working with

Susan Mazzullo, a dedicated instructor for several undergraduate statistical classes, has been a pleasure.

I am also grateful to the faculty members in other departments who have contributed to my academic journey. Dr. Chixiang Chen deserves thanks for proofreading my papers and providing invaluable suggestions. Dr. Tianzhou Ma's leadership in bi-weekly group meetings has provided fresh insights and perspectives. Dr. Xin He, a member of my dissertation committee, has been unfailingly kind and supportive.

I am deeply thankful to my friends in Maryland, including Chuan Bi, Yunjiang Ge, Hwiyoung Lee, Tong Lu, Yezhi Pan, Qiong Wu, Zhenyao Ye, Nathan Yu, Yiran Zhang, Xiaocui Zhang, Xuze Zhang, Zhiwei Zhao, Xiaoyu Zhou, and more.

Lastly, I would like to express my profound appreciation to my parents, Guangping Yang and Xiaohui Huang. Thank you for introducing me to this remarkable and beautiful world, enabling me to discover my passion and complete this degree. Your love and support have been my driving force.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Estimation of Large Covariance Matrix

Technological innovation has had a profound impact on scientific discoveries by integrating the analysis of large-scale datasets into various fields. Statistically, a covariance matrix or the precision matrix (i.e., the inverse of a covariance matrix) is crucial in understanding the intricate relationships among massive and dependent covariates. Thus, considerable attention has been concentrated on the use of an appropriate covariance (or precision or correlation) matrix estimator. A practical issue arises from the fact that the sample size n in research is often inadequate in comparison to the number of covariates, say p. Thus, when the number of covariates exceeds the sample size, i.e., $p > n$, the sample covariance matrix, which is a natural choice of the covariance matrix estimator in the case $p < n$, fails to be a reliable estimator due to the absence of positive definiteness (Dykstra, 1970) and the inconsistency of eigenvalues and eigenvectors (Johnstone, 2001; Johnstone and Lu, 2009; Johnstone and Paul, 2018). Consequently, Fan (2005) asserted that it is essentially challenging to obtain an appropriate covariance estimator without imposing a structure assumption on a large covariance matrix.

1.2 Classical Covariance Structures

Focus has been devoted to structured or patterned matrices, not only because they hold significance in theoretical analyses for multivariate models but also because they serve as fundamental tools in numerous applications. Many notable instances of structured matrices have found applications in multivariate analysis.

One straightforward example involves testing a *given* (known) covariance matrix statistically based on multivariate normal observations (Korin, 1968). Another example is the adoption of a *sphericity* structure by Mauchly (1940), where the diagonals are identical and the off-diagonals are zero. Wilks (1946) then proposed an *intraclass*, or called *uniform*, or *complete symmetry* structure, characterized by equal diagonals and equal off-diagonals. Wilks developed likelihood ratio tests (LRTs) to statistically evaluate hypotheses related to population means or population covariance structures. The estimation and hypothesis testing problems under the uniform covariance structures were further explored by Roy and Murthy (1960); Geisser (1964); Aitkin et al. (1968); Haq (1974); Mathai and Katiyar (1979); Clement et al. (1981); Bhoj (1987), and others. Building upon Wilks' uniform structure, Votaw (1948) introduced two types of *compound symmetry* structures, wherein the blocks exhibit uniform structures and derived the LRTs for the corresponding hypotheses. The exact distributions of the LRTs were examined by Consul (1968) and Mathai and Rathie (1970). From a symmetry perspective, matrices with *Toeplitz* or *circular Toeplitz* structures have proven essential in various applications, including physics, mathematics, and signal processing (Olkin and Press, 1969; Olkin, 1972).

Many classical covariance structures can be generalized to *linear structured* covariance matrices, wherein only a few covariance parameters are unknown (Srivastava, 1966; Anderson,

1969; Anderson et al., 1970; Anderson, 1973). Furthermore, *reducible* and *totally reducible* matrices were investigated by Rogers and Young (1975), Young (1976), and Sinha and Wieand (1979).

In addition to parameterizing the covariance matrix unknown up to serval parameters, another approach to design the covariance matrix is by adopting *blocks* in the covariance structure. For instance, Afifi and Elashoff (1969) partitioned the covariance matrix based on the dichotomous and continuous variables. The block versions of Wilks' uniform structure were discussed by Fleiss (1966), Arnold (1973), and Rogers and Young (1974), while the block forms of Votaw's compound symmetry structures were extended by Szatrowski (1982) and Žežula et al. (2018).

1.3 Recent Developments of Covariance Structures

Recent statistical methods have made successful advancements in addressing the estimation problem associated with high-dimensional structured covariance matrices within the asymptotic framework where the number of covariates p grows with the sample size n together. For example, a *shrinkage* method involves shrinking the sample covariance matrix to the identity matrix (Ledoit and Wolf, 2004). Several *penalty* methods have also been established to estimate large-scale covariance and precision matrices (Friedman et al., 2008; Rothman et al., 2008; Lam and Fan, 2009; Ravikumar et al., 2011). Under the assumption of *bandability*, i.e., the entries in the covariance matrix decrease to zero from the diagonal to the off-diagonal direction, the banding or (its smooth version) tapering covariance estimator has been proposed (Wu and Pourahmadi, 2003; Bickel and Levina, 2008a; Bickel and Gel, 2011). Additionally, various versions of *thresholding* covariance estimators, along with their theoretical properties such as optimal convergence

rates under different matrix norms, have been studied (Karoui, 2008; Bickel and Levina, 2008b; Rothman et al., 2009; Cai and Liu, 2011; Cai and Yuan, 2012).

However, the sparsity assumption may be violated in many real-life applications. For instance, the financial returns and house prices studies (Fan et al., 2016), and the interactive features in biomedical studies (Chen et al., 2018). The *spiked sparse* covariance matrix has been explored by Johnstone (2001). Alternatively, the *conditional sparsity* assumption, i.e., a structure assumption combined with sparsity and low-rank approximation, has been proposed, and the large covariance matrices have been consistently estimated using *multi-factor* models (Fan et al., 2008, 2013, 2018). Under another non-sparsity assumption, i.e., the Toeplitz structure (a special case of bandable covariance structure), Cai et al. (2013) and Cai et al. (2016) have studied the problem of estimating large covariance matrices and optimal convergence rates. Expository literature on structured covariance matrix estimation can be found in Pourahmadi (2013), Wainwright (2019), and Fan et al. (2020).

1.4 Interconnected Community Structure

The accumulating availability of community detection, network division, or clustering techniques has led to enormous scientific discoveries and challenges in many studies involving large-scale networks, such as biologics, biomedicine, plant science, computer science, finance, social networks (Newman, 2006). Characterizing the structures or latent patterns in networked systems is crucial as they may quantitatively describe the complex interactions among high-dimensional features or variables (Wu et al., 2021). Novel discoveries based on structured networks could deepen our understanding of the scientific mechanisms behind them, see various real examples in

He et al. (2015), He et al. (2019), Pal et al. (2020), and Perrot-Dockès et al. (2022). Conventional community detection methods typically impose an *independence* assumption that the latent communities are not correlated, resulting in *block-diagonal* structures in networks, particularly when the number of features in networks is extremely high (Zhao, 2017; Lee and Wilkinson, 2019). However, modeling the networks with independent communities may mislead subsequent statistical analysis due to oversimplification. Alternatively, an *interconnected community* structure is much more flexible for network analysis by allowing non-null connections among features at the community level. In other words, we may add non-null off-diagonal blocks to the structure of networks to represent correlations among communities. Wu et al. (2021) examined the relationships between the independent communities and interconnected communities, as well as between overlapped communities and interconnected communities.

1.5 A New Interconnected Community Structure with Uniform Blocks

In the present dissertation, we propose an interconnected community structure that exists in a wide range of high-dimensional datasets. This structure can be implemented in population covariance matrices, correlation matrices, or weighted adjacency matrices in various fields, including genetics, proteomics, brain imaging, and RNA expression data, among many others.

From a sample perspective, an *interconnected community structure with well-organized blocks* is widely observed. It exhibits several characteristics: (1) it is *latent*, meaning that network or community detection algorithms need to be employed to analyze the raw data in a preliminary study; (2) it is *non-sparse*, i.e., the elements of the covariance or correlation matrix have small but non-zero values; (3) it demonstrates an almost constant-valued block form, where the elements

within each block are nearly identical, exhibiting low variability; and (4) it may contain *singletons* or *isolated nodes*. When the singletons are identified, we refer to the proposed structure as an *interconnected community structure with generalized well-organized blocks*. In other words, we can form an interconnected community structure with well-organized blocks from a generalized version by extracting the features that are reordered by community detection algorithms and excluding the singletons. The population versions of the interconnected community structures with (generalized) well-organized blocks are referred to as the *interconnected community structure with (generalized) uniform blocks*.

Yeast Genome Study. One example of the interconnected community structure with well-organized blocks (without singletons) is observed in the yeast genome study (Spellman et al., 1998). We plot a heat map of the sample correlation matrix for 724 selected genes in Figure 1.1(B), where a K-medoids clustering algorithm was applied to the raw dataset, see Figure 1.1(A). Figure 1.1(C) represents the population correlation matrix may have 8 by 8 uniform blocks.

Proteomics Study. Another example is from a proteomics study (Yildiz et al., 2007) in Figure 1.1(D). The sample correlation matrix is analyzed, revealing an interconnected community structure with generalized well-organized blocks among 184 protein features using the NICE algorithm (Chen et al., 2018) in Figure 1.1(E). Extracting the structure without singletons results in a sample correlation matrix with 107 protein features, shown in Figure 1.1(F). It suggests a population correlation matrix with the corresponding uniform-block structure in Figure 1.1(G).

Seed Quality Study. A 923 by 923 sample correlation matrix is calculated from raw seed quality data (Perrot-Dockès et al., 2022), plotted in Figure 1.1(H). An interconnected community structure (without singletons), as shown in Figure 1.1(I) in the sample correlation matrix, is

provided by a hierarchical clustering algorithm. This implies a population correlation matrix with the corresponding 7 by 7 uniform-block structure. See the illustration in Figure 1.1(J).

NMR Study. We observe a 249 by 249 sample correlation matrix calculated from a subset of raw NMR data (Ritchie et al., 2023) in Figure 1.1(K). An interconnected community structure with generalized well-organized blocks is revealed in the sample correlation matrix by a community detection algorithm (Chen et al., 2023). See Figure 1.1(L). A 170 by 170 sample correlation matrix (of residuals) extracted in Figure 1.1(M), which implies a population correlation matrix with the corresponding 5 by 5 uniform-block structure in Figure 1.1(N).

Exposome and Metabolites Study. Another real-data example of the pattern of well-organized blocks in sample correlation matrices is found in a study involving the exposome and metabolites (ISGlobal, 2021). Figures 1.1(O) and (S) are the heat maps of the correlation matrices among exposome and metabolite features, respectively. In this study, there are 169 exposome variables and 221 metabolite variables for 1192 subjects. In particular, Figure 1.1(P) contains the heat map of the sample correlation matrix calculated by the clustering method (Wu et al., 2021) for exposome variables, while Figure 1.1(T) contains the heat map of the sample correlation matrix calculated by the clustering method for metabolite variables. Omitting the singletons, 89 out of 169 exposome variables are identified to form a 9 by 9 well-organized blocks (Figure 1.1(Q)) and 141 out of 221 metabolite variables (Figure 1.1(U)) are identified to form a 7 by 7 well-organized blocks. We may assume that their population correlation matrices have block forms, as shown in Figure 1.1(R) and (V), respectively.

EPSI Study. The presence of an interconnected community pattern is detected in the sample correlation matrix for an echo-planar spectroscopic imaging (EPSI) dataset, including 445 combinations of neurometabolites and regions-of-interest (ROIs) detected by the NICE algorithm

(Chen et al., 2018) and shown in Figure 1.1(W) and (X). Among these 445 combinations, the sample correlation matrix of 227 combinations is extracted and implies the underlying population correlation matrix has the structure of uniform blocks in Figure 1.1(Y) and (Z).

Other Studies. In addition to the above studies, analogs of the interconnected community pattern were uncovered in a cutaneous melanoma dataset from the genome-wide association study (He et al., 2015), a neuroimaging activation and connectivity dataset, a DNA methylation dataset (Chen et al., 2016), and a gene expression profiling dataset from the host peripheral blood study (Chen et al., 2018). Recently, it was extracted from a multiple myeloma dataset (He et al., 2019), a citation network study (Pal et al., 2020), and an RNA-seq dataset from the cancer genome atlas study of acute myeloid leukemia (Wu et al., 2021).

Throughout the current research, we concentrate on the population covariance matrices having interconnected community structures with uniform blocks, as illustrated in Figure 1.1(C), (G), (J), (N), (R), (V), and (Z). The proposed interconnected community structure with uniform blocks has two-fold advantage. On the one hand, it leads to a dramatically reduced number of covariance parameters (from $p(p+1)/2$ to $K+(K+1)K/2$ where K is the number of diagonal blocks) but remains adequate flexibility to deal with the arbitrary dependency between covariates. On the other hand, the non-sparse uniform correlations between blocks might represent some stability (e.g., with respect to time or location in applications (Votaw et al., 1950)) and symmetry in experimental quantities, therefore, improving the interpretations.

1.6 Organization

The rest of the dissertation is organized as follows. The definitions and properties of uniform-block matrices and their applications in hypothesis tests are presented in Chapter 2. An estimation procedure for the covariance matrix that is assumed to have the uniform-block structure is conducted under both low- and high-dimensional frameworks in Chapter 3. Multivariate outcomes having an interconnected community structure with uniform blocks are analyzed by using autoregressive regression models in Chapter 4. A semi-confirmatory factor model with uniform-block interconnected community structure is studied and an estimation procedure is proposed for the unknown matrices and factor scores in Chapter 5. All technical proofs and supplementary materials are appended at the end of the dissertation.

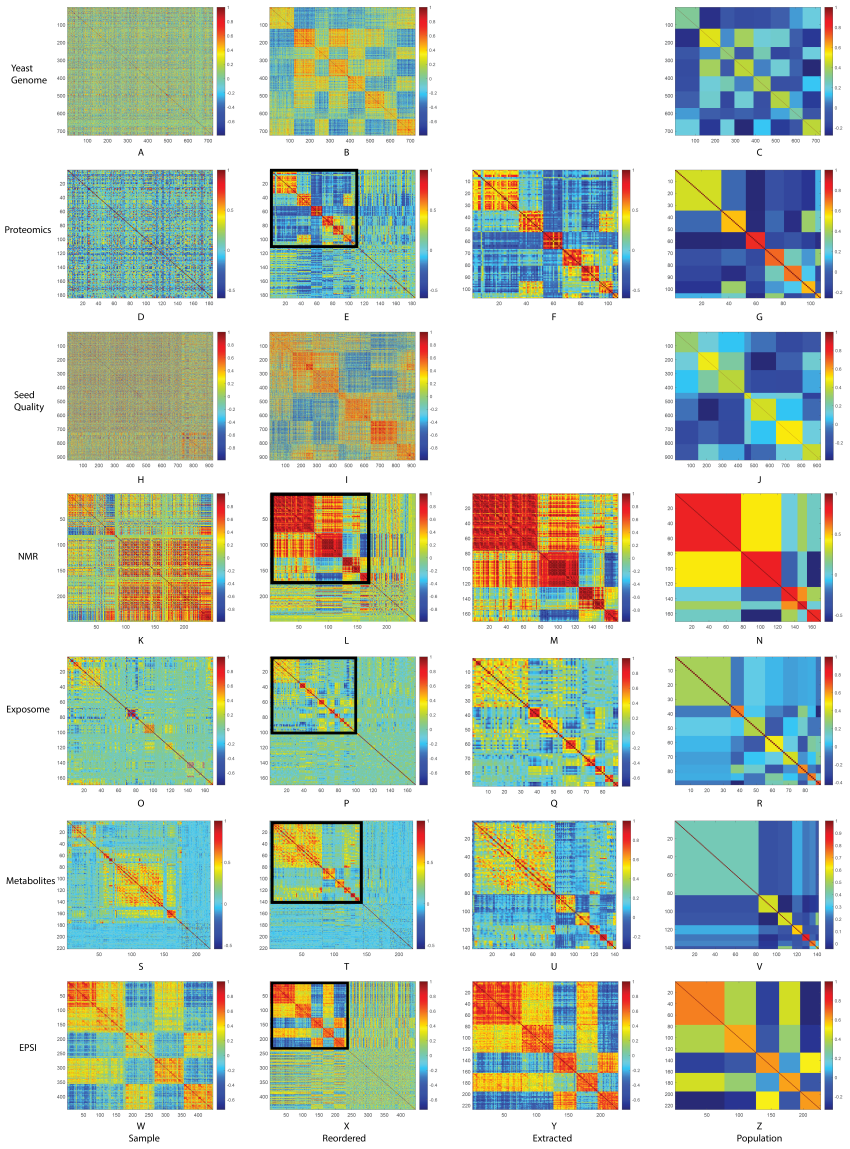


Figure 1.1: The plots in the first column are the heat maps of sample correlation matrices for the raw datasets. The plots in the second column are the heat maps of sample correlation matrices with reordered features/variables using community detection algorithms. We call these structures interconnected community structures with (generalized) well-organized blocks (if the singletons exist). The plots in the third column are the heat maps of sample correlation matrices with features/variables that are extracted from the reordered ones. We call these structures interconnected community structures with well-organized blocks. The plots in the fourth column are the heat maps of population correlation matrices that are assumed to have the interconnected community structures with uniform blocks.

Chapter 2: A New Representation of Uniform-Block Matrix and Applications

2.1 Introduction

Covariance matrices with specific structures or patterns have been extensively studied for their crucial roles in theoretical and practical applications of multivariate analysis. Numerous examples of structured covariance matrices have been employed in multivariate analysis. Mauchly (1940) proposed a *spherical* covariance matrix with identical positive variance parameters along the diagonal and zero correlation parameters off the diagonal. Wilks (1946) extended the sphericity structure to have equal non-zero values for the off-diagonal correlation parameters, terming it the *uniform* (*intraclass* or *complete symmetry*) structure in the application to parallel forms of a test in educational studies. Furthermore, Votaw (1948) expanded Wilks' complete symmetry structure by incorporating the interchangeability of mutually exclusive subsets of variables, introducing two types of *compound symmetry* covariance structures, which were utilized in medical experiments (Votaw et al., 1950). In addition to the spherical and intraclass symmetric structures, Olkin and Press (1969) proposed another covariance structure known as *circular symmetry*, which was applied in physical studies and time series analysis.

A number of technological breakthroughs have led to significantly large-dimensional variables in real-world practice, necessitating the consideration of more complex covariance structures to reduce dimensionality. Customarily, various covariance structures have been developed,

including the *bandability* (Wu and Pourahmadi, 2003; Bickel and Levina, 2008a), the *sparsity* (Karoui, 2008; Bickel and Levina, 2008b; Cai and Liu, 2011), and the combination of sparsity and *low-rank* (Fan et al., 2008, 2011). Alternatively, to address the high dimensionality problem, covariance matrices can be assumed to have a block structure, where the number of unknown parameters is remarkably smaller than the original dimension. For instance, Rogers and Young (1974) generalized Wilks' intraclass structure to an arbitrary order in an educational study, such that all diagonal blocks have the same intraclass form, as do all off-diagonal blocks. Szatrowski (1976) studied covariance matrices with *block compound symmetry* structures, including type I and type II, and applied them to the analysis of educational testing data (Szatrowski, 1982). Olkin (1972) introduced circular symmetry structures in blocks and proposed a more general structure known as *block circular symmetry* for applications in physics. Roy and Leiva (2011), Roy et al. (2015), Roy et al. (2016), and Žežula et al. (2018) have extensively investigated a block structure referred to as *blocked compound symmetry* or *equicorrelation (partition)* (Leiva, 2007; Roy and Leiva, 2008), and applied it in brain imaging and bone densitometry studies. In this chapter, our focus is on a covariance or correlation matrix with a particular block structure that is commonly observed in empirical applications.

We concentrate on investigating a specific block pattern called *uniform-block* (UB) structure, motivated by its numerous real-world applications in Chapter 1. Specifically, the UB structure is characterized by diagonal and off-diagonal elements within each diagonal submatrix being equal to two constants and all elements within each off-diagonal submatrix being equal to a constant. A partitioned matrix having a uniform-block structure is denoted as a *uniform-block* matrix. The concept of UB structures is not completely new and has been introduced by various researchers in different contexts. For instance, Geisser (1963) referred to it as the *uniform case*

of order m and derived an information test statistic for the population mean vector, when the covariance matrix of a normal population has a UB structure of order $m = 1$ or 2. Morrison (1972) extended Geisser's information test statistic to a more general order. Huang and Yang (2010) investigated the random sampling issues in the presence of a UB structure in the correlation matrix. Cadima et al. (2010) referred to it as a k*-group block structure* and studied the eigendecomposition of correlation matrices with a UB structure. Roustant and Deville (2017) named a correlation matrix with UB structure a *parametric block correlation matrix with* p *blocks*, and provided necessary and sufficient conditions for its positive definiteness. Roustant et al. (2020) investigated the Gaussian process regression problems using the name of *generalized compound symmetry block covariance matrices* for UB matrices. Recently, Archakov and Hansen (2022) examined this structured matrix, referring to it as a *block matrix with block partition*, and provided canonical forms for both symmetric and nonsymmetric cases.

However, to the best of our knowledge, there have been limited comprehensive studies on the algebraic properties of UB matrices, which restricts their applications in various fields, including statistics, biometrics, economics, finance, and others. For example, Geisser (1963) was the first to derive the null distributions of the information test statistics concerning the population mean vector(s) for both single and multiple samples, given a covariance matrix with a 2 by 2 UB structure. Specifically, Geisser (1963) derived an analogous version of Hotelling's (generalized) T^2-statistic regarding the population mean vector based on a single normal sample, and an analogous version of Hotelling's (generalized) T_0^2-statistic, known also as the Hotelling-Lawley trace (Lawley, 1938; Hotelling, 1947, 1951), for testing the equality of population mean vectors based on multiple normal samples. Under the null hypotheses, both of Geisser's information test statistics follow identical distributions as linear combinations of independent F-variates. Al-

though Geisser (1963) and Morrison (1972) also extended these results to a general case with an arbitrary number of diagonal blocks, proofs were omitted.

In this study, we presented the algebraic properties of UB matrices through a novel block Hadamard product representation. In essence, given a vector consisting of the block sizes, a UB matrix can be uniquely determined by a diagonal matrix and a symmetric matrix of much smaller dimensions. Moreover, these two lower-dimensional matrices (and the block-size vector) can be viewed as the "coordinates" of a UB matrix since many important algebraic calculations on UB matrices only depend on their "coordinate" matrices. As a result, this representation greatly simplifies the algebraic operations, including the power computation, inverse calculation, eigenvalues determination, and determinant evaluation of a UB matrix, by leveraging its "coordinate" matrices. As an application in statistics, we revisited and rigorously established the exact null distributions of Geisser's information test statistics for a general number of orders, including single and multiple sample cases.

We organize the remainder of this chapter as follows. Section 2.2 presents the definitions and properties of UB matrices. Section 2.3 and Section 2.4 demonstrate the exact null distributions of Geisser's information test statistics for one-sample and multiple-sample cases, respectively. Lastly, we summarize our findings and provide remarks and discussions in Section 2.5. Technical proofs are given in Chapter A. Throughout this chapter, let $\top$ denote the transpose of a vector or matrix. Let $\mathbf{I}_n, \mathbf{J}_n \in \mathbb{R}^{n\times n}$ denote the identity matrix and all-one matrix, respectively. Let $\mathbf{0}_{n\times m}, \mathbf{1}_{n\times m} \in \mathbb{R}^{n\times m}$ denote the all-zero matrix and all-one matrix, respectively. Let $\mathrm{diag}(\cdot)$ and $\mathrm{Bdiag}(\cdot)$ denote the diagonal matrix and the block-diagonal matrix, respectively. Let $\mathrm{tr}(\cdot)$ and $\det(\cdot)$ denote the trace and determinant of a square matrix, respectively. Let $\mathrm{sum}(\cdot)$ denote the sum of all elements of a matrix. Let $\mathrm{corr}(\mathbf{\Sigma}) = \mathrm{diag}^{-1/2}(\sigma_{11}, \ldots, \sigma_{pp}) \times \mathbf{\Sigma} \times$

$\mathrm{diag}^{-1/2}(\sigma_{11}, \ldots, \sigma_{pp})$ denote the correlation matrix of covariance matrix $\boldsymbol{\Sigma}$ with diagonal elements $\sigma_{11}, \ldots, \sigma_{pp}$.

2.2 Uniform-Block Structure and Uniform-Block Matrix

In this section, we begin by defining a uniform-block structure and matrix. Next, we introduce a block Hadamard product representation for uniform-block matrices, which unveils their algebraic properties.

Definition 2.2.1 (partition-size vector and partitioned matrix by a partition-size vector)**.** *Given a p by p matrix $\boldsymbol{N} \in \mathbb{R}^{p\times p}$ and a positive integer $K \in \mathbb{Z}_+$ such that $K < p$, we define:*

(1) a column vector $\boldsymbol{p} = (p_1, \ldots, p_K)^\top \in \mathbb{Z}_+^K$ is a partition-size vector, if $p_k > 1$ for every k and $p = p_1 + \cdots + p_K$;

(2) the K by K partitioned matrix $(\boldsymbol{N}_{kk'})$ of $\boldsymbol{N}$ is the partitioned matrix of $\boldsymbol{N}$ by $\boldsymbol{p}$, if the (k, k')-th block $\boldsymbol{N}_{kk'}$ has dimensions p_k by $p_{k'}$ for $k, k' = 1, \ldots, K$.

Definition 2.2.2 (uniform-block structure and matrix)**.** *Given a partition-size vector $\boldsymbol{p} = (p_1, \ldots, p_K)^\top$ and the K by K partitioned matrix $(\boldsymbol{N}_{kk'})$ of a symmetric matrix $\boldsymbol{N}$ by $\boldsymbol{p}$, we define:*

(1) the structure of $(\boldsymbol{N}_{kk'})$ is a uniform-block structure, if there exist real numbers a_{kk} and $b_{kk'}$ satisfying that the diagonal block $\boldsymbol{N}_{kk} = a_{kk}\boldsymbol{I}_{p_k} + b_{kk}\boldsymbol{J}_{p_k}$ for every $k = k'$ and the off-diagonal block $\boldsymbol{N}_{kk'} = b_{kk'}\boldsymbol{1}_{p_k\times p_{k'}}$ with $b_{k'k} = b_{kk'}$ for every $k \neq k'$;

(2) the partitioned matrix $(\boldsymbol{N}_{kk'})$ is a uniform-block matrix if it has the structure of uniform-block. Furthermore, let $\boldsymbol{N}[\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p}] = (\boldsymbol{N}_{kk'})$ denote this uniform-block matrix, where $\boldsymbol{A} = \mathrm{diag}(a_{11}, \ldots, a_{KK})$ is a K by K diagonal matrix and $\boldsymbol{B} = (b_{kk'})$ is a K by K symmetric matrix with $b_{k'k} = b_{kk'}$ for every $k \neq k'$.

Remark (non-symmetric uniform-block structure and matrix)**.** We impose the condition of symmetry on a uniform-block structure or matrix, as defined in Definition 2.2.2, because we will be considering covariance or correlation matrices with this structure. However, it is worth noting that a non-symmetric uniform-block structure and matrix can also be defined by removing the condition $b_{k'k} = b_{kk'}$ for every $k \neq k'$, i.e., allowing $\mathbf{B}$ to be an arbitrary K by K matrix. Nonetheless, throughout this chapter, unless explicitly stated otherwise, we refer to a uniform-block structure or matrix as symmetric.

Following Definition 2.2.2, we introduce two important instances of UB matrices: the partitioned matrices of an identity matrix $\mathbf{I}_p$ and an all-one matrix $\mathbf{J}_p$ are UB matrices, by a predetermined partition-size vector $\boldsymbol{p} = (p_1, \ldots, p_K)^\top$. For simplicity, we will use $\mathbf{I}[\boldsymbol{p}]$ and $\mathbf{J}[\boldsymbol{p}]$ instead of $\mathbf{I}[\mathbf{I}_K, \mathbf{0}_{K\times K}, \boldsymbol{p}]$ and $\mathbf{J}[\mathbf{0}_{K\times K}, \mathbf{J}_K, \boldsymbol{p}]$ throughout the chapter.

$$\mathbf{I}[\boldsymbol{p}] = \mathbf{I}[\mathbf{I}_K, \mathbf{0}_{K\times K}, \boldsymbol{p}] = \mathbf{I}_p = \begin{pmatrix} \mathbf{I}_{p_1} & \mathbf{0}_{p_1\times p_2} & \cdots & \mathbf{0}_{p_1\times p_K} \\ \mathbf{0}_{p_2\times p_1} & \mathbf{I}_{p_2} & \cdots & \mathbf{0}_{p_2\times p_K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{p_K\times p_1} & \mathbf{0}_{p_K\times p_2} & \cdots & \mathbf{I}_{p_K} \end{pmatrix},$$

$$\mathbf{J}[\boldsymbol{p}] = \mathbf{J}[\mathbf{0}_{K\times K}, \mathbf{J}_K, \boldsymbol{p}] = \mathbf{J}_p = \begin{pmatrix} \mathbf{J}_{p_1} & \mathbf{1}_{p_1\times p_2} & \cdots & \mathbf{1}_{p_1\times p_K} \\ \mathbf{1}_{p_2\times p_1} & \mathbf{J}_{p_2} & \cdots & \mathbf{1}_{p_2\times p_K} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_{p_K\times p_1} & \mathbf{1}_{p_K\times p_2} & \cdots & \mathbf{J}_{p_K} \end{pmatrix}.$$

Using the notations $\mathbf{I}[\boldsymbol{p}]$ and $\mathbf{J}[\boldsymbol{p}]$, we propose the following novel block Hadamard product representation of a UB matrix, which extremely simplifies the algebraic calculations involving

UB matrices.

Lemma 2.2.1 (block Hadamard product representation of a UB matrix)**.** *Given a pre-determined a partition-size vector* $\boldsymbol{p} = (p_1, \ldots, p_K)^\top$*, suppose the* K *by* K *partitioned matrix* $(\boldsymbol{N}_{kk'})$ *of a* p *by* p *symmetric matrix* $\boldsymbol{N}$ *by* $\boldsymbol{p}$ *is a UB matrix* $\boldsymbol{N}[\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p}]$*, where* $\boldsymbol{A} = diag(a_{11}, \ldots, a_{KK})$ *is a diagonal matrix and* $\boldsymbol{B} = (b_{kk'})$ *is a symmetric matrix with* $b_{k'k} = b_{kk'}$ *for every* $k \neq k'$*. Then,*

$$\boldsymbol{N}[\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p}] = \boldsymbol{A} \circ \boldsymbol{I}[\boldsymbol{p}] + \boldsymbol{B} \circ \boldsymbol{J}[\boldsymbol{p}],$$

holds uniquely for $\boldsymbol{A}$ *and* $\boldsymbol{B}$*, where* $\circ$ *denotes the block Hadamard product satisfying that* $\boldsymbol{A} \circ \boldsymbol{I}[\boldsymbol{p}]$ *is the block-diagonal matrix* $\mathrm{Bdiag}(a_{11}\boldsymbol{I}_{p_1}, \ldots, a_{KK}\boldsymbol{I}_{p_K})$ *and* $\boldsymbol{B} \circ \boldsymbol{J}[\boldsymbol{p}]$ *is the symmetric block matrix* $\left(b_{kk'}\boldsymbol{I}_{p_k \times p_{k'}}\right)$*.*

Remark (block Hadamard product representation)**.** The matrix operator $\circ$ can be regarded as a specialized form of block Hadamard product that is specifically tailored for block matrices, as discussed in works (Horn et al., 1991; Günther and Klotz, 2012). We provide an illustration of Lemma 2.2.1 in Figure 2.1. A proof of Lemma 2.2.1 is available in Chapter A. From the proof, it is evident that the block Hadamard product representation holds (except for uniqueness) when $p_k = 1$ for some k. Furthermore, the representation also holds for non-symmetric uniform-block matrices.

By Lemma 2.2.1, the block Hadamard product representation is crucial for a UB matrix, as it provides an explicit expression involving the diagonal matrix $\mathbf{A}$, the symmetric matrix $\mathbf{B}$, and the partitioned-size vector $\boldsymbol{p}$. Therefore, we suggest utilizing the notation $\mathbf{N}[\mathbf{A}, \mathbf{B}, \boldsymbol{p}]$ in Definition 2.2.2 (instead using the usual notation $\mathbf{N}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$) throughout this chapter to emphasize

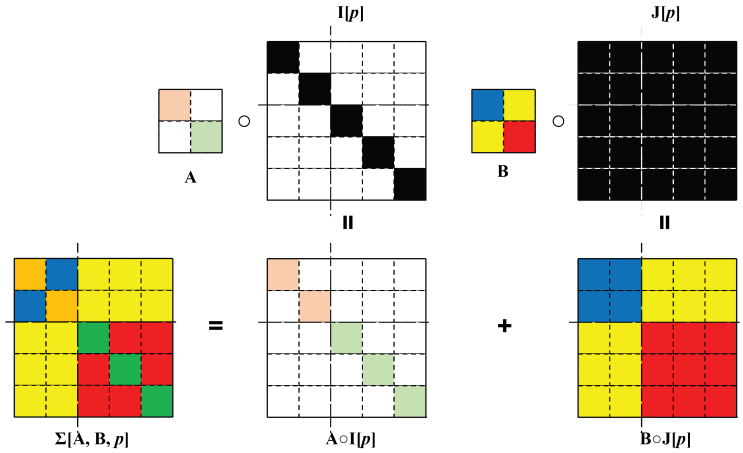


Figure 2.1: Illustration of the block Hadamard product representation of a UB matrix $\mathbf{\Sigma}[\mathbf{A},\mathbf{B},\boldsymbol{p}] = \mathbf{A}\circ\mathbf{I}[\boldsymbol{p}] + \mathbf{B}\circ\mathbf{J}[\boldsymbol{p}]$, where $\boldsymbol{p} = (2,3)^\top$, $K = 2$, $p = 5$, each square represents an element of $\mathbf{\Sigma}[\mathbf{A},\mathbf{B},\boldsymbol{p}]$, different colors represent different values.

the importance of this representation. As demonstrated below, $\mathbf{A}$, $\mathbf{B}$, and $\boldsymbol{p}$ are sufficient and necessary for determining the expressions for the power, inverse (if it exists), and eigenvalues of a UB matrix $\mathbf{N}[\mathbf{A},\mathbf{B},\boldsymbol{p}]$.

Corollary 2.2.1 (algebraic properties of UB matrices)**.** *Given a common partition-size vector* $\boldsymbol{p} = (p_1,\ldots,p_K)^\top$, *suppose* $\boldsymbol{N} = \boldsymbol{N}[\boldsymbol{A},\boldsymbol{B},\boldsymbol{p}]$, $\boldsymbol{N}_1 = \boldsymbol{N}_1[\boldsymbol{A}_1,\boldsymbol{B}_1,\boldsymbol{p}]$ *and* $\boldsymbol{N}_2 = \boldsymbol{N}_2[\boldsymbol{A}_2,\boldsymbol{B}_2,\boldsymbol{p}]$ *are UB matrices with* K *by* K *diagonal matrices* $\boldsymbol{A} = \mathrm{diag}(a_{11},\ldots,a_{KK})$, $\boldsymbol{A}_1$, $\boldsymbol{A}_2$, K *by* K *symmetric matrices* $\boldsymbol{B} = (b_{kk'})$, $\boldsymbol{B}_1$, $\boldsymbol{B}_2$. *Let* $\boldsymbol{\Delta} = \boldsymbol{A} + \boldsymbol{B}\times\boldsymbol{P} \in \mathbb{R}^{K\times K}$ *with* $\boldsymbol{P} = \mathrm{diag}(p_1,\ldots,p_K)$.

(1) (Addition/Subtraction) suppose $\boldsymbol{N}^* = \boldsymbol{N}_1 \pm \boldsymbol{N}_2$, *then the partitioned matrix of* $\boldsymbol{N}^*$ *by* $\boldsymbol{p}$ *is a UB matrix, denoted by* $\boldsymbol{N}^*[\boldsymbol{A}^*,\boldsymbol{B}^*,\boldsymbol{p}]$, *where* $\boldsymbol{A}^* = \boldsymbol{A}_1 \pm \boldsymbol{A}_2$ *and* $\boldsymbol{B}^* = \boldsymbol{B}_1 \pm \boldsymbol{B}_2$;

(2) (Product) suppose $\boldsymbol{N}^* = \boldsymbol{N}_1 \times \boldsymbol{N}_2$, *in general,* $\boldsymbol{N}^*$ *is not a UB matrix. But, if* $\boldsymbol{N}_1$ *and* $\boldsymbol{N}_2$ *are commute, i.e.,* $\boldsymbol{N}_1\times\boldsymbol{N}_2 = \boldsymbol{N}_2\times\boldsymbol{N}_1$, *then* $\boldsymbol{N}^*$ *is a UB matrix, denoted by* $\boldsymbol{N}^*[\boldsymbol{A}^*,\boldsymbol{B}^*,\boldsymbol{p}]$, *where* $\boldsymbol{A}^* = \boldsymbol{A}_1\times\boldsymbol{A}_2 = \boldsymbol{A}^{*,\top}$ *and* $\boldsymbol{B}^* = \boldsymbol{A}_1\times\boldsymbol{B}_2 + \boldsymbol{B}_1\times\boldsymbol{A}_2 + \boldsymbol{B}_1\times\boldsymbol{P}\times\boldsymbol{B}_2 = \boldsymbol{B}^{*,\top}$; *in particular,*

(2-1) (Square) suppose $\boldsymbol{N}^* = \boldsymbol{N}\times\boldsymbol{N}$, *then the partitioned matrix of* $\boldsymbol{N}^*$ *by* $\boldsymbol{p}$ *is a UB matrix, denoted by* $\boldsymbol{N}^*[\boldsymbol{A}^*,\boldsymbol{B}^*,\boldsymbol{p}]$, *where* $\boldsymbol{A}^* = \boldsymbol{A}\times\boldsymbol{A}$ *and* $\boldsymbol{B}^* = \boldsymbol{A}\times\boldsymbol{B} + \boldsymbol{B}\times\boldsymbol{A} + \boldsymbol{B}\times\boldsymbol{P}\times\boldsymbol{B}$;

(2-2) (Power) suppose $\boldsymbol{N}^* = \boldsymbol{N} \times \cdots \times \boldsymbol{N} = \boldsymbol{N}^m$ *with integer* $m \geq 2$*, then the partitioned matrix of* $\boldsymbol{N}^*$ *by* $\boldsymbol{p}$ *is a UB matrix, denoted by* $\boldsymbol{N}^*\left[\boldsymbol{A}^{(m)}, \boldsymbol{B}^{(m)}, \boldsymbol{p}\right]$*, where* $\boldsymbol{A}^{(1)} = \boldsymbol{A}$*,* $\boldsymbol{B}^{(1)} = \boldsymbol{B}$*,* $\boldsymbol{A}^{(m')} = \boldsymbol{A}^{(m'-1)} \times \boldsymbol{A}$ *and* $\boldsymbol{B}^{(m')} = \boldsymbol{A}^{(m'-1)} \times \boldsymbol{B} + \boldsymbol{B}^{(m'-1)} \times \boldsymbol{A} + \boldsymbol{B}^{(m'-1)} \times \boldsymbol{P} \times \boldsymbol{B}$ *for* $m' = 2, \ldots, m$*;*

(3) (Eigenvalues) $\boldsymbol{N}[\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p}]$ *has* p *real eigenvalues in total, those are* a_{kk} *with multiplicity* $(p_k - 1)$ *for* $k = 1, \ldots, K$ *and the rest* K *eigenvalues are identical with those of* $\boldsymbol{\Delta}$*;*

(4) (Determinant) $\boldsymbol{N}[\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p}]$ *has the determinant* $\left(\prod_{k=1}^{K} a_{kk}^{p_k - 1}\right) \times \det(\boldsymbol{\Delta})$*;*

(5) (Inverse) suppose $\boldsymbol{N}$ *is invertible and* $\boldsymbol{N}^* = \boldsymbol{N}^{-1}$*, then the partitioned matrix of* $\boldsymbol{N}^*$ *by* $\boldsymbol{p}$ *is a UB matrix, denoted by* $\boldsymbol{N}^*[\boldsymbol{A}^*, \boldsymbol{B}^*, \boldsymbol{p}]$*, where* $\boldsymbol{A}^* = \boldsymbol{A}^{-1}$ *and* $\boldsymbol{B}^* = -\boldsymbol{\Delta}^{-1} \times \boldsymbol{B} \times \boldsymbol{A}^{-1}$*.*

(6) (Canonical Form) let $\bar{p}_0 = 0$*,* $\bar{p}_k = \sum_{k'=1}^{k} p_k$ *for* $k = 1, \ldots, K$ *(then* $\bar{p}_K = p$*), and* λ_j *denote the* j*-th eigenvalue of* $\boldsymbol{N}[\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p}]$*, where* $\lambda_1 = \cdots = \lambda_{\bar{p}_1 - 1} = a_{11}$*,* $\lambda_{\bar{p}_1 + 1} = \cdots = \lambda_{\bar{p}_2 - 1} = a_{22}$*, …,* $\lambda_{\bar{p}_{K-1}+1} = \cdots = \lambda_{\bar{p}_K - 1} = a_{KK}$ *and the rest* $\lambda_{\bar{p}_1}, \lambda_{\bar{p}_2}, \ldots, \lambda_{\bar{p}_K}$ *are identical with the eigenvalues of* $\boldsymbol{\Delta}$ *(in the decreasing order). Thus, there exists an* p *by* p *orthogonal matrix* $\boldsymbol{\Gamma}$ *satisfying that* $\boldsymbol{\Gamma} \times \boldsymbol{N}[\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p}] \times \boldsymbol{\Gamma}^\top = diag(\lambda_1, \lambda_2, \ldots, \lambda_p)$ *and* $\boldsymbol{\Gamma}$ *can be constructed by* K *Helmert submatrices and* K *row vectors as follows:*

$$\boldsymbol{\Gamma} = \begin{pmatrix} \tilde{\boldsymbol{H}}_1 & \boldsymbol{0}_{(p_1-1)\times p_2} & \cdots & \boldsymbol{0}_{(p_1-1)\times p_K} \\ \xi_{1,1}\boldsymbol{1}_{1\times p_1} & \xi_{1,2}\boldsymbol{1}_{1\times p_2} & \cdots & \xi_{1,K}\boldsymbol{1}_{1\times p_K} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0}_{(p_K-1)\times p_1} & \boldsymbol{0}_{(p_K-1)\times p_2} & \cdots & \tilde{\boldsymbol{H}}_K \\ \xi_{K,1}\boldsymbol{1}_{1\times p_1} & \xi_{K,2}\boldsymbol{1}_{1\times p_2} & \cdots & \xi_{K,K}\boldsymbol{1}_{1\times p_K} \end{pmatrix},$$

where $\tilde{\boldsymbol{H}}_k \in \mathbb{R}^{(p_k-1)\times p_k}$ *is the submatrix of a standard Helmert matrix of order* p_k *without the first row (Lancaster, 1965) and* $\boldsymbol{\xi}_k = (\xi_{k,1}, \xi_{k,2}, \ldots, \xi_{k,K})^\top \in \mathbb{R}^{K\times 1}$ *is the eigenvector (normalized to*

the unit length) of $\boldsymbol{\Delta}$ *corresponding to the eigenvalue* $\lambda_{\bar{p}_k}$ *for every* k.

Remark (sufficient and necessary condition for positive definiteness)**.** Let $\mathbf{P} = \mathrm{diag}(p_1, \ldots, p_K)$. We observe that the term $\boldsymbol{\Delta} = \mathbf{A} + \mathbf{B} \times \mathbf{P}$ plays a critical role in determining the eigenvalues, determinant, and inverse of an invertible UB matrix $\mathbf{N}[\mathbf{A}, \mathbf{B}, \boldsymbol{p}]$. Although $\boldsymbol{\Delta}$ is not symmetric in general, it has K real eigenvalues because $\boldsymbol{\Delta} = \left(\mathbf{A}\mathbf{P}^{-1} + \mathbf{B}\right)\mathbf{P}$ is similar to a real symmetric matrix $\mathbf{P}^{1/2}\left(\mathbf{A}\mathbf{P}^{-1} + \mathbf{B}\right)\mathbf{P}^{1/2}$ (because they have the same characteristic polynomial), which has K real eigenvalues. Therefore, $\mathbf{N}[\mathbf{A}, \mathbf{B}, \boldsymbol{p}]$ is positive definite or invertible, if and only if $\mathbf{A}$ is positive definite (i.e., $a_{kk} > 0$ for every k) and $\boldsymbol{\Delta}$ has K positive eigenvalues.

Remark (quadratic subspace)**.** Consider the trace as an inner product, and let $\mathscr{A}$ denote the finite-dimensional Hilbert space of p by p real symmetric matrices. A subspace $\mathscr{B}$ of $\mathscr{A}$ is said to be a *quadratic subspace* of $\mathscr{A}$, if $B \in \mathscr{B}$ implies that $B^2 \in \mathscr{B}$ (Seely, 1971). By the square property in Corollary 2.2.1, the collection of all UB matrices having a common partition-size vector forms a quadratic subspace. Quadratic subspaces are useful in studying the completeness of minimal sufficient statistics in a family of multivariate normal distributions (Seely, 1971, 1977; Zmyślony, 1980). For example, Szatrowski (1980) explored the relationship between the quadratic subspace and the explicit representation of maximum likelihood estimators for covariance matrices in a normal model. Roy et al. (2016) proved the optimal properties of the unbiased estimator that they derived for estimating a blocked compound symmetry covariance matrix.

Remark (algebraic properties for non-symmetric UB matrices)**.** Most results in Corollary 2.2.1 also hold for non-symmetric uniform-block matrices. Specifically, the sum, difference, and product of two non-symmetric UB matrices are still a non-symmetric UB matrix with the same expressions of $\mathbf{A}^*$ and $\mathbf{B}^*$ as in Corollary 2.2.1. The determinant and inverse (if it exists) of a

non-symmetric UB matrix are also a non-symmetric UB matrix with the same expressions of $\mathbf{A}^*$ and $\mathbf{B}^*$. However, it is worth noting that although a non-symmetric UB matrix $\mathbf{N}[\mathbf{A}, \mathbf{B}, \boldsymbol{p}]$ still has p eigenvalues, i.e., a_{kk} with multiplicity $(p_k - 1)$ for every k and K eigenvalues of $\boldsymbol{\Delta}$, some of the K eigenvalues of $\boldsymbol{\Delta}$ may be complex. Subsequently, we may rearrange the K Helmert submatrices in $\boldsymbol{\Gamma}$ below the remaining K row vectors, resulting in a block-diagonal canonical form for a non-symmetric UB matrix (please see Theorem 1 in Archakov and Hansen (2022)). If $\boldsymbol{\Delta}$ is diagonalizable, the canonical form will have a diagonal structure, and $\boldsymbol{\Gamma}$ remains the same as in Corollary 2.2.1, where $\lambda_{\bar{p}_1}, \ldots, \lambda_{\bar{p}_K}$ may be ordered in decreasing real parts.

The results in Corollary 2.2.1 highlight the following two-fold advantage of using the block Hadamard product representations of UB matrices. First, calculations on K by K matrices $\mathbf{A}$ and $\mathbf{B}$ can replace calculations on a larger p by p matrix $\mathbf{N}$, where K is typically much smaller than p, e.g., a proteomics study has $K = 7$ and $p = 107$ and a brain imaging study has $K = 5$ and $p = 227$ in Chapter 3. This reduction in matrix size can significantly improve computational efficiency. Second, the results involving addition or subtraction of UB matrices with a common partition-size vector, as well as operations such as taking the square (or power), computing eigenvalues, determinant, and inverse (if it exists) of a UB matrix can be expressed in terms of the "coordinates" $\mathbf{A}$, $\mathbf{B}$, and $\boldsymbol{p}$. These results greatly facilitate the use of UB matrices in various fields of applications. For example, in Chapter 3, we proposed the best-unbiased covariance- and precision-matrix estimators when the number of diagonal blocks K is fixed, as well as a modified hard-thresholding covariance matrix estimator when K grows with the sample size, respectively.

Before proceeding to hypothesis testing problems in the next sections, we specify the relationships between a covariance matrix and its precision and correlation matrix.

Corollary 2.2.2 (covariance matrix with a UB structure)**.** *Given a partition-size vector* $\boldsymbol{p} = (p_1, \ldots, p_K)^\top$*, suppose* $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}[\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p}]$ *is a* p *by* p *positive definite covariance matrix with a uniform-block structure, where* $\boldsymbol{A} = \operatorname{diag}(a_{11}, \ldots, a_{KK})$ *and* $\boldsymbol{B} = (b_{kk'})$ *with* $b_{k'k} = b_{kk'}$ *for every* $k \neq k'$*. Then, the partitioned matrix of* $\boldsymbol{\Theta} = \boldsymbol{\Sigma}^{-1}$ *by* $\boldsymbol{p}$ *is a UB matrix, denoted by* $\boldsymbol{\Theta}[\boldsymbol{A}_{\boldsymbol{\Theta}}, \boldsymbol{B}_{\boldsymbol{\Theta}}, \boldsymbol{p}]$*; the partitioned matrix of* $\boldsymbol{\Xi} = \operatorname{corr}(\boldsymbol{\Sigma})$ *by* $\boldsymbol{p}$ *is a UB matrix, denoted by* $\boldsymbol{\Xi}[\boldsymbol{A}_{\boldsymbol{\Xi}}, \boldsymbol{B}_{\boldsymbol{\Xi}}, \boldsymbol{p}]$*, where*

$$\begin{cases} \boldsymbol{A}_{\boldsymbol{\Theta}} = \boldsymbol{A}^{-1} \\ \boldsymbol{B}_{\boldsymbol{\Theta}} = -\boldsymbol{\Delta}^{-1} \times \boldsymbol{B} \times \boldsymbol{A}^{-1} \end{cases}, \begin{cases} \boldsymbol{A}_{\boldsymbol{\Xi}} = \boldsymbol{C}^{-1/2} \times \boldsymbol{A} \times \boldsymbol{C}^{-1/2} \\ \boldsymbol{B}_{\boldsymbol{\Xi}} = \boldsymbol{C}^{-1/2} \times \boldsymbol{B} \times \boldsymbol{C}^{-1/2} \end{cases},$$

with $\boldsymbol{\Delta} = \boldsymbol{A} + \boldsymbol{B} \times \boldsymbol{P}$*,* $\boldsymbol{P} = \operatorname{diag}(p_1, \ldots, p_K)$*,* $\boldsymbol{C} = \operatorname{diag}(c_{11}, \ldots, c_{KK})$*, and* $c_{kk} = a_{kk} + b_{kk}$ *for every* k*.*

2.3 Testing A Specific Mean for One-Sample

In the case where the number of diagonal blocks $K = 1$ or $K = 2$, Geisser (1963) proposed an information test statistic for testing a specific mean vector based on a multivariate normal sample and derived its exact null distribution in closed form. The distribution of Geisser's information test statistic under the null hypothesis is identical to the distribution of a sum of several independent F-variates. However, for the general case $K > 2$, Geisser (1963) provided an algorithm for calculating the information test statistic and explicitly formulated the exact null distribution, but omitted the proofs. In this section, we present the exact null distribution of the one-sample Geisser's information test statistic using the notations of UB matrices: this exact null distribution is equivalent to the distribution of a linear combination of mutually independent F-variates,

where the last variate is exactly the Hotelling's T^2 statistic.

Specifically, given p-dimensional normal vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{\text{i.i.d.}}{\sim} N(\boldsymbol{\mu}, \boldsymbol{\Sigma}[\mathbf{A}, \mathbf{B}, \boldsymbol{p}])$, the null and alternative hypotheses are given by

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0, \tag{2.3.1}$$

where the covariance matrix is known to have a UB stricture, $\mathbf{A} = \operatorname{diag}(a_{11}, \ldots, a_{KK})$ is an unknown diagonal matrix, $\mathbf{B} = (b_{kk'})$ is an unknown symmetric matrix with $b_{k'k} = b_{kk'}$ for every $k \neq k'$, $\boldsymbol{p} = (p_1, \ldots, p_K)^\top$ is a known partition-size vector, and $\boldsymbol{\mu}_0 \in \mathbb{R}^p$ is a pre-determined vector. To guarantee positive definiteness of $\boldsymbol{\Sigma}[\mathbf{A}, \mathbf{B}, \boldsymbol{p}]$, we assume $\mathbf{A}$ is positive definite and $\boldsymbol{\Delta} = \mathbf{A} + \mathbf{B} \times \mathbf{P}$ has positive eigenvalues only, with $\mathbf{P} = \operatorname{diag}(p_1, \ldots, p_K)$.

Before deriving Geisser's information test statistic for a specific mean vector, we introduce the maximum likelihood estimator of $\boldsymbol{\Sigma}[\mathbf{A}, \mathbf{B}, \boldsymbol{p}]$ based on a multivariate normal sample. Let $\bar{\boldsymbol{X}} = n^{-1}(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)$, $\mathbf{S} = (n-1)^{-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^\top$, and $(\mathbf{S}_{kk'})$ denote the sample mean, the (unbiased) sample covariance matrix, and the partitioned matrix of $\mathbf{S}$ by $\boldsymbol{p}$, respectively. If the sample size is larger than the total number of unknown covariance parameters, i.e., $n > K + (K+1)K/2$, then we can obtain the best-unbiased estimators of $\mathbf{A}$ and $\mathbf{B}$, denoted by $\widehat{\mathbf{A}} = \operatorname{diag}(\widehat{a}_{11}, \ldots, \widehat{a}_{KK})$ and $\widehat{\mathbf{B}} = \left(\widehat{b}_{kk'}\right)$ with $\widehat{b}_{k'k} = \widehat{b}_{kk'}$ for every $k \neq k'$, respectively, where $\widehat{a}_{kk}$ and $\widehat{b}_{kk'}$ are given by

$$\widehat{a}_{kk} = \frac{p_k \times \operatorname{tr}(\mathbf{S}_{kk}) - \operatorname{sum}(\mathbf{S}_{kk})}{p_k \times (p_k - 1)}, \quad \widehat{b}_{kk'} = \begin{cases} \dfrac{\operatorname{sum}(\mathbf{S}_{kk'})}{p_k \times p_{k'}}, & k \neq k' \\ \dfrac{\operatorname{sum}(\mathbf{S}_{kk'}) - \operatorname{tr}(\mathbf{S}_{kk'})}{p_k \times (p_{k'} - 1)}, & k = k' \end{cases} \tag{2.3.2}$$

for every k and every k, k' respectively (see the details in Chapter 3). It is clear that the maximum likelihood estimator $\widehat{a}_{kk}$ is exactly the average of the off-diagonal elements within the (k,k)-th diagonal block of $(\mathbf{S}_{kk'})$; $\widehat{b}_{kk}$ is exactly the average of diagonal elements within the (k,k)-th diagonal block minus $\widehat{a}_{kk}$; and $\widehat{b}_{kk'}$ is the average of all elements within the (k,k')-th off-diagonal block. By Corollary 2.2.2, the plug-in estimators of $\boldsymbol{\Sigma}\left[\mathbf{A},\mathbf{B},\boldsymbol{p}\right]$ and $\boldsymbol{\Theta}\left[\mathbf{A}_{\boldsymbol{\Theta}},\mathbf{B}_{\boldsymbol{\Theta}},\boldsymbol{p}\right]$ are $\widehat{\boldsymbol{\Sigma}}\left[\widehat{\mathbf{A}},\widehat{\mathbf{B}},\boldsymbol{p}\right]$ and $\widehat{\boldsymbol{\Theta}}\left[\widehat{\mathbf{A}}_{\boldsymbol{\Theta}},\widehat{\mathbf{B}}_{\boldsymbol{\Theta}},\boldsymbol{p}\right]$, where $\widehat{\mathbf{A}}$ and $\widehat{\boldsymbol{\Delta}}=\widehat{\mathbf{A}}+\widehat{\mathbf{B}}\times\mathbf{P}$ are assumed to be positive definite and have positive eigenvalues only, respectively, and $\widehat{\mathbf{A}}_{\boldsymbol{\Theta}}$ and $\widehat{\mathbf{B}}_{\boldsymbol{\Theta}}$ are given by $\widehat{\mathbf{A}}_{\boldsymbol{\Theta}}=\widehat{\mathbf{A}}^{-1}$ and $\widehat{\mathbf{B}}_{\boldsymbol{\Theta}}=-\widehat{\boldsymbol{\Delta}}^{-1}\times\widehat{\mathbf{B}}\times\widehat{\mathbf{A}}^{-1}$.

Theorem 2.3.1 (exact null distribution of Geisser's one-sample information test statistic)**.** *Geisser's one-sample test statistic for the hypotheses in* (2.3.1) *is given by*

$$U=n\times\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)^{\top}\times\widehat{\boldsymbol{\Theta}}\left[\widehat{\boldsymbol{A}}_{\boldsymbol{\Theta}},\widehat{\boldsymbol{B}}_{\boldsymbol{\Theta}},\boldsymbol{p}\right]\times\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right).$$

Under H_0, it follows a distribution $\mathcal{U}$ that is identical with the distribution of

$$\sum_{k=1}^{K}\left(p_k-1\right)F^{(k)}_{(p_k-1),(p_k-1)(n-1)}+T^2,$$

where $T^2=K(n-1)(n-K)^{-1}F^{(K+1)}_{K,n-K}$ is the Hotelling's T^2-statistic and $F^{(k)}_{df_1,df_2}$ are $(K+1)$ mutually independent F-variates with degrees of freedom df_1 and df_2 for $k=1,\ldots,K+1$.

Remark (information test statistic U)**.** Geisser's information criterion was proposed by Geisser (1963) to test the hypotheses in (2.3.1) using analysis of variance tables. In the distribution $\mathcal{U}$, the last variate is precisely Hotelling's (generalized) T^2-statistic, which is most likely used in multivariate inference (Anderson, 1992).

Remark (related distributions under H_0)**.** (1) As $n \to \infty$, U asymptotically follows a chi-square distribution χ^2_p where $p = p_1 + \cdots + p_K$ (Geisser, 1963);

(2) Given a significance level α and an arbitrary vector $\boldsymbol{a} \in \mathbb{R}^p$, the $100(1-\alpha)\%$ simultaneous confidence interval for a measurable function $\boldsymbol{a}^\top \boldsymbol{\mu}$ has the form

$$\boldsymbol{a}^\top \bar{\boldsymbol{X}} + \pm \sqrt{\mathcal{U}(\alpha) \times \boldsymbol{a}^\top \widehat{\boldsymbol{\Sigma}} \left[\widehat{\mathbf{A}}, \widehat{\mathbf{B}}, \boldsymbol{p}\right] \boldsymbol{a}/n},$$

where $\mathcal{U}(\alpha)$ denotes the upper α-th percentile of the distribution $\mathcal{U}$ (Morrison, 1972);

(3) An approximate of $\mathcal{U}$ is suggested as the distribution of $C_1 F_{(p,C_2)}$ by Morrison (1971), where the scale coefficient C_1 and the second degree of freedom C_2 are determined by equating the first two cumulants of $C_1 F_{(p,C_2)}$ to those of $\mathcal{U}$. The specific values of C_1 and C_2 for $K = 2$ can be found in Spjøtvoll (1972) and Young (1976). Furthermore, Dyer (1982) considered the distribution of the sum of generalized F variates and Lee and Hu (1996) extended the above result to independent F-variates with arbitrary coefficients and degrees of freedom.

Remark (related distributions under H_1)**.** (1) The non-null distribution was analogous to the null distribution $\mathcal{U}$, except that one or more F-variates are non-central (Geisser, 1963);

(2) An approximate non-null distribution of U can be represented as $D_1 F_{(p,D_2)}(\delta)$, where the noncentrality parameter $\delta = n\left(\boldsymbol{\mu} - \boldsymbol{\mu}_0\right)^\top \boldsymbol{\Theta}\left[\mathbf{A}_{\boldsymbol{\Theta}}, \mathbf{B}_{\boldsymbol{\Theta}}, \boldsymbol{p}\right]\left(\boldsymbol{\mu} - \boldsymbol{\mu}_0\right)$, and the scale coefficient D_1 and the second degree of freedom D_2 are determined by equating the first two cumulants of $D_1 F_{(p,D_2)}(\delta)$ to the non-null distribution of U.

2.4 Testing the Equality of Means for Multiple-Sample

We consider a general M-sample mean test ($M > 1$), where the samples are drawn from M normal distributions with means $\boldsymbol{\mu}^{(m)} \in \mathbb{R}^p$, for $m = 1, \ldots, M$, and an equal covariance matrix with a UB structure $\boldsymbol{\Sigma}[\mathbf{A}, \mathbf{B}, \boldsymbol{p}] \in \mathbb{R}^{p \times p}$.

Specifically, suppose the m-th sample $\boldsymbol{X}_1^{(m)}, \ldots, \boldsymbol{X}_{n_m}^{(m)}$ has a size of n_m, for $m = 1, 2, \ldots, M$. Thus, the grand sample size is denoted by $n = n_1 + \cdots + n_M$ and we assume $n > \max\{M, K + (K+1)K/2\}$. Let $\bar{\boldsymbol{X}}^{(m)} = n_m^{-1}\left(\boldsymbol{X}_1^{(m)} + \cdots + \boldsymbol{X}_{n_m}^{(m)}\right)$, $\bar{\boldsymbol{X}} = n^{-1}\left(n_1\bar{\boldsymbol{X}}^{(1)} + \cdots + n_M\bar{\boldsymbol{X}}^{(M)}\right)$, and $\mathbf{S} = (n-M)^{-1}\sum_{m=1}^{M}\sum_{j=1}^{n_m}\left(\boldsymbol{X}_j^{(m)} - \bar{\boldsymbol{X}}^{(m)}\right)\left(\boldsymbol{X}_j^{(m)} - \bar{\boldsymbol{X}}^{(m)}\right)^\top$ denote the m-th sample mean, the grand sample mean, and the (pooled) unbiased estimator of the common covariance matrix, respectively. The maximum likelihood estimators $\widehat{\mathbf{A}}$, $\widehat{\mathbf{B}}$ can be obtained similarly to those in (2.3.2), yielding the estimators $\widehat{\mathbf{A}}_{\boldsymbol{\Theta}}$, $\widehat{\mathbf{B}}_{\boldsymbol{\Theta}}$, and $\widehat{\boldsymbol{\Theta}}\left[\widehat{\mathbf{A}}_{\boldsymbol{\Theta}}, \widehat{\mathbf{B}}_{\boldsymbol{\Theta}}, \boldsymbol{p}\right]$, respectively. Therefore, the null and alternative hypotheses can be written as

$$H_0^{(M)} : \boldsymbol{\mu}^{(1)} = \cdots = \boldsymbol{\mu}^{(M)} \quad \text{versus} \quad H_1^{(M)} : \boldsymbol{\mu}^{(m')} \neq \boldsymbol{\mu}^{(m)} \text{ for some } m' \,. \tag{2.4.1}$$

Theorem 2.4.1 (exact null distribution of Geisser's multiple-sample information test statistic)**.** *Geisser's multiple-sample information test statistic for the hypotheses in* (2.4.1) *is given by*

$$U_M = \sum_{m=1}^{M} n_m \left(\bar{\boldsymbol{X}}^{(m)} - \bar{\boldsymbol{X}}\right)^\top \widehat{\boldsymbol{\Theta}}\left[\widehat{\boldsymbol{A}}_{\boldsymbol{\Theta}}, \widehat{\boldsymbol{B}}_{\boldsymbol{\Theta}}, \boldsymbol{p}\right]\left(\bar{\boldsymbol{X}}^{(m)} - \bar{\boldsymbol{X}}\right).$$

Under $H_0^{(M)}$, it follows a distribution $\mathcal{U}_M$ that is identical with the distribution of

$$\sum_{k=1}^{K}(M-1)(p_k-1)F^{(k)}_{(M-1)(p_k-1),(n-M)(p_k-1)} + T_0^2$$

where T_0^2 is the Hotelling's T_0^2-statistic, $F^{(k)}_{df_1,df_2}$ are K mutually independent F-variates (and independent from T_0^2) with degrees of freedom df_1 and df_2 for $k = 1, \ldots, K$.

Remark (information test statistic U_M)**.** The Hotelling's (generalized) T_0^2-statistic, also known as the Hotelling-Lawley trace, is commonly used to test the equality of multiple populations means, assuming these multiple normal populations have the same (arbitrary) population covariance matrices (Lawley, 1938; Hotelling, 1947, 1951). However, it is intractable to obtain the exact null or non-null distribution of Hotelling's T_0^2-statistic, and various approximations have been proposed in the literature (Ito, 1956, 1960; Pillai and Young, 1971; Siotani, 1971; McKeon, 1974).

2.5 Discussion

In this chapter, we concentrate on the algebraic properties of a specific type of block matrices, where each block is uniform. We chose to parameterize the matrices in this way for two key reasons. First, the uniform-block pattern has been popularly discovered in plenty of large-scale biological data. Second, from a biological perspective, the variables that are clustered into the same community may exhibit stochastic equivalence or comparable patterns, while variables from different communities may have coherent connections at the community level. Compared to the conventional diagonal or block-diagonal structure, the proposed uniform-block structure offers more flexibility and is better suited for real data analysis, since the information contained

in the non-zero off-diagonal blocks can potentially provide valuable insights into the scientific mechanisms.

In addition to defining a uniform-block structure, we have discovered a unique block Hadamard product representation for a uniform-block matrix. This representation plays an important role because it allows for the transformation of a large-scale uniform-block matrix into two lower-dimensional matrices and an integer-valued vector. The block Hadamard product representation simplifies the computations related to uniform-block matrices. With these algebraic properties, the uniform-block matrices are applicable to various statistical problems. For example, covariance estimation with the uniform-block structure in Chapter 3, hypothesis testing for the information test statistics, the multivariate linear regression models in Chapter 4, and confirmatory factor analysis models in Chapter 5.

In conclusion, a uniform-block matrix (or structure), its associated algebraic properties, and the block Hadamard product representation have broad applications in a range of fields, including linear algebra, statistics, economics, and many others.

Chapter 3: Covariance Matrix Estimation for High-Throughput Biomedical Data with Dependence Structure of Interconnected Communities

3.1 Introduction

Technological innovations in biomedicine have facilitated the generation of high-dimensional datasets with simultaneous measurements of up to millions of biological features (Fan and Lv, 2008). In the past few decades, numerous statistical methods have been developed to analyze these large-dimensional datasets. Estimating a covariance matrix (or a precision matrix) is fundamental to these analyses (Fan et al., 2016; Cai et al., 2016; Wainwright, 2019) because a covariance matrix not only describes the complex interactive relations among variables but also leads to accurate inferential and predictive results for clinical outcomes (He et al., 2019; Ke et al., 2022). Since the dimensionality of the variables is much larger than the sample size, we resort to advanced statistical methods rather than traditional covariance estimation strategies (Johnstone, 2001; Johnstone and Paul, 2018). The shrinkage and thresholding methods can provide a reliable and robust covariance estimator under the sparsity assumption (Ledoit and Wolf, 2004; Bickel and Levina, 2008b; Rothman et al., 2009; Cai and Liu, 2011). In addition, prior knowledge of a covariance structure can greatly improve the accuracy of estimation and statistical inference (Fan, 2005; Bien, 2019). For example, recent methods can accommodate Toeplitz, banded,

block-diagonal covariance structures (Cai et al., 2013; Bickel and Levina, 2008a; Devijver and Gallopin, 2018).

In the present research, we consider a widely prevalent covariance structure in a large body of high-dimensional datasets (see the examples in Figure 3.1, extra examples in Chapter B, and more examples in Figure 1.1). A well-organized block pattern is widely observed in most commonly used biomedical data types, including genetics, proteomics, brain imaging, and RNA expression data, among many others (Spellman et al., 1998; Yildiz et al., 2007; Chiappelli et al., 2019; Chen et al., 2016; He et al., 2019, 2015; Wu et al., 2021). This block pattern exhibits several properties of highly organized networks. For example, there is high modality as some variables are clustered in the multiple and coherently correlated communities; there is small-worldness as these communities are interconnected; and the network is scale-free as the remaining variables are isolated if singletons (see the top parts of Figure 3.1(B) and (C)) are detected (Newman, 2006). Therefore, we can specify this well-organized block pattern by assigning the high-dimensional variables to multiple interconnected communities (see the middle parts of Figure 3.1(B) and (C)), which are more informative, and a set of detected singletons. The interconnected community structures might not be directly available from the high-dimensional biomedical data, but they can be estimated by several clustering algorithms and network detection methods (Magwene, 2021; Wu et al., 2021). Although there are potential benefits to leveraging the estimated interconnected community structure to enhance the estimation of large covariance matrices, existing statistical methods are restricted to establishing a connection between this structure and the covariance or precision parameters and providing precise estimates. To address this methodological gap, we propose a novel statistical procedure that enables closed-form estimators of large covariance and precision matrices and supports robust statistical inference.

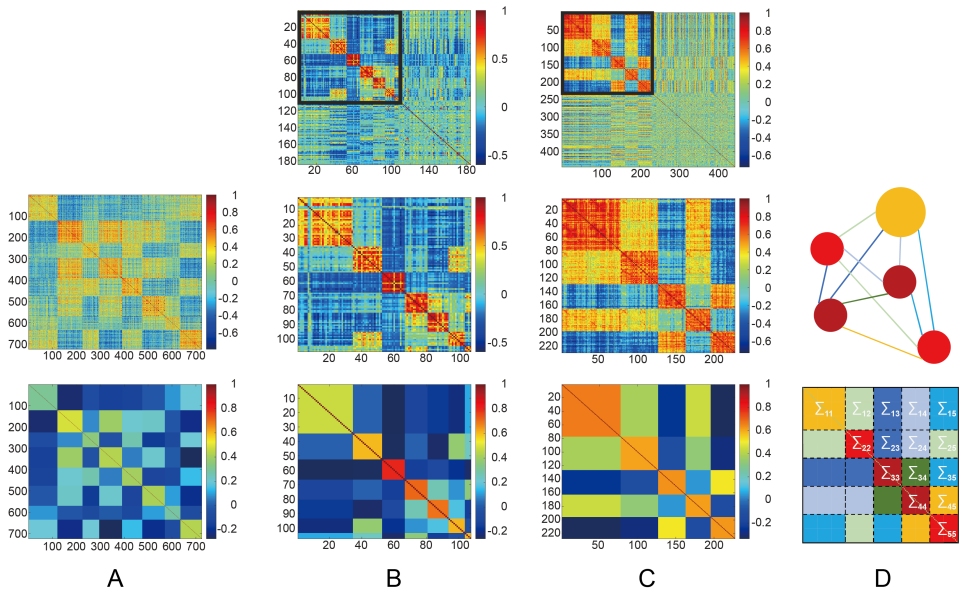


Figure 3.1: A: top, heat map of the sample correlation matrix calculated by a K-medoids clustering algorithm for Spellman's yeast genome study, showing 8 by 8 well-organized blocks (without singletons) in the structure; bottom: an assumed population correlation matrix with an 8 by 8 uniform-block structure. B and C: top, heat maps of the sample correlation matrices calculated by a network detection algorithm for proteomics and brain imaging datasets, respectively, showing the block patterns (with singletons); middle, heat maps in the black frames in the top parts, showing 7 by 7 and 5 by 5 well-organized blocks respectively in the structure; bottom, assumed population correlation matrices with a 7 by 7 and a 5 by 5 uniform-block structures, respectively. D: top, illustration of a network model with 5 communities; bottom, a corresponding population matrix with a 5 by 5 uniform-block structure.

We propose a parametric covariance model that subdivides the covariance matrix into blocks or submatrices and assigns each block to either a community or an interconnection between two communities based on the observed interconnected community structure. Linking the covariance parameters and the underlying network topological structure, we can facilitate a closed-form estimator of each covariance parameter (using only the elements within the corresponding block) and can establish the asymptotic properties for the proposed estimators. Specifically, we derive the explicit estimators using advanced matrix theories including the block Hadamard product representation of a covariance matrix with the above structure, and the thresholding covariance regularization in the high-dimensional setting, where the number of covariance parameters in all blocks exceeds the sample size.

Our method makes at least three novel contributions. Firstly, we have developed a fast (closed-form) and accurate procedure for estimating a large covariance (and precision) matrix with a particular structure that is applicable to various high-throughput biomedical data including genomics, metabolomics, proteomics, neuroimaging data, and many others. Our method quantitatively characterizes the interconnected community structure by estimating parameters in both diagonal sub-matrices (i.e., communities) and off-diagonal sub-matrices (i.e., interactions between communities), and thus better reveals the interactive mechanisms of a complex biosystem. By utilizing this interconnected community structure, our large covariance matrix estimation procedure also outperforms comparable methods in terms of accuracy and is numerically robust to model misspecification. Consequently, our approach can lead to a more precise selection of biological features of interest (e.g., cancer-related gene expressions) in the context of multiple testing which relies on the accurate and reliable large covariance- and precision-matrix estimation (Fan et al., 2012; Fan and Han, 2017). Secondly, we have derived the exact variance estimators

of the covariance parameter estimators and established the asymptotic properties, which enables us to evaluate covariance patterns and provide confidence intervals. Lastly, we have extended our method to accommodate scenarios where the number of diagonal blocks exceeds the sample size, allowing for scalability to accommodate ultra-high-dimensional datasets.

The rest of the chapter is organized as follows. Section 3.2 introduces our proposed method. We first mathematically define the uniform-block structure and the uniform-block matrix in Section 3.2.1. Then, in Section 3.2.2, we derive the best-unbiased covariance- and precision-matrix estimators for small K by taking advantage of the block Hadamard product representation. We generalize the estimating procedure to large K with a diverging number of diagonal blocks in Section 3.2.3. Section 3.3 contains thorough numerical evaluations of our method under various scenarios. Section 3.4 illustrates the use of our method in two real-world applications. We provide a discussion in Section 3.5. Additional data examples, the exact covariance estimators for the covariance parameters in the blocks, more simulation studies, and all the technical proofs are given in Chapter B.

3.2 Methodology

3.2.1 A parametric covariance model with a uniform-block structure

Suppose $\mathbf{X} \in \mathbb{R}^{n\times p}$ is an n by p observed data matrix containing n independent and identically distributed p-variate normal vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with mean $\boldsymbol{\mu} \coloneqq \mathrm{E}(\boldsymbol{X}_1) \in \mathbb{R}^p$, positive definite covariance matrix $\boldsymbol{\Sigma} \coloneqq \mathrm{cov}(\boldsymbol{X}_1) \in \mathbb{R}^{p\times p}$ (denoted by $\boldsymbol{\Sigma} \succ 0$), and precision matrix $\boldsymbol{\Omega} \coloneqq \boldsymbol{\Sigma}^{-1}$. Let $\mathbf{S}^*$ and $\mathbf{S}$ denote the biased and unbiased versions of the sample covariance matrix, respectively: $\mathbf{S} \coloneqq (n-1)^{-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^\top$ and $\mathbf{S}^* \coloneqq n^{-1}(n-1)\mathbf{S}$ with

$\bar{\boldsymbol{X}} \coloneqq n^{-1}(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)$, where $\mathbf{M}^\top$ denotes the transpose of a matrix (or a vector) $\mathbf{M}$. Moreover, we require that the covariance matrix has the uniform-block structure described below. The parameterization of this covariance structure is illustrated in two steps.

We specify the parametric covariance matrix based on our previous discussions of the uniform-block structure. First, we use a vector to characterize the community sizes. Specifically, given the dimension p of the covariance matrix $\boldsymbol{\Sigma}$ and the number of communities K, let $p_1, \ldots, p_K$ be positive integers satisfying $p_k > 1$ $(k = 1, \ldots, K)$ and $p = p_1 + \cdots + p_K$ and let $\boldsymbol{p} \coloneqq (p_1, \ldots, p_K)^\top$ be the *partition-size vector*, which is assumed to be fixed throughout the paper. Given $\boldsymbol{p}$, we can express $\boldsymbol{\Sigma}$ *partitioned by* $\boldsymbol{p}$ *in block form*:

$$(\boldsymbol{\Sigma}_{kk'}) \coloneqq \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \ldots & \boldsymbol{\Sigma}_{1K} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \ldots & \boldsymbol{\Sigma}_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{K1} & \boldsymbol{\Sigma}_{K2} & \ldots & \boldsymbol{\Sigma}_{KK} \end{pmatrix}, \tag{3.2.1}$$

where $\boldsymbol{\Sigma}_{kk'} \in \mathbb{R}^{p_k \times p_{k'}}$ $(k, k' = 1, \ldots, K)$. Second, following (3.2.1), we specify the diagonal submatrix $\boldsymbol{\Sigma}_{kk} \coloneqq a_{kk}\mathbf{I}_{p_k} + b_{kk}\mathbf{J}_{p_k}$ for every k and the off-diagonal submatrix $\boldsymbol{\Sigma}_{kk'} \coloneqq b_{kk'}\mathbf{1}_{p_k \times p_{k'}}$ with $b_{kk'} = b_{k'k}$ for every $k \neq k'$, where $\mathbf{I}_s$, $\mathbf{J}_s$, and $\mathbf{1}_{s\times t}$ are an s by s identity matrix, an s by s all-one matrix, and an s by t all-one matrix, respectively. Using $K < p_1 + \cdots + p_K = p$, we can represent a large covariance matrix $\boldsymbol{\Sigma}$ by a smaller diagonal matrix $\mathbf{A} \coloneqq \operatorname{diag}(a_{11}, \ldots, a_{KK})$, a

smaller symmetric matrix $\mathbf{B} \coloneqq (b_{kk'})$ with $b_{kk'} = b_{k'k}$ for every $k \neq k'$, and a known vector $\boldsymbol{p}$:

$$\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p}) \coloneqq (\boldsymbol{\Sigma}_{kk'}) = \begin{pmatrix} a_{11}\mathbf{I}_{p_1} + b_{11}\mathbf{J}_{p_1} & b_{12}\mathbf{1}_{p_1 \times p_2} & \dots & b_{1K}\mathbf{1}_{p_1 \times p_K} \\ b_{21}\mathbf{1}_{p_2 \times p_1} & a_{22}\mathbf{I}_{p_2} + b_{22}\mathbf{J}_{p_2} & \dots & b_{2K}\mathbf{1}_{p_2 \times p_K} \\ \vdots & \vdots & \ddots & \vdots \\ b_{K1}\mathbf{1}_{p_K \times p_1} & b_{K2}\mathbf{1}_{p_K \times p_2} & \dots & a_{KK}\mathbf{I}_{p_K} + b_{KK}\mathbf{J}_{p_K} \end{pmatrix}. \tag{3.2.2}$$

In this chapter, we say that the pattern in (3.2.2) is a *uniform-block structure*. If $\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$ has the structure in (3.2.2), it is a *uniform-block matrix*. This covariance parameterization strategy based on $\boldsymbol{\Sigma}_{kk}$ and $\boldsymbol{\Sigma}_{kk'}$ has been widely used in the statistics literature. For example, the generalized estimation equations where the working correlation structure has a compound symmetry as well as linear mixed-effects models with a random intercept both have this pattern. In practice, this parameterization strategy can characterize the covariance structure well using a parsimonious model (as shown in Figure 3.1). In Section 3.3, we demonstrate that the performance of this parameterization is robust under misspecification and matrix perturbation. By building the parsimonious and effective covariance-matrix specification, we can develop reliable and accurate covariance-matrix estimations using the likelihood approach and achieve optimistic properties.

Notice that the partition-size vector $\boldsymbol{p}$ is assumed to be known throughout this chapter. In practice, it can be learned by a preliminary algorithm (for example, a K-medoids clustering algorithm (Magwene, 2021) or a network detection algorithm (Wu et al., 2021)) before estimating the covariance matrix. Also, the above definition of a uniform-block matrix does not guarantee its positive definiteness in general. Thus, additional constraints should be imposed on the uniform-block matrix $\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$ to ensure that it is a valid covariance matrix. We defer the discussion of

these constraints to the next section.

3.2.2 Matrix estimation for the uniform-block structure with a small K

Given a partition-size vector $\boldsymbol{p}$ and the parametric covariance matrix $\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$ with the uniform-block structure (3.2.2), we define a q-dimensional parameter vector

$$\boldsymbol{\theta} \coloneqq (a_{11}, \ldots, a_{KK}, b_{11}, \ldots, b_{1K}, b_{22}, \ldots, b_{KK})^\top$$

consisting of the covariance parameters in the blocks. That is, the parameters of interest are in the upper triangular part of $\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$. Also, $q = K + K(K+1)/2$. Thus, the problem of estimating a p by p symmetric covariance matrix reduces to that of estimating the q-dimensional parameter vector $\boldsymbol{\theta}$. In practice, q is considerably smaller than $p(p+1)/2$, thereby remarkably reducing the dimensionality of the parameters of interest.

The small K setting can be specified as follows: $q < n$ while K, p, and q are fixed and p can be greater than n. In other words, both the number of diagonal blocks K and the number of parameters in the blocks $q = K + K(K+1)/2$ are smaller than the sample size n. Moreover, we require that K and the dimension of the covariance matrix p, which is proportional to K, are fixed, so that q is also fixed. Also, p can be large enough to exceed n with small K. In this section, we first introduce an explicit maximum likelihood estimator of $\boldsymbol{\theta}$ with asymptotic properties and then refine it to be the best-unbiased estimator, whose exact variance estimator is also provided in closed form. Finally, we provide the best unbiased covariance- and precision-matrix estimators for small K.

We start with the maximum likelihood estimation. Specifically, let $(\mathbf{S}^*_{kk'})$ be the block form

of the biased sample covariance matrix $\mathbf{S}^*$ partitioned by $\boldsymbol{p}$. If the data $\mathbf{X}$ are normally distributed, then the log-likelihood function of the data can be expressed by

$$\ell_n(\boldsymbol{\theta}; \mathbf{X}) \propto \frac{n}{2} \log \left(\det \left[\{\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})\}^{-1} \right] \right) - \frac{n}{2} \operatorname{tr} \left[(\mathbf{S}^*_{kk'}) \times \{\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})\}^{-1} \right].$$

A typical way to estimate $\boldsymbol{\theta}$ in the literature is to derive the score function by taking the first-order partial derivative of the log-likelihood function with respect to $\boldsymbol{\theta}_j$:

$$\frac{\partial}{\partial \boldsymbol{\theta}_j} \ell_n(\boldsymbol{\theta}; \mathbf{X}) = \frac{n}{2} \operatorname{tr} \left[\{\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p}) - (\mathbf{S}^*_{kk'})\} \times \frac{\partial \{\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})\}^{-1}}{\partial \boldsymbol{\theta}_j} \right] \quad (j = 1, \ldots, q), \quad (3.2.3)$$

where $\boldsymbol{\theta}_j \in \{a_{11}, \ldots, a_{KK}, b_{11}, \ldots, b_{1K}, b_{22}, \ldots, b_{KK}\}$ denotes the jth component of $\boldsymbol{\theta}$ and $\partial \{\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})\}^{-1} / \partial \boldsymbol{\theta}_j \in \mathbb{R}^{p \times p}$ is a matrix whose entries are functions of $\boldsymbol{\theta}_j$.

However, solving the score equation (3.2.3) is challenging. Although the unknown entries a_{kk} and $b_{kk'}$ are uniformly and elegantly arranged in $\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$, they are entangled in a complex way in the precision matrix $\boldsymbol{\Omega} = \{\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})\}^{-1}$. In other words, $\boldsymbol{\theta}_j$ is implicit in $\boldsymbol{\Omega}$ so that the closed form of $\boldsymbol{\Omega}$ is not accessible in general. The complexity of the calculation increases if the scale of the precision matrix increases. Alternatively, the existing numerical algorithms for solving $\boldsymbol{\theta}$ (for example, the method of averaging (Szatrowski, 1980)) rely on iterative updating schemes, which require a long computational time and may lead to unstable estimates. These facts motivate us to reconsider the possibility of deriving a closed-form estimator of $\boldsymbol{\theta}$. Thus, we aim to find an explicit expression for $\boldsymbol{\Omega}$ in terms of a_{kk} and $b_{kk'}$ by taking advantage of the special covariance structure in (3.2.2). More precisely, we speculate that $\boldsymbol{\Omega}$ has an analogous form to (3.2.2), which indeed can be fulfilled by realizing the following representation of the

block Hadamard product.

Lemma 3.2.1. *Given a partition-size vector* $\boldsymbol{p} = (p_1, \ldots, p_K)^\top$ *satisfying* $p_k > 1$ *for every* k *and* $p = p_1 + \cdots + p_K$, $\boldsymbol{I}_p$ *and* $\boldsymbol{J}_p$ *partitioned by* $\boldsymbol{p}$ *are uniform-block matrices, expressed by* $\boldsymbol{I}(\boldsymbol{p}) \coloneqq \boldsymbol{I}_p(\boldsymbol{I}_K, \boldsymbol{0}_{K\times K}, \boldsymbol{p}) = \mathrm{Bdiag}(\boldsymbol{I}_{p_1}, \ldots, \boldsymbol{I}_{p_K})$ *and* $\boldsymbol{J}(\boldsymbol{p}) \coloneqq \boldsymbol{J}_p(\boldsymbol{0}_{K\times K}, \boldsymbol{1}_{K\times K}, \boldsymbol{p}) = (\boldsymbol{1}_{p_k\times p_{k'}})$, *respectively, where* $\boldsymbol{0}_{r\times s}$ *denotes the* r *by* s *zero matrix and* $\mathrm{Bdiag}(\cdot)$ *denotes a block-diagonal matrix. Suppose a* p *by* p *matrix* $\boldsymbol{N}$ *partitioned by* $\boldsymbol{p}$ *is a uniform-block matrix of the form* (3.2.2)*, expressed by* $\boldsymbol{N}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p})$*, then the following representation is unique,*

$$\boldsymbol{N}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p}) = \boldsymbol{A} \circ \boldsymbol{I}(\boldsymbol{p}) + \boldsymbol{B} \circ \boldsymbol{J}(\boldsymbol{p}),$$

where $\circ$ *denotes the block Hadamard product satisfying that* $\boldsymbol{A}\circ\boldsymbol{I}(\boldsymbol{p}) \coloneqq \mathrm{Bdiag}\left(a_{11}\boldsymbol{I}_{p_1}, \ldots, a_{KK}\boldsymbol{I}_{p_K}\right)$ *and* $\boldsymbol{B} \circ \boldsymbol{J}(\boldsymbol{p}) \coloneqq \left(b_{kk'}\boldsymbol{I}_{p_k\times p_{k'}}\right)$.

Based on Lemma 3.2.1, we derive several basic properties of a uniform-block matrix, which are summarized in Chapter 2. These properties reveal how $\mathbf{A}$, $\mathbf{B}$, and $\boldsymbol{p}$ determine the algebraic operations for a uniform-block matrix $\mathbf{N}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$ and how a collection of uniform-block matrices with the same $\boldsymbol{p}$ form a quadratic subspace (Seely, 1971). If we view $\mathbf{A}$, $\mathbf{B}$, and $\boldsymbol{p}$ as the "coordinates" of a uniform-block matrix, then using the notation $\mathbf{N}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$ can simplify the mathematical operations between p by p uniform-block matrices into those between their corresponding lower-dimensional K by K "coordinates". Following Chapter 2, we get two useful results for the covariance and precision matrices.

Corollary 3.2.1. *(1) Given* $\boldsymbol{\Sigma}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p})$ *defined in* (3.2.2)*,* $\boldsymbol{\Sigma}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p}) \succ 0$ *if and only if* $\boldsymbol{A} \succ 0$ *and* $\boldsymbol{\Delta}$ *has positive eigenvalues only, where* $\boldsymbol{\Delta} \coloneqq \boldsymbol{A} + \boldsymbol{B} \times \boldsymbol{P}$ *and* $\boldsymbol{P} \coloneqq \mathrm{diag}(p_1, \ldots, p_K)$.

(2) Suppose $\boldsymbol{\Sigma}(\boldsymbol{A},\boldsymbol{B},\boldsymbol{p}) \succ 0$ *and* $\boldsymbol{\Omega} = \{\boldsymbol{\Sigma}(\boldsymbol{A},\boldsymbol{B},\boldsymbol{p})\}^{-1}$ *is the precision matrix. Then,* $\boldsymbol{\Omega}$ *partitioned by* $\boldsymbol{p}$ *is a uniform-block matrix as well, which can be written by* $\boldsymbol{\Omega}(\boldsymbol{A}_{\boldsymbol{\Omega}},\boldsymbol{B}_{\boldsymbol{\Omega}},\boldsymbol{p}) = \boldsymbol{A}_{\boldsymbol{\Omega}} \circ \boldsymbol{I}(\boldsymbol{p}) + \boldsymbol{B}_{\boldsymbol{\Omega}} \circ \boldsymbol{J}(\boldsymbol{p})$*, where*

$$\boldsymbol{A}_{\boldsymbol{\Omega}} = \boldsymbol{A}^{-1}, \quad \boldsymbol{B}_{\boldsymbol{\Omega}} = -\boldsymbol{\Delta}^{-1} \times \boldsymbol{B} \times \boldsymbol{A}^{-1}. \tag{3.2.4}$$

The first assertion of Corollary 3.2.1 finalizes the additional constraints that guarantee that $\boldsymbol{\Sigma}(\mathbf{A},\mathbf{B},\boldsymbol{p}) \succ 0$. The second assertion confirms that the precision matrix $\boldsymbol{\Omega} = \{\boldsymbol{\Sigma}(\mathbf{A},\mathbf{B},\boldsymbol{p})\}^{-1}$ partitioned by $\boldsymbol{p}$ is a uniform-block matrix, expressed by $\boldsymbol{\Omega}(\mathbf{A}_{\boldsymbol{\Omega}},\mathbf{B}_{\boldsymbol{\Omega}},\boldsymbol{p})$. Furthermore, (3.2.4) provides the relations between $\mathbf{A}_{\boldsymbol{\Omega}}$, $\mathbf{B}_{\boldsymbol{\Omega}}$, and $\mathbf{A}$, $\mathbf{B}$.

Therefore, applying the representation of the precision matrix in Corollary 3.2.1, we can rewrite the partial derivative of the log-likelihood in (3.2.3) as

$$\frac{\partial}{\partial \boldsymbol{\theta}_j}\ell_n(\boldsymbol{\theta};\mathbf{X}) = \frac{n}{2}\,\mathrm{tr}\left[\{\boldsymbol{\Sigma}(\mathbf{A},\mathbf{B},\boldsymbol{p}) - (\mathbf{S}^*_{kk'})\} \times \left\{\frac{\partial \mathbf{A}_{\boldsymbol{\Omega}}}{\partial \boldsymbol{\theta}_j} \circ \mathbf{I}(\boldsymbol{p}) + \frac{\partial \mathbf{B}_{\boldsymbol{\Omega}}}{\partial \boldsymbol{\theta}_j} \circ \mathbf{J}(\boldsymbol{p})\right\}\right], \tag{3.2.5}$$

where $\boldsymbol{\theta}_j \in \{a_{11},\ldots,a_{KK},b_{11},\ldots,b_{1K},b_{22},\ldots,b_{KK}\}$ $(j = 1,\ldots,q)$. In contrast to (3.2.3), the derivatives in (3.2.5) can be calculated by:

$$\frac{\partial \mathbf{A}_{\boldsymbol{\Omega}}}{\partial a_{kk}} = -a_{kk}^{-2}\mathbf{E}_{kk}, \quad \frac{\partial \mathbf{B}_{\boldsymbol{\Omega}}}{\partial a_{kk}} = \boldsymbol{\Delta}^{-1}\mathbf{E}_{kk}\boldsymbol{\Delta}^{-1}\mathbf{B}\mathbf{A}^{-1} + a_{kk}^{-2}\boldsymbol{\Delta}^{-1}\mathbf{B}\mathbf{E}_{kk} \quad (k = 1,\ldots,K),$$

$$\frac{\partial \mathbf{A}_{\boldsymbol{\Omega}}}{\partial b_{kk'}} = \mathbf{0}_{K\times K}, \quad \frac{\partial \mathbf{B}_{\boldsymbol{\Omega}}}{\partial b_{kk'}} = \begin{cases} -\boldsymbol{\Delta}^{-1}\mathbf{E}_{kk}\mathbf{P}\boldsymbol{\Delta}^{-1}\mathbf{P}^{-1}, & k = k' \\ -\boldsymbol{\Delta}^{-1}(\mathbf{E}_{kk'} + \mathbf{E}_{k'k})\mathbf{P}\boldsymbol{\Delta}^{-1}\mathbf{P}^{-1}, & k \neq k' \end{cases} \quad (k,k' = 1,\ldots,K),$$

where $\mathbf{E}_{kk'} \in \mathbb{R}^{K\times K}$ denotes a matrix with 1 in the (k,k') entry and 0 otherwise. The explicit

forms of the derivatives highlight the advantage of (3.2.5) over (3.2.3). Owing to (3.2.5), we can explicitly derive the analytic form of the maximum likelihood estimator for $\boldsymbol{\theta}$:

$$\widetilde{a}^*_{kk} \coloneqq \frac{\mathrm{tr}(\mathbf{S}^*_{kk})}{p_k - 1} - \frac{\mathrm{sum}(\mathbf{S}^*_{kk})}{p_k \times (p_k - 1)}, \quad \widetilde{b}^*_{kk'} \coloneqq \begin{cases} \dfrac{\mathrm{sum}(\mathbf{S}^*_{kk'})}{p_k \times p_{k'}}, & k \neq k' \\ \dfrac{\mathrm{sum}(\mathbf{S}^*_{kk'}) - \mathrm{tr}(\mathbf{S}^*_{kk'})}{p_k \times (p_{k'} - 1)}, & k = k' \end{cases}, \tag{3.2.6}$$

for every k and k', where $\mathrm{sum}(\mathbf{M}) = \sum_{j=1}^{r} \sum_{j'=1}^{s} m_{jj'}$ denotes the sum of all entries in $\mathbf{M} \coloneqq (m_{jj'}) \in \mathbb{R}^{r \times s}$. Technical details are referred to Chapter B. We denote the maximum likelihood estimator of $\boldsymbol{\theta}$ as

$$\widetilde{\boldsymbol{\theta}}^* \coloneqq (\widetilde{a}^*_{11}, \ldots, \widetilde{a}^*_{KK}, \widetilde{b}^*_{11}, \ldots, \widetilde{b}^*_{1K}, \widetilde{b}^*_{22}, \ldots, \widetilde{b}^*_{KK})^\top.$$

The strong consistency, the asymptotic efficiency, and normality for $\widetilde{\boldsymbol{\theta}}^*$ can similarly be derived by standard procedures in the literature (Ferguson, 1996; van der Vaart and Wellner, 1996; van der Vaart, 2000). Please refer to Chapter B.

Despite its asymptotic property, $\widetilde{\boldsymbol{\theta}}^*$ is biased under a finite sample size and, therefore, is not a uniformly minimum variance unbiased estimator. Namely, $\widetilde{\boldsymbol{\theta}}^*$ is not the best unbiased estimator for $\boldsymbol{\theta}$. Thus, it is natural to consider an unbiased estimator of $\boldsymbol{\theta}$ by replacing $\mathbf{S}^*$ with the unbiased version $\mathbf{S}$ in (3.2.6). Specifically, let $(\mathbf{S}_{kk'})$ be the block form of $\mathbf{S}$ partitioned by $\boldsymbol{p}$. Substituting $\mathbf{S}_{kk'}$ for $\mathbf{S}^*_{kk'}$ in (3.2.6), for every k and k', we have

$$\widetilde{a}_{kk} \coloneqq \frac{\mathrm{tr}(\mathbf{S}_{kk})}{p_k - 1} - \frac{\mathrm{sum}(\mathbf{S}_{kk})}{p_k \times (p_k - 1)}, \quad \widetilde{b}_{kk'} \coloneqq \begin{cases} \dfrac{\mathrm{sum}(\mathbf{S}_{kk'})}{p_k \times p_{k'}}, & k \neq k' \\ \dfrac{\mathrm{sum}(\mathbf{S}_{kk'}) - \mathrm{tr}(\mathbf{S}_{kk'})}{p_k \times (p_{k'} - 1)}, & k = k' \end{cases}. \tag{3.2.7}$$

We denote the above unbiased estimator as $\widetilde{\boldsymbol{\theta}} \coloneqq (\widetilde{a}_{11}, \ldots, \widetilde{a}_{KK}, \widetilde{b}_{11}, \ldots, \widetilde{b}_{1K}, \widetilde{b}_{22}, \ldots, \widetilde{b}_{KK})^\top$.

Alternatively, since $\mathbf{S} = n(n-1)^{-1}\mathbf{S}^*$, we have a relation between the maximum likelihood estimator $\widetilde{\boldsymbol{\theta}}^*$ and the unbiased estimator $\widetilde{\boldsymbol{\theta}}$:

$$\widetilde{a}_{kk} = \frac{n}{n-1}\widetilde{a}^*_{kk}, \quad \widetilde{b}_{kk'} = \frac{n}{n-1}\widetilde{b}^*_{kk'} \quad (k, k' = 1, \ldots, K), \quad \widetilde{\boldsymbol{\theta}} = \frac{n}{n-1}\widetilde{\boldsymbol{\theta}}^*.$$

Moreover, the following theorem summarizes the optimal property of $\widetilde{\boldsymbol{\theta}}$.

Theorem 3.2.1 (Optimal property of $\widetilde{\boldsymbol{\theta}}$)**.** *$\widetilde{\boldsymbol{\theta}}$ is the best-unbiased estimator of $\boldsymbol{\theta}$, and*

$$\widetilde{\boldsymbol{\mu}} \coloneqq \bar{\boldsymbol{X}}, \quad \widetilde{\boldsymbol{\Sigma}}(\widetilde{\boldsymbol{A}}, \widetilde{\boldsymbol{B}}, \boldsymbol{p}) = \widetilde{\boldsymbol{A}} \circ \boldsymbol{I}(\boldsymbol{p}) + \widetilde{\boldsymbol{B}} \circ \boldsymbol{J}(\boldsymbol{p}) \tag{3.2.8}$$

are the best unbiased estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p})$, where $\widetilde{\boldsymbol{A}} \coloneqq \operatorname{diag}(\widetilde{a}_{11}, \ldots, \widetilde{a}_{KK})$ and $\widetilde{\boldsymbol{B}} \coloneqq (\widetilde{b}_{kk'})$ with $\widetilde{b}_{kk'} = \widetilde{b}_{k'k}$ for every $k \neq k'$ are the best unbiased estimators of $\boldsymbol{A}$ and $\boldsymbol{B}$.

Geisser (1963) and Morrison (1972) derived the same estimator $\widetilde{\boldsymbol{\theta}}$ based on an analysis of the variance table, but they did not show the above optimal property. Furthermore, after calculating $\widetilde{\boldsymbol{\Delta}} \coloneqq \widetilde{\mathbf{A}} + \widetilde{\mathbf{B}} \times \mathbf{P}$ with the best unbiased estimators $\widetilde{\mathbf{A}}$ and $\widetilde{\mathbf{B}}$, we derive the best-unbiased estimator for the precision matrix, as summarized in the following corollary.

Corollary 3.2.2. *If both $\widetilde{\boldsymbol{A}}$ and $\widetilde{\boldsymbol{\Delta}}$ are positive definite, then*

$$\widetilde{\boldsymbol{\Omega}}(\widetilde{\boldsymbol{A}}_{\boldsymbol{\Omega}}, \widetilde{\boldsymbol{B}}_{\boldsymbol{\Omega}}, \boldsymbol{p}) = \widetilde{\boldsymbol{A}}_{\boldsymbol{\Omega}} \circ \boldsymbol{I}(\boldsymbol{p}) + \widetilde{\boldsymbol{B}}_{\boldsymbol{\Omega}} \circ \boldsymbol{J}(\boldsymbol{p}) \tag{3.2.9}$$

is the best unbiased estimator of $\boldsymbol{\Omega}(\boldsymbol{A}_{\boldsymbol{\Omega}}, \boldsymbol{B}_{\boldsymbol{\Omega}}, \boldsymbol{p})$, $\widetilde{\boldsymbol{A}}_{\boldsymbol{\Omega}} \coloneqq \widetilde{\boldsymbol{A}}^{-1}$ and $\widetilde{\boldsymbol{B}}_{\boldsymbol{\Omega}} \coloneqq -\widetilde{\boldsymbol{\Delta}}^{-1} \times \widetilde{\boldsymbol{B}} \times \widetilde{\boldsymbol{A}}^{-1}$.

The covariance matrix of the proposed estimator $\widetilde{\boldsymbol{\theta}}$ can be calculated from the Fisher information matrix based on the asymptotic normality. Given a small K (say, $K \leq 3$, empirically),

the Fisher information matrix and its inverse may have explicit expressions. For some real applications in which K is relatively large (say, $K > 3$), a calculation of the inverse of the Fisher information matrix might be burdensome and unstable. Alternatively, we provide the exact variance and covariance estimators of the elements of $\widetilde{\boldsymbol{\theta}}$ in Corollary 3.2.3 and Chapter B under a finite sample size, respectively.

Corollary 3.2.3. *The exact variance estimators of $\widetilde{a}_{kk}$ and $\widetilde{b}_{kk'}$ are*

$$\mathrm{var}(\widetilde{a}_{kk}) = \frac{2a_{kk}^2}{(n-1)(p_k-1)},$$

$$\mathrm{var}(\widetilde{b}_{kk'}) = \begin{cases} \frac{2}{(n-1)p_k(p_k-1)} \left\{(a_{kk}+p_k b_{kk})^2 - (2a_{kk}+p_k b_{kk})b_{kk}\right\}, & k = k' \\ \frac{1}{2(n-1)p_k p_{k'}} \left\{p_k p_{k'}(b_{kk'}^2 + b_{k'k}^2) + 2(a_{kk}+p_k b_{kk})(a_{k'k'}+p_{k'} b_{k'k'})\right\}, & k \neq k' \end{cases}$$

for every k and k'. The exact covariance estimators of $\widetilde{a}_{\ell\ell}$ and $\widetilde{b}_{kk'}$ for every ℓ, k and k' are given in Chapter B.

At the end of this section, we briefly introduce how we can extend the above estimation procedure to situations with singletons. In addition to the communities assumed to follow the uniform-block structure (for example, the blocks in the black frames in the top parts of Figure 3.1(B) and (C)), the other singletons on the diagonal are also commonly detected in real applications (see the diagonal entries outside the black frames in the top parts of Figure 3.1(B) and (C)), such as a proteomics study (Yildiz et al., 2007) and a brain imaging study (Chiappelli et al., 2019). Notice that a structure of well-organized communities with singletons may imply that a population covariance matrix of the form $\boldsymbol{\Sigma}_{\text{(full)}} := \{\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p}), \mathbf{D}_1; \mathbf{D}_1^\top, \mathbf{D}_2\}$, where $\mathbf{D}_1 \in \mathbb{R}^{p\times d}$, $\mathbf{D}_2 \in \mathbb{R}^{d\times d}$, and $\boldsymbol{\Sigma}_{\text{(full)}} \in \mathbb{R}^{(p+d)\times(p+d)}$, and d denotes the number of singletons. Following the

decomposition idea in Fan et al. (2013), we can estimate $\mathbf{\Sigma}_{\text{(full)}}$ in two steps. First, we partition the unbiased sample covariance matrix $\mathbf{S}_{\text{(full)}} \coloneqq \{(\mathbf{S}_{kk'}), \mathbf{S}_1; \mathbf{S}_1^\top, \mathbf{S}_2\}$ in the same way. Second, we obtain the best unbiased estimator $\widetilde{\mathbf{\Sigma}}(\widetilde{\mathbf{A}}, \widetilde{\mathbf{B}}, \boldsymbol{p})$ based on $(\mathbf{S}_{kk'})$ and obtain the hard-thresholding estimator $\widetilde{\mathbf{R}}_{\text{(full)}}$ of the matrix $\mathbf{R}_{\text{(full)}} \coloneqq \{\mathbf{0}_{p\times p}, \mathbf{S}_1; \mathbf{S}_1^\top, \mathbf{S}_2\}$ (Bickel and Levina, 2008a), yielding the consistent estimator $\widetilde{\mathbf{\Sigma}}_{\text{(full)}} \coloneqq \text{Bdiag}\{\widetilde{\mathbf{\Sigma}}(\widetilde{\mathbf{A}}, \widetilde{\mathbf{B}}, \boldsymbol{p}), \mathbf{0}_{d\times d}\} + \widetilde{\mathbf{R}}_{\text{(full)}}$ under a matrix norm if p is fixed and d may grow with the sample size n.

3.2.3 Matrix estimation for the uniform-block structure with a large K

In Section 3.2.2, we estimated the covariance matrix with the uniform-block structure and its precision matrix for small K. However, there are applications where the covariance matrices are uniform-block matrices with more diagonal blocks than the sample size. More specifically, a large K occurs when K, $q > n$ and K, p, and q grow with n. In other words, the number of diagonal blocks K is greater than the sample size n, and so is the number of covariance parameters in the blocks: $q = K + K(K+1)/2$. Moreover, we require that K, the dimension p of the covariance matrix (which is proportional to K), and q grow with n and diverge as n goes to infinity. In this section, we generalize the proposed estimation procedure for small K and introduce a consistent covariance-matrix estimator by modifying the hard-thresholding method with large K. Denote $\|\mathbf{M}\|_{\text{F}} = (\sum_{j=1}^r \sum_{j'=1}^r m_{jj'}^2)^{1/2}$ and $\|\mathbf{M}\|_{\text{S}} = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{M}\mathbf{x}\|_2$ as the Frobenius norm and spectral norm of $\mathbf{M} \coloneqq (m_{jj'}) \in \mathbb{R}^{r\times r}$ respectively, where $\|\mathbf{x}\|_2 \coloneqq (\sum_{j=1}^r x_j^2)^{1/2}$ for $\mathbf{x} \coloneqq (x_1, \ldots, x_r)^\top \in \mathbb{R}^r$.

For normal data $\mathbf{X} \in \mathbb{R}^{n\times p}$ with population mean $\boldsymbol{\mu} = \mathbf{0}_{p\times 1}$, the unbiased sample covariance matrix $\mathbf{S} = \mathbf{X}^\top\mathbf{X}/n$, and the population covariance matrix with the uniform-block structure

$\mathbf{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$ for large K, we propose a new thresholding approach based on the work by Bickel and Levina (2008b). Since covariance matrix $\mathbf{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$ is fully determined by $\mathbf{A}$, $\mathbf{B}$, and $\boldsymbol{p}$ according to Lemma 3.2.1, we threshold the estimates of $\mathbf{A}$ and $\mathbf{B}$, rather than $\mathbf{S}$, to yield a covariance-matrix estimate. Specifically, given a thresholding level $\lambda = \lambda_n > 0$, let

$$\widehat{a}_{kk}(\lambda) \coloneqq \widetilde{a}_{kk} \times \mathbb{I}(|\widetilde{a}_{kk}| > \lambda), \quad \widehat{b}_{kk'}(\lambda) \coloneqq \widetilde{b}_{kk'} \times \mathbb{I}(|\widetilde{b}_{kk'}| > \lambda) \tag{3.2.10}$$

be the hard-thresholding estimators of a_{kk} and $b_{kk'}$, respectively, where $\widetilde{a}_{kk}$ and $\widetilde{b}_{kk'}$ are the best unbiased estimators for every k and k' and $\mathbb{I}(\cdot)$ is the indicator function. We regard the new covariance-matrix estimator

$$\widehat{\mathbf{\Sigma}}_\lambda(\widehat{\mathbf{A}}_\lambda, \widehat{\mathbf{B}}_\lambda, \boldsymbol{p}) = \widehat{\mathbf{A}}_\lambda \circ \mathbf{I}(\boldsymbol{p}) + \widehat{\mathbf{B}}_\lambda \circ \mathbf{J}(\boldsymbol{p}) \tag{3.2.11}$$

as the *modified hard-thresholding estimator* of $\mathbf{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$, where $\widehat{\mathbf{A}}_\lambda \coloneqq \operatorname{diag}\{\widehat{a}_{11}, \ldots, \widehat{a}_{KK}\}$, $\widehat{\mathbf{B}}_\lambda \coloneqq \{\widehat{b}_{kk'}(\lambda)\}$ with $\widehat{b}_{kk'}(\lambda) = \widehat{b}_{k'k}(\lambda)$ for every $k \neq k'$. The consistency of this modified hard-thresholding estimator is summarized below.

Theorem 3.2.2 (Consistency of the modified hard-thresholding estimator)**.** *Given a positive-definite covariance matrix with the uniform-block structure* $\mathbf{\Sigma}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p}) = (\sigma_{jj'})$ *defined as* (3.2.2)*, suppose* $K = K(n) > n \to \infty$ *and* $\log(K)/n \to 0$ *as* $n \to \infty$*, and there exist constants* $0 < C_{p_0}, C_{q_0} < \infty$ *such that* $\sum_{j=1}^{p} \sum_{j'=1}^{p} |\sigma_{jj'}|^{p_0} \leq C_{p_0}$ *and* $\max_j \sum_{j'=1}^{p} |\sigma_{jj'}|^{q_0} \leq C_{q_0}$ *for*

$0 < p_0 < 2$ *and* $0 < q_0 < 1$*, then*

$$\|\widehat{\boldsymbol{\Sigma}}_\lambda(\widehat{\boldsymbol{A}}_\lambda, \widehat{\boldsymbol{B}}_\lambda, \boldsymbol{p}) - \boldsymbol{\Sigma}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p})\|_F^2 \leq O_P(1) C_{p_0} \left\{ \frac{\log(K)}{n} \right\}^{1-\frac{p_0}{2}},$$
$$\|\widehat{\Sigma}_\lambda(\widehat{\boldsymbol{A}}_\lambda, \widehat{\boldsymbol{B}}_\lambda, \boldsymbol{p}) - \boldsymbol{\Sigma}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p})\|_S \leq O_P(1) C_{q_0} \left\{ \frac{\log(K)}{n} \right\}^{\frac{1-q_0}{2}},$$

where we choose $\lambda = C\{\log(K)/n\}^{1/2}$ *for some positive constant* C*.*

The performance of the modified hard-thresholding estimator (3.2.10) relies on the choice of λ, which can be determined by applying the resampling rule in Bickel and Levina (2008a,b).

3.3 Numerical Studies

3.3.1 Simulations

To evaluate the performance of the proposed method comprehensively, we simulate data and benchmark them against comparable estimation methods for large covariance (and precision) matrices in the following three scenarios.

In Scenario 1 (Section 3.3.2), we generate normal data using a covariance matrix with a uniform-block structure $\boldsymbol{\Sigma}_{0,1}(\mathbf{A}_0, \mathbf{B}_0, \boldsymbol{p}_1)$ for n subjects for small K. That is, the number of diagonal blocks K is small so that the number of covariance parameters in the blocks q is smaller than the sample size n, whereas the dimension of the covariance matrix p can be greater than n. We first focus on evaluating the finite-sample performance of the parameter vector estimator $\widetilde{\boldsymbol{\theta}}$ in (3.2.7) by comparing the estimates with the ground truth. We also assess the accuracy of the exact covariance estimator for $\widetilde{\boldsymbol{\theta}}$, as presented in Chapter B. We then present the performance of the covariance estimator $\widetilde{\boldsymbol{\Sigma}}_1(\widetilde{\mathbf{A}}_1, \widetilde{\mathbf{B}}_1, \boldsymbol{p}_1)$ in (3.2.8) and the precision estimator $\widetilde{\boldsymbol{\Omega}}_1(\widetilde{\mathbf{A}}_{\boldsymbol{\Omega},1}, \widetilde{\mathbf{B}}_{\boldsymbol{\Omega},1}, \boldsymbol{p}_1)$

in (3.2.9) by comparing their losses in the matrix norms with those of existing covariance- and precision-matrix estimators.

In Scenario 2 (Section 3.3.3), we simulate normal data using the covariance matrix $\mathbf{\Sigma}_{0,2}(\mathbf{A}_0, \mathbf{B}_0, \boldsymbol{p}_2)$ with the structure of uniform blocks for large K. That is, the number of diagonal blocks K is greater than and grows with the sample size n, as does the number of parameters in the blocks q and the dimension of the covariance matrix p. Then, we compare the modified hard-thresholding estimator $\widehat{\mathbf{\Sigma}}_2(\widehat{\mathbf{A}}_2, \widehat{\mathbf{B}}_2, \boldsymbol{p}_2)$ in (3.2.11) with its competitors by computing the losses in the matrix norms.

In Scenario 3 (Section 3.3.4), we perform a misspecification analysis for $\widetilde{\mathbf{\Sigma}}_3(\widetilde{\mathbf{A}}_3, \widetilde{\mathbf{B}}_3, \boldsymbol{p}_3)$ and $\widetilde{\mathbf{\Omega}}_3(\widetilde{\mathbf{A}}_{\mathbf{\Omega},3}, \widetilde{\mathbf{B}}_{\mathbf{\Omega},3}, \boldsymbol{p}_3)$ when the covariance matrix is not a uniform-block matrix.

3.3.2 Scenario 1: comparison for small K covariance matrix

We first set the true covariance uniform-block matrix as $\mathbf{\Sigma}_{0,1}(\mathbf{A}_0, \mathbf{B}_0, \boldsymbol{p}_1) = \mathbf{A}_0 \circ \mathbf{I}(\boldsymbol{p}_1) + \mathbf{B}_0 \circ \mathbf{J}(\boldsymbol{p}_1)$, where the number of diagonal blocks $K = 5$; the partition-size vector

$$\boldsymbol{p}_1 \coloneqq (p_{\text{ind}}, p_{\text{ind}}, p_{\text{ind}}, p_{\text{ind}}, p_{\text{ind}})^\top = p_{\text{ind}} \times \mathbf{1}_{K\times 1}$$

with individual component $p_{\text{ind}} = 30$, 45, or 60; the number of parameters in the blocks $q = K + K(K+1)/2 = 20$; the dimension of the covariance matrix $p = K \times p_{\text{ind}} = 150$, 225, or

300; and $\mathbf{A}_0 \coloneqq \mathrm{diag}(a_{0,11}, \ldots, a_{0,KK})$ and $\mathbf{B}_0 \coloneqq (b_{0,kk'})$ with $b_{0,kk'} = b_{0,k'k}$ for $k \neq k'$:

$$\mathbf{A}_0 = \mathrm{diag}(0.016, 0.214, 0.749, 0.068, 0.100), \quad \mathbf{B}_0 = \begin{pmatrix} 6.731 & -1.690 & 0.696 & -2.936 & 1.913 \\ & 5.215 & 3.815 & -1.010 & 0.703 \\ & & 4.328 & -3.357 & -0.269 \\ & & & 6.788 & 0.000 \\ & & & & 3.954 \end{pmatrix}.$$

The true precision matrix $\boldsymbol{\Omega}_{0,1}(\mathbf{A}_{\boldsymbol{\Omega},0}, \mathbf{B}_{\boldsymbol{\Omega},0}, \boldsymbol{p}_1) = \mathbf{A}_{\boldsymbol{\Omega},0} \circ \mathbf{I}(\boldsymbol{p}_1) + \mathbf{B}_{\boldsymbol{\Omega},0} \circ \mathbf{J}(\boldsymbol{p}_1)$ is given by (3.2.4), that is, $\mathbf{A}_{\boldsymbol{\Omega},0} \coloneqq \mathbf{A}_0^{-1}$ and $\mathbf{B}_{\boldsymbol{\Omega},0} \coloneqq -(\mathbf{A}_0 + \mathbf{B}_0 \times \mathbf{P}_{0,1})^{-1} \times \mathbf{B}_0 \times \mathbf{A}_0^{-1}$, where $\mathbf{P}_{0,1} \coloneqq \mathrm{diag}(p_{\text{ind}}, p_{\text{ind}}, p_{\text{ind}}, p_{\text{ind}}, p_{\text{ind}})$. Specifically, we let $n = 50$, 100, or 150 satisfying $q = 20 < n$. We generate the data matrix $\mathbf{X}$ by drawing an independent and identically distributed sample from $N(\mathbf{0}_{p\times 1}, \boldsymbol{\Sigma}_{0,1}(\mathbf{A}_0, \mathbf{B}_0, \boldsymbol{p}_1))$ and repeat this procedure 1000 times.

For each replicate, we calculate $\widetilde{a}_{kk}$ and $\widetilde{b}_{kk'}$ using (3.2.7) to obtain $\widetilde{\mathbf{A}}_1 \coloneqq \mathrm{diag}(\widetilde{a}_{11}, \ldots, \widetilde{a}_{KK})$, $\widetilde{\mathbf{B}}_1 \coloneqq (\widetilde{b}_{kk'})$, $\widetilde{\mathbf{A}}_{\boldsymbol{\Omega},1} \coloneqq \widetilde{\mathbf{A}}_1^{-1}$, and $\widetilde{\mathbf{B}}_{\boldsymbol{\Omega},1} \coloneqq -(\widetilde{\mathbf{A}}_1 + \widetilde{\mathbf{B}}_1 \times \mathbf{P}_{0,1})^{-1} \times \widetilde{\mathbf{B}}_1 \times \widetilde{\mathbf{A}}_1^{-1}$ and then calculate their standard errors by substituting the estimates $\widetilde{a}_{kk}$ and $\widetilde{b}_{kk'}$ for a_{kk} and $b_{kk'}$, respectively, in Corollary 3.2.3. We also calculate the estimates of the covariance matrix and precision matrices using (3.2.8) and (3.2.9) and denote them by $\widetilde{\boldsymbol{\Sigma}}_{\text{prop.}} \coloneqq \widetilde{\boldsymbol{\Sigma}}_1(\widetilde{\mathbf{A}}_1, \widetilde{\mathbf{B}}_1, \boldsymbol{p}_1)$ and $\widetilde{\boldsymbol{\Omega}}_{\text{prop.}} \coloneqq \widetilde{\boldsymbol{\Omega}}_1(\widetilde{\mathbf{A}}_{\boldsymbol{\Omega},1}, \widetilde{\mathbf{B}}_{\boldsymbol{\Omega},1}, \boldsymbol{p}_1)$, respectively. In addition, we estimate the covariance matrix using the conventional methods including the soft-thresholding method (soft; Antoniadis and Fan, 2001), the hard-thresholding method (hard), the adaptive-thresholding method (adaptive), and the principal orthogonal complement thresholding method (POET), as shown in Figure 3.2. We also estimate the precision matrix using the graphical lasso method (glasso; Friedman et al., 2008), the

bandwidth test (banded; An et al., 2014), Bayesian frameworks with G-Wishart prior (BayesG; Banerjee and Ghosal, 2014) or with the k-banded Cholesky prior (BayesKBC; Lee and Lee, 2021), as also shown in the figure. Finally, we evaluate the performance of all methods using the losses in the Frobenius and spectral norms respectively, that is, $\|\widetilde{\boldsymbol{\Sigma}}_* - \boldsymbol{\Sigma}_{0,1}(\mathbf{A}_0, \mathbf{B}_0, \boldsymbol{p}_1)\|_{\mathrm{F}}$, $\|\widetilde{\boldsymbol{\Sigma}}_* - \boldsymbol{\Sigma}_{0,1}(\mathbf{A}_0, \mathbf{B}_0, \boldsymbol{p}_1)\|_{\mathrm{S}}$, $\|\widetilde{\boldsymbol{\Omega}}_* - \boldsymbol{\Omega}_{0,1}(\mathbf{A}_{\boldsymbol{\Omega},0}, \mathbf{B}_{\boldsymbol{\Omega},0}, \boldsymbol{p}_1)\|_{\mathrm{F}}$, and $\|\widetilde{\boldsymbol{\Omega}}_* - \boldsymbol{\Omega}_{0,1}(\mathbf{A}_{\boldsymbol{\Omega},0}, \mathbf{B}_{\boldsymbol{\Omega},0}, \boldsymbol{p}_1)\|_{\mathrm{S}}$ for the method $*$.

For the 1000 replicates, we assess the relative bias, the Monte Carlo standard deviation, the average standard error, and the empirical coverage probability based on 95% Wald-type confidence intervals for each covariance parameter, as presented in Table 3.1. The results in Table 3.1 demonstrate that the proposed best-unbiased estimators $\widetilde{a}_{kk}$ and $\widetilde{b}_{kk'}$ achieve satisfactory performance: the relative bias in the parameter estimations are generally small in contrast to the Monte Carlo standard deviations; as the sample size n increases, the average standard errors become smaller for all covariance parameters; the average standard errors are approximately equal to the Monte Carlo standard deviations with comparable corresponding 95% empirical coverage probabilities. We demonstrate the performance of covariance- and precision-matrix estimators in terms of the Frobenius and spectral norms and computational times for the proposed procedure and the competing methods in Figure 3.2. Our estimating procedure outperforms the existing methods as it is much faster and has smaller matrix norms losses. Compared with the conventional precision-matrix estimators, the proposed precision-matrix estimator has much lower losses because the true precision matrix has many non-sparse blocks (that is, the off-diagonal entries in $\mathbf{B}_{\boldsymbol{\Omega},0}$ are very small but different from 0). As the dimension p increases, the matrix norms losses between the proposed covariance estimate and $\boldsymbol{\Sigma}_{0,1}(\mathbf{A}_0, \mathbf{B}_0, \boldsymbol{p}_1)$ increase while those between the proposed precision estimate and $\boldsymbol{\Omega}_{0,1}(\mathbf{A}_{\boldsymbol{\Omega},0}, \mathbf{B}_{\boldsymbol{\Omega},0}, \boldsymbol{p}_1)$ decrease slightly. One possible explanation

is that we fix n and q, then $\widetilde{\boldsymbol{\Sigma}}_{\text{prop.}}$ is determined by $\widetilde{\mathbf{A}}_1$, $\widetilde{\mathbf{B}}_1$, and $\boldsymbol{p}_1$ while $\widetilde{\boldsymbol{\Omega}}_{\text{prop.}}$ is determined by $\widetilde{\mathbf{A}}_1^{-1}$, $\widetilde{\mathbf{B}}_1$, $(\widetilde{\mathbf{A}}_1 + \widetilde{\mathbf{B}}_1 \times \mathbf{P}_{0,1})^{-1}$, and $\boldsymbol{p}_1$.

		$p = 150$				$p = 225$				$p = 300$			
n		bias	MCSD	ASE	CP	bias	MCSD	ASE	CP	bias	MCSD	ASE	CP
50	$a_{0,11}$	0	0.1	0.1	95.5	0	0	0	95.2	0	0	0	95.7
	$a_{0,22}$	0	0.8	0.8	93.9	0	0.7	0.7	95	0	0.6	0.6	94.7
	$a_{0,33}$	0.1	2.8	2.8	95.7	0.1	2.3	2.3	95.3	0	1.9	2	96.3
	$a_{0,44}$	0	0.3	0.3	94.8	0	0.2	0.2	95.3	0	0.2	0.2	95.4
	$a_{0,55}$	0	0.4	0.4	94.6	0	0.3	0.3	94.7	0	0.3	0.3	94.3
	$b_{0,11}$	-1.7	139.4	135.7	91.4	0.8	135.3	136.2	93.1	-0.4	134.1	135.9	93.2
	$b_{0,12}$	6	86.9	87.5	94.8	4.3	85.7	87.2	95.1	1	86.2	88.2	95
	$b_{0,13}$	5.6	79.4	78	95.4	3.5	77.7	77.9	95.9	2.3	75.9	78	96.9
	$b_{0,14}$	-2.6	110.3	105.1	91.8	-0.8	104	105.3	93.7	-0.1	103.6	104.8	95
	$b_{0,15}$	-2	80.7	78.2	93.5	1.9	78.1	78.4	94	-3.3	77.6	78.4	93.7
	$b_{0,22}$	-0.9	105.7	105.3	93.2	-7.4	105.7	104	91.2	5.7	106.3	106.6	93.9
	$b_{0,23}$	-0.1	88.6	87.1	92.7	-3.1	86.6	86.3	93.3	4.1	89.5	87.8	93.7
	$b_{0,24}$	1.3	85.6	86.1	95.4	-2.7	86.2	85.8	95.3	-0.4	88.7	86.4	95.2
	$b_{0,25}$	3.2	66.1	65.5	95.7	-3.3	63.4	64.9	95.4	2.9	67.9	66.1	95.7
	$b_{0,33}$	0.7	88.1	88.1	93.6	-0.5	86.6	87.7	93.3	3.3	90.1	88.3	93.5
	$b_{0,34}$	0.4	91.6	91.1	94.4	-2.6	91.3	91.3	93	1.2	91.4	90.9	93.5
	$b_{0,35}$	2.9	61.4	59.3	95.6	-0.9	58.1	59.1	95.8	1.9	60.2	59.5	96.3
	$b_{0,44}$	0.1	140.5	137.2	92.1	2.6	137	137.7	94.2	-4.9	132.2	136.2	93.1
	$b_{0,45}$	-0.5	78.6	74	94.2	-1.5	73.5	73.9	96.2	-2.1	73.8	73.8	96
	$b_{0,55}$	-1.6	80.7	79.6	92.7	-2.2	81.4	79.5	92.7	0.4	80.4	80	93
100	$a_{0,11}$	0	0	0	95.1	0	0	0	94.9	0	0	0	95.2
	$a_{0,22}$	0	0.6	0.6	95.8	0	0.4	0.5	96.2	0	0.4	0.4	94.8
	$a_{0,33}$	0	2.1	2	94.2	0	1.6	1.6	94.9	0.1	1.4	1.4	95.3
	$a_{0,44}$	0	0.2	0.2	96.3	0	0.1	0.1	95.5	0	0.1	0.1	95.4
	$a_{0,55}$	0	0.3	0.3	95.3	0	0.2	0.2	95.1	0	0.2	0.2	95.3
	$b_{0,11}$	0.3	97.6	95.7	93.5	7.1	97.3	96.7	93.5	-1.1	93.7	95.5	94.6
	$b_{0,12}$	1.2	59.6	61.8	95.4	-2.1	61.3	62.4	95.8	2.3	62.5	61.9	94.7
	$b_{0,13}$	1.1	52.2	54.7	96.8	-0.4	54.3	55	95.6	1.2	55	54.8	94.9
	$b_{0,14}$	-0.4	72.5	74.1	95.4	-3.1	75.4	74.3	94.3	-0.2	74.5	73.9	94.8
	$b_{0,15}$	-0.8	55.6	55.2	93.8	3.7	56	55.9	96	0.7	56.8	55.3	94.8
	$b_{0,22}$	-1.6	71.5	74	95.1	2.3	73.4	74.5	94	1.8	76.5	74.4	94
	$b_{0,23}$	-2.3	58.3	61	94.1	-0.1	60.6	61.3	94.2	1.5	64.3	61.4	93.8
	$b_{0,24}$	1.7	59.9	60.6	95.3	3.3	61.7	60.6	95.5	-0.1	61.8	60.7	94.5
	$b_{0,25}$	1.8	45.4	46.1	95.6	1.9	48.3	46.6	94.8	1.7	46.2	46.3	95
	$b_{0,33}$	-2.6	58.9	61.5	95.3	-1.6	60.7	61.5	93.9	0.5	64	61.8	94.1
	$b_{0,34}$	1.4	63.3	64	94.9	3.5	62.7	63.7	94.8	0.6	66.7	64	93.5
	$b_{0,35}$	1.2	39.8	41.6	96.2	1.5	43.9	41.9	94.8	2	41.9	41.7	95.4
	$b_{0,44}$	0.8	93.1	96.6	96.3	-3.9	93.8	95.9	94.7	-0.8	98.8	96.4	93
	$b_{0,45}$	0	50.2	52.1	96	-1.9	54.5	52.2	93.3	-1.2	53.9	52	94.7
	$b_{0,55}$	-0.8	57.1	56.1	94	4.1	56.7	56.8	94.9	0	56.6	56.2	93.9
150	$a_{0,11}$	0	0	0	95	0	0	0	96	0	0	0	94.6
	$a_{0,22}$	0	0.5	0.5	93.7	0	0.4	0.4	95.5	0	0.3	0.3	94.1
	$a_{0,33}$	0	1.6	1.6	95.1	0	1.3	1.3	93.6	0	1.1	1.1	96.3
	$a_{0,44}$	0	0.1	0.1	94.7	0	0.1	0.1	95.1	0	0.1	0.1	96
	$a_{0,55}$	0	0.2	0.2	95.3	0	0.2	0.2	95.2	0	0.2	0.2	94.4
	$b_{0,11}$	2.6	75.1	78.3	95.6	-2.8	77.3	77.7	94.3	-5.8	76.5	77.3	94.3
	$b_{0,12}$	1.1	49.7	50.6	95.1	1.3	49.4	50.2	95.1	1.1	51.6	50.3	93.7
	$b_{0,13}$	2.3	44.1	44.9	96.1	1	44.4	44.7	96.2	-1.5	45.9	44.4	94
	$b_{0,14}$	-3.6	58.3	60.6	95.9	-0.5	61.9	60.4	94.2	4.6	60.3	59.8	93.7
	$b_{0,15}$	0.6	46.1	45.1	94.5	-1.6	45.9	44.9	94.9	-1.4	44.9	44.9	93.7
	$b_{0,22}$	1.1	61.7	60.6	95	-2.7	60.5	60.2	93.6	1.2	61.7	60.6	93.9
	$b_{0,23}$	1.1	51.4	50.1	94.6	-0.8	49.5	49.9	94.2	0.9	50	50	94.2
	$b_{0,24}$	-0.5	48.6	49.6	95.8	-2	48.3	49.5	95.3	-1.2	49.5	49.4	94.4
	$b_{0,25}$	1.4	36.7	37.7	95.2	-1.1	37.2	37.5	94.5	2.1	38.7	37.8	95.6
	$b_{0,33}$	1.1	52.1	50.6	94.4	1.3	49.5	50.5	94.7	-0.2	49.3	50.3	94.4
	$b_{0,34}$	-1	52.4	52.4	94.9	-3.6	51.5	52.6	95.1	1	51.4	52.1	94.4
	$b_{0,35}$	2.3	33.2	34.1	95.6	-1.8	33.7	34.1	95.3	1	35.5	34	94.1
	$b_{0,44}$	1.5	77.2	78.8	95	3.2	81.2	79	93.9	-3.9	79.7	78.2	93.2
	$b_{0,45}$	-2.6	42.3	42.5	95.5	2.7	43.1	42.5	94.5	0.4	42.9	42.4	95.2
	$b_{0,55}$	-0.6	46.9	45.8	93.4	-0.7	46.8	45.8	94.2	0.7	46.4	45.9	94.5

Table 3.1: Estimation results $(\times 100)$ for $a_{0,kk}$ and $b_{0,kk'}$ in Scenario 1 under $n = 50$, 100, or 150, where bias denotes the average bias, MCSD denotes the Monte Carlo standard deviation, ASE denotes the average standard errors, CP denotes the coverage probability based on a 95% Wald-type confidence interval.

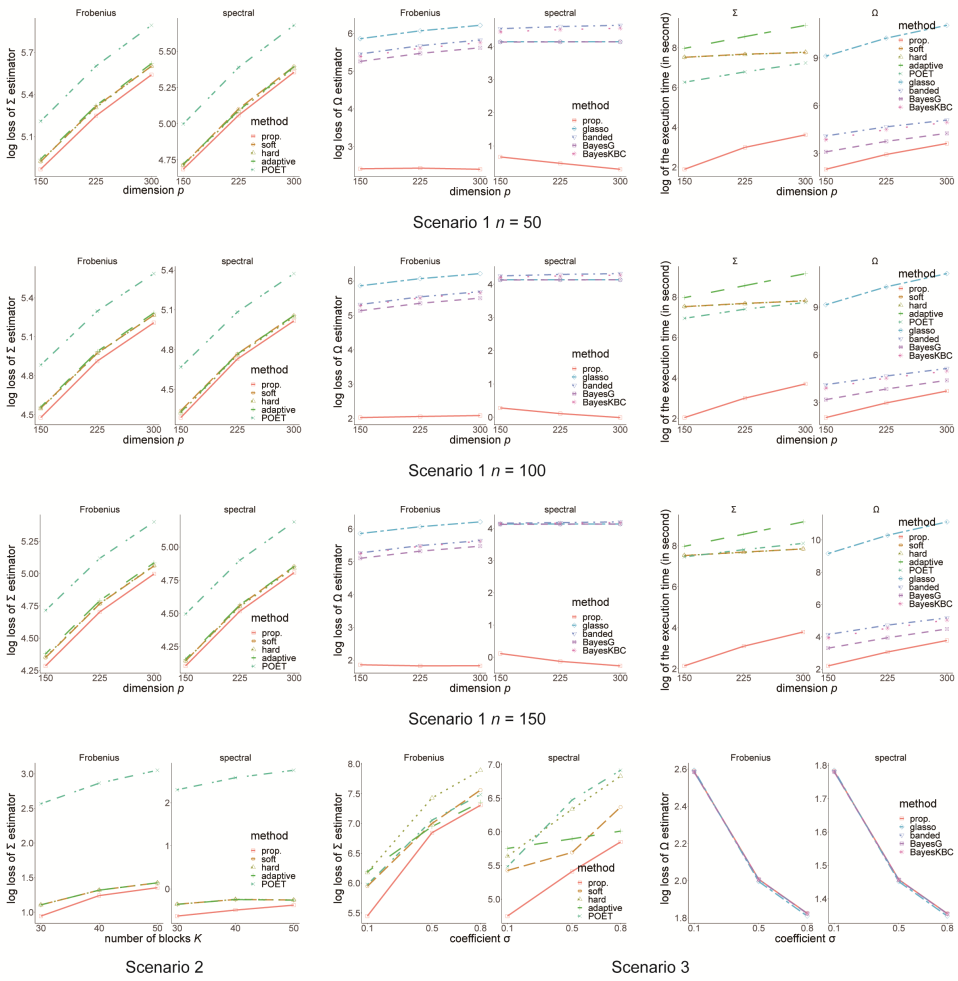


Figure 3.2: Results of the losses in the Frobenius and spectral norms and the execution time for the proposed covariance- and precision-matrix estimators and the conventional estimators in simulation studies. The proposed estimators outperform the competitors regarding fewer losses and shorter computational times.

3.3.3 Scenario 2: comparison for large K covariance matrix

The true covariance matrix is $\boldsymbol{\Sigma}_{0,2}(\mathbf{A}_{0,K}, \mathbf{B}_{0,K}, \boldsymbol{p}_2) = \mathbf{A}_{0,K} \circ \mathbf{I}(\boldsymbol{p}_2) + \mathbf{B}_{0,K} \circ \mathbf{J}(\boldsymbol{p}_2)$, where $K = 30$, 40, or 50; $\boldsymbol{p}_2 \coloneqq (p_{\text{ind}}, \ldots, p_{\text{ind}})^\top = p_{\text{ind}} \times \mathbf{1}_{K \times 1}$ with the individual component $p_{\text{ind}} = 10$; $q = K + K(K+1)/2 = 495$, 860, or 1325; $p = K \times p_{\text{ind}} = 300$, 400, or 500; and $\mathbf{A}_{0,K}$ and $\mathbf{B}_{0,K}$ are generated depending on the value of K. For each K, we first generate a K by K diagonal matrix $\mathbf{A}_{0,K}$ and a symmetric matrix $\mathbf{B}_{0,K}$ satisfying $\boldsymbol{\Sigma}_{0,2}(\mathbf{A}_{0,K}, \mathbf{B}_{0,K}, \boldsymbol{p}_2) \succ 0$, or equivalently,

$\mathbf{A}_{0,K} \succ 0$ and $\mathbf{\Delta}_{0,K} \coloneqq \mathbf{A}_{0,K} + \mathbf{B}_{0,K} \times \mathbf{P}_{0,K}$ is invertible, where $\mathbf{P}_{0,K} \coloneqq \text{diag}(p_{\text{ind}}, \ldots, p_{\text{ind}}) \in \mathbb{R}^{K \times K}$. Then, we simulate the data matrix $\mathbf{X}$ based on an independently and identically distributed sample of size $n = 30$ from $N(\mathbf{0}_{p\times 1}, \mathbf{\Sigma}_{0,2}(\mathbf{A}_{0,K}, \mathbf{B}_{0,K}, \boldsymbol{p}_2))$, where all K, p, $q \geq n$. For each K, the above generation procedure is repeated 1000 times. To estimate $\mathbf{\Sigma}_{0,2}(\mathbf{A}_{0,K}, \mathbf{B}_{0,K}, \boldsymbol{p}_2)$, we adopt the modified hard-thresholding estimator in (3.2.11). The thresholding level λ is chosen by following a similar procedure to that in Bickel and Levina (2008a,b). We also compare our method with the estimation methods for large covariance matrices used in Scenario 1 (Section 3.3.2). For 1000 replicates, we plot the average losses in terms of matrix norms in Figure 3.2. The results in Figure 3.2 show that the proposed modified estimator produces the smallest loss by taking advantage of the underlying structure.

3.3.4 Scenario 3: simulation analysis under model misspecification

In this scenario, we assess the performance of the proposed covariance- and precision-matrix estimators under the model misspecification when the true covariance matrix does not have a uniform-block structure. Specifically, we set the true covariance matrix $\mathbf{\Upsilon}_{0,\sigma} \coloneqq \mathbf{\Sigma}_{0,3}(\mathbf{A}_0, \mathbf{B}_0, \boldsymbol{p}_3) + \mathbf{M}_\sigma \succ 0$, where $K = 5$; $\boldsymbol{p}_3 \coloneqq p_{\text{ind}} \times \mathbf{1}_{K\times 1}$ with $p_{\text{ind}} = 30$; $q = 20$; $p = K \times p_{\text{ind}} = 150$; and $\mathbf{A}_0$ and $\mathbf{B}_0$ are identical with those in Scenario 1 (Section 3.3.2). Matrix $\mathbf{M}_\sigma$ follows a Wishart distribution with p degrees of freedom and parameter $\sigma\mathbf{I}_p$, where $\sigma = 0.1$, 0.5, or 0.8. It is clear that if $\sigma = 0$, then $\mathbf{\Upsilon}_{0,\sigma} = \mathbf{\Sigma}_{0,3}(\mathbf{A}_0, \mathbf{B}_0, \boldsymbol{p}_3)$ is a uniform-block matrix; if $\sigma > 0$, then $\mathbf{\Upsilon}_{0,\sigma}$ and the true precision matrix $\mathbf{\Upsilon}_{0,\sigma}^{-1}$ are not uniform-block matrices because the uniformity does not hold. For each σ, we generate $\mathbf{X}$ based on a random sample with a size of $n = 50$ (satisfying $q = 20 < n$) by drawing from $N(\mathbf{0}_{p\times 1}, \mathbf{\Upsilon}_{0,\sigma})$. We fit the data using the covariance-matrix estimator in (3.2.8),

the precision-matrix estimator in (3.2.9), and the other estimators for large covariance and precision matrices described in Scenario 1 (Section 3.3.2). We simulate 1000 replicates for each σ. The average losses $\|\widetilde{\mathbf{\Sigma}}_* - \mathbf{\Upsilon}_{0,\sigma}\|_{\mathrm{F}}$, $\|\widetilde{\mathbf{\Sigma}}_* - \mathbf{\Upsilon}_{0,\sigma}\|_{\mathrm{S}}$, $\|\widetilde{\mathbf{\Omega}}_* - \mathbf{\Upsilon}_{0,\sigma}^{-1}\|_{\mathrm{F}}$, and $\|\widetilde{\mathbf{\Omega}}_* - \mathbf{\Upsilon}_{0,\sigma}^{-1}\|_{\mathrm{S}}$ are calculated among 1000 replicates for the method $*$. The results are plotted in Figure 3.2.

The results in Figure 3.2 show that the proposed covariance-matrix estimator works well under the misspecified covariance structure. Compared with the traditional covariance-matrix estimators, the losses in terms of matrix norms are smaller for our proposed method. The losses in terms of matrix norms for our precision-matrix estimator are comparable with those for the other methods. One possible reason is that the inverse of a non-uniform-block matrix is not a uniform-block matrix so the proposed precision-matrix estimator cannot take benefit from the underlying structure.

In summary, our method can robustly and accurately estimate covariance matrices with the dependence structure of interconnected communities. Since recent multiple testing correction methods, e.g., to control the false discovery proportion (FDP; Fan et al., 2012; Fan and Han, 2017), are based on the covariance matrix estimate, we also evaluated the influence of covariance estimation on the accuracy of feature selection (see the extra simulation study in Chapter B). Our simulation results showed that our approach can largely improve the sensitivity while preserving the FDP compared to competing methods.

3.4 Data Examples

3.4.1 Proteomics data analysis

We apply the proposed method to estimate the covariance matrix of high-throughput proteomics data for cancer research (Yildiz et al., 2007). Specifically, 288 subjects (180 male and 108 female, aged 62.4 ± 9.4 years) participated in this case–control study. Matrix-assisted laser desorption ionization mass spectrometry was utilized to identify the abundant peptides in human serum between case and control samples. After preprocessing (Chen et al., 2009), 184 features in the serum were selected as candidate proteins and peptides. Of these, 107 were identified by the network detection algorithm (Chen et al., 2018) and used to form a sample correlation matrix with 7 by 7 well-organized blocks (see the first two heat maps of Figure 3.3(A)). Our aim was to estimate the large correlation matrix and to understand the interactive relations between these 107 features.

We focus on the $p = 107$ features in the $K = 7$ communities for the $n = 288$ participants, therefore $q = 35$. The partition-size vector $\boldsymbol{p} = (34, 18, 14, 14, 13, 10, 4)^\top$ was provided by the network detection algorithm (Chen et al., 2018). Let $\mathbf{A}_0 = \mathrm{diag}(a_{0,11}, \ldots, a_{0,KK})$ and $\mathbf{B}_0 = (b_{0,kk'})$ with $b_{0,kk'} = b_{0,k'k}$ for $k \neq k'$ denote the K by K unknown diagonal matrix and symmetric matrix, respectively. We assume that the p by p population correlation matrix is a uniform-block matrix whose diagonal entries are 1 (that is, $a_{0,kk} + b_{0,kk} = 1$ for every k), expressed by $\mathbf{R}_0(\mathbf{A}_0, \mathbf{B}_0, \boldsymbol{p})$. Since $q = 35 < n = 288$, the estimates and standard errors of $a_{0,kk}$ and $b_{0,kk'}$ can be obtained by the best unbiased estimators (3.2.7) and Corollary 3.2.3, respectively. We summarize the results in the third and fourth plots of Figure 3.3(A), where the sum of the

estimates of $a_{0,kk}$ and $b_{0,kk}$ is equal to 1 for every k because the diagonal entries in the sample correlation matrix are 1. The fourth plot of Figure 3.3(A) shows that the 95% confidence intervals of the correlations between the $(1,2)$, $(2,4)$, and $(6,7)$ blocks contain 0. The $(1,3)$, $(1,4)$, $(1,5)$, $(1,6)$, $(2,3)$, $(2,5)$, $(3,4)$, $(3,5)$, $(3,6)$, $(3,7)$, and $(5,7)$ blocks have negative correlations while the remaining blocks have positive correlations.

3.4.2 Brain imaging data analysis

The second example is a brain imaging study based on echo-planar spectroscopic imaging, which can simultaneously measure multiple neurometabolites in the regions of a whole brain (Chiappelli et al., 2019). The data were collected from 78 participants (39 male and 39 female, aged 42.1 ± 18.8 years). Five neurometabolites were identified in 89 brain regions: choline, myo-inositol, creatine-containing compounds, N-acetylaspartate, and glutamate–glutamine. We first estimated the sample correlation matrix for the combinations of the neurometabolites and brain regions (that is, $445 = 5 \times 89$) and applied the approach developed by Chen et al. (2018) to extract a structure for the latent well-organized blocks with 227 combinations from the possible 445 (see the first heat map of Figure 3.3(B)). This structure has five diagonal blocks and 10 off-diagonal blocks (see the second heat map of Figure 3.3(B)). We next applied the proposed method to estimate the correlation matrix.

For this application, we obtained $n = 78$ and $\boldsymbol{p} = (77, 49, 36, 33, 32)^\top$ from the preliminary network detection algorithm (Chen et al., 2018), so $K = 5$, $p = 227$, and $q = 20$. We estimated $a_{0,kk}$ and $b_{0,kk'}$ with (3.2.7) and their standard errors with Corollary 3.2.3. The results are plotted in the third and fourth plots of Figure 3.3(B). The third plot of Figure 3.3(B)

shows all the estimated correlations. The correlations within all diagonal blocks and those between the $(1, 2)$, $(1, 4)$, $(2, 4)$, and $(3, 5)$ blocks are positive with the 95% confidence intervals not containing 0. The correlations between the $(1, 3)$, $(1, 5)$, and $(2, 5)$ blocks are negative with the 95% confidence intervals not containing 0. The $(2, 3)$, $(3, 4)$, and $(4, 5)$ the blocks have the 95% confidence intervals for the correlations containing 0. The five diagonal blocks are as follows: (1) 74 regions with choline (including three regions with other metabolites), (2) 49 regions with myo-inositol, (3) 26 regions with N-acetylaspartate (including 10 regions with glutamate–glutamine), (4) 33 regions with creatine-containing compounds, and (5) another 32 regions with N-acetylaspartate. The off-diagonal blocks show the positive relations among choline, myo-inositol, and creatine-containing compounds, and also between N-acetylaspartate and glutamate–glutamine across the brain. There is also a global negative correlation relation between two sets of metabolites across the brain: (1) choline, myo-inositol, and creatine-containing compounds and (2) glutamate–glutamine and N-acetylaspartate. The region-level correlations may assist in providing an understanding of the neurophysiological mechanisms relating to metabolites in the central nervous system.

3.5 Discussion

We have developed a computationally efficient method for estimating large covariance and precision matrices with a uniform-block structure. In our empirical analyses of multiple types of high-throughput biomedical data, we found that most of these datasets, including gene expression, proteomics, neuroimaging, and exposome data, among many others, exhibit a latent yet well-organized block pattern (see the examples in Figure 3.1 and Chapter B). By leverag-

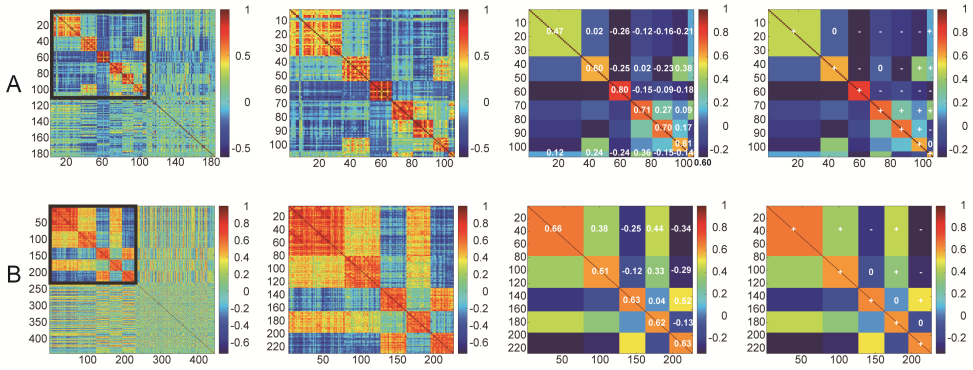


Figure 3.3: A: the first two are the heat maps of the sample correlation matrices for the proteomics dataset, the third one exhibits the estimates of $b_{0,kk'}$ (their standard errors are between 0.03 and 0.07; for each k, the estimate of $a_{0,kk}$ is one minus the estimate of $b_{0,kk}$ with standard errors around 0.01), and the fourth one contains the confidence intervals for $b_{0,kk'}$ ($+$ refers to the 95% confidence interval on the right of 0, $-$ refers to the 95% confidence interval on the left of 0, 0 refers to the 95% confidence interval containing 0). B: the first two are the heat maps of the sample correlation matrices for the brain imaging dataset, the third one exhibits the estimates of $b_{0,kk'}$ (their s.e. are between 0.07 and 0.11; for each k, the estimate of $a_{0,kk}$ is equal to one minus the estimate of $b_{0,kk}$ with standard errors around 0.01), and the fourth one contains the confidence intervals for $b_{0,kk'}$ ($+$ refers to the 95% confidence interval on the right of 0, $-$ refers to the 95% confidence interval on the left of 0, 0 refers to the 95% confidence interval containing 0).

ing the uniform-block structure, we provide an accurate estimate of the parameter vector for a large covariance matrix with a drastically reduced number of parameters. We further derive the covariance- and precision-matrix estimators in closed forms, which significantly reduce the computational burden and improve the accuracy of statistical inference.

In a uniform-block structure, we assign one parameter for the diagonal entries and one parameter for the off-diagonal entries in a diagonal block, and one parameter for all entries in an off-diagonal block. This parameterization strategy is driven by the fact that the intra-block and the inter-block variances are relatively small in a large sample correlation matrix (see the box plots in Chapter B). Given the strong block patterns in the correlation matrices (see the examples in Figure 3.1 and Chapter B), this parameterization strategy appears valid. It is also analogous to the commonly used compound symmetry covariance structure in linear mixed-effect models and

generalized estimating equation models. In addition, we demonstrated that our method generally performs well when there are multiple parameters within each block, as in Scenario 3 (Section 3.3.4). Therefore, our approach is, overall, robust and fast for high-dimensional biomedical data as it considers the commonly observed block-wise covariance structures.

Chapter 4: Modeling Multivariate Outcomes with Interconnected Modules: Evaluating the Impact of Alcohol Intake on Plasma Metabolomics

4.1 Introduction

Simultaneous measurement of hundreds of thousands of biological features has revealed complex scientific mechanisms across various fields (He et al., 2019; Ke et al., 2022). Numerous statistical methodologies have been developed to address the challenges associated with high-dimensional data analysis. For instance, shrinkage and penalty techniques have been employed in linear regression models, leading to the establishment of theoretical properties for the sparse estimators (Tibshirani, 1996; Efron et al., 2004; Hastie et al., 2015). Similarly, analogous regularization methods have successfully extended to estimate large-scale covariance or precision matrices while incorporating assumptions of bandability, sparsity, or low-rank structures (Wu and Pourahmadi, 2003; Bickel and Levina, 2008a,b; Bickel and Gel, 2011; Fan et al., 2011, 2013). From an inference perspective, multiplicity-adjusted procedures have been proposed to enable the simultaneous testing of numerous hypotheses, irrespective of whether the underlying test statistics are independent (Benjamini and Hochberg, 1995; Storey et al., 2004; Efron, 2004) or heavily correlated (Benjamini and Yekutieli, 2001; Efron, 2007; Leek and Storey, 2008; Fan et al., 2012; Fan and Han, 2017).

Conventionally, statistical methods for handling high-dimensional variables can be broadly categorized into two classes. In the first class, high-dimensional variables are considered as predictors. For example, regression shrinkage methods, e.g., the lasso (Tibshirani, 1996) and its many variants, are commonly used to select the variables of interest (Yuan and Lin, 2007; Fan and Lv, 2008; He et al., 2019). In the second class, high-dimensional variables are treated as multivariate outcomes in regression models. For example, high-dimensional imaging data are often modeled as outcomes while spatial dependence is taken into account. In addition, multiple testing methods widely used in omics data analysis can be categorized into this class (Leek and Storey, 2008; Fan et al., 2012; Fan and Han, 2017). In this chapter, our focus lies within the second category of data analysis, as we aim to investigate the effect of alcohol intake on metabolomic profiles. Specifically, our goal is to incorporate the structured dependence among the multivariate outcomes into exposures or predictors of interest.

The challenge of incorporating the dependence structure into a multivariate regression model is twofold. First, the pattern of dependence structure can be latent and complex, which implies that we need to detect or estimate this structure before modeling. For example, gene co-expression network analysis (Wu et al., 2021) and seed quality analysis (Perrot-Dockès et al., 2022). On the other hand, given a detected dependence structure of data, it is also challenging to utilize the pattern in a multivariate regression model, since the covariance structure of (marginal) outcomes is implicitly influenced by the detected pattern. In our empirical analyses of various high-dimensional datasets, including gene expression, proteomics, brain imaging, and seed quality, among others in Chapter 3, a particular structure, called the *interconnected community structure*, is most prevalent (please see a real example in Figure 4.1(C)). In particular, the pattern has several characteristics: it is *latent*, i.e., it is the output of applying network detection

algorithms to the raw data; it is *non-sparse*, i.e., elements of the covariance or correlation matrix have small but non-zero values; it has an *almost constant-valued* block form, i.e., elements of each block are almost the same, exhibiting low variability; and it may have many *singletons* or isolated nodes (see a real example with the singletons in Figure 4.1(B)). But, as we discussed, due to the existence of this structured covariance or correlation matrix in data, it is essential to incorporate the corresponding structured dependence into a statistical model to obtain reliable and accurate results. A few questions naturally arise: given this pattern in the covariance or correlation structure, (1) does the matrix consisting of all dependence parameters have a specific structure? (2) what is the relationship between the dependence structure and the covariance or correlation structure at the population level? (3) how can we estimate the dependence parameters and the other parameters, e.g., the regression coefficients, simultaneously?

To address these questions, we propose a *Multivariate Autoregressive regression model with Uniform-block Dependence* (MAUD) that incorporates the above latent dependence structure found in data into a parametric regression model, and we furthermore develop an estimation procedure for the dependence parameters and regression coefficients and establish the finite- and large-sample properties of the proposed estimators.

The MAUD makes at least three contributions. First, biological features in outcomes are divided into communities (or groups, or clusters) based on a preliminary study (e.g., a network detection algorithm), and the dependence parameters both within and between communities are defined in the MAUD. We notice that the dependence structure is naturally a *uniform-block* structure (see Chapter 2) rather than an arbitrary one. A uniform-block structure brings two advantages. One is that structures preserved under common algebraic operations, i.e., the corresponding covariance, correlation, and precision matrices also have uniform-block structures with explicit

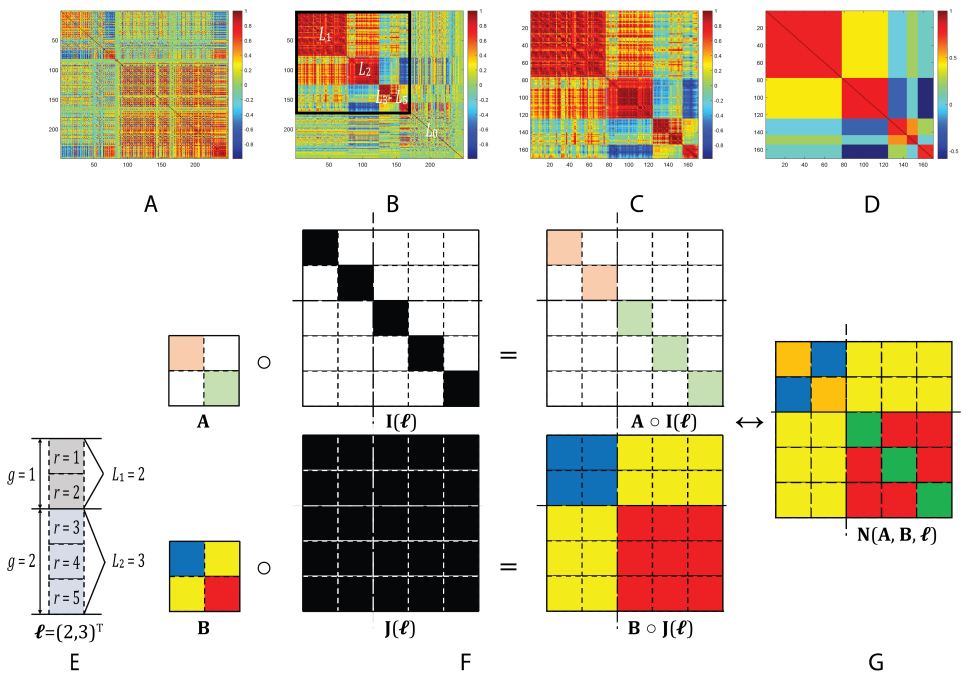


Figure 4.1: A: 249 by 249 sample correlation matrix (of residuals) calculated from a subset of raw NMR data (Ritchie et al., 2023); B: an interconnected community structure with generalized (with singletons) well-organized blocks in the sample correlation matrix provided by a community detection algorithm (Chen et al., 2023); C: a 170 by 170 sample correlation matrix (of residuals) extracted from the black frame in B, which is called an interconnected community structure with well-organized blocks; D: a population correlation matrix with an interconnected community structure with (5 by 5) uniform blocks; E: a partition-size vector $\boldsymbol{\ell} = (L_1, L_2)^\top = (2, 3)^\top$ with $R = 5$ and $G = 2$; F: an illustration of the block Hadamard product representation $\mathbf{N}(\mathbf{A}, \mathbf{B}, \boldsymbol{\ell}) = \mathbf{A} \circ \mathbf{I}(\boldsymbol{\ell}) + \mathbf{B} \circ \mathbf{J}(\boldsymbol{\ell})$; G: an illustration of a uniform-block matrix $\mathbf{N}(\mathbf{A}, \mathbf{B}, \boldsymbol{\ell})$, where the cells with different colors in the matrices represent the elements with different values, the cells with different colors in $\boldsymbol{\ell}$ represent the features in different communities.

expressions in terms of the dependence parameters. The other one is that computations become efficient, i.e., computations of uniform-block matrices are simplified in the estimation procedure due to both dimensionality reduction and closed-form expressions. The second contribution is that the regression coefficient estimator is derived by the ordinary least-squares (OLS) method, which is identical to the generalized least-square (GLS) estimator and the feasible generalized least-square (FGLS) estimator, so it has an exact covariance matrix. Simultaneously, the dependence parameter estimators are derived by the maximum likelihood (ML) method and have an

asymptotic covariance matrix in closed form. The third contribution is that the covariance matrix of the regression coefficient estimator is exact and explicit in terms of the dependence parameters and the design matrix. In summary, compared to the conventional models that use a diagonal or block-diagonal structure to model dependence in network data (i.e., ignoring the non-null off-diagonal elements), the proposed MAUD offers more accurate estimates and inferences.

This chapter continues with Section 4.2, which is concerned with our methodology of dependence modeling, estimation, and statistical inference: Section 4.2.1 introduces a simultaneous autoregressive regression model for multivariate outcomes; Section 4.2.2 specifies the multivariate autoregressive regression model with a uniform-block structure and establishes the relationship between the dependence parameters and the covariance matrix; Section 4.2.3 describes the estimation procedure with theoretical properties; Section 4.2.4 presents the hypothesis tests for the regression coefficients and dependence parameters. In Section 4.3, we conduct simulation studies to evaluate the proposed methodology. We then apply the proposed method to real data in Section 4.4 and provide a discussion in Section 4.5. All proofs, additional figures, and tables are given in Chapter C.

4.2 Methodology

4.2.1 A simultaneous autoregressive regression model

Let $\mathbf{Y}^* = (\mathbf{y}_1^*, \ldots, \mathbf{y}_n^*) \in \mathbb{R}^{S \times n}$ denote the data matrix, where $\mathbf{y}_i^* = (y_{i1}, \ldots, y_{iS})^\top \in \mathbb{R}^{S \times 1}$ denotes the S-dimensional multivariate outcomes of features for a participant $i = 1, \ldots, n$. Let $\mathbf{y}^{(r)} \in \mathbb{R}^{1 \times n}$ denote the r-th row vector of $\mathbf{Y}^*$. We assume that the dependence structure of the S-dimensional multivariate features reflects an interconnected community structure (see Fig-

ure 4.1(B)). Without loss of generality, the r-th feature (within the first R features) can be assigned to community g with $\phi : \{1, 2, \ldots, R\} \to \{1, \ldots, G\}, r \mapsto g$ as the community assigning function, i.e., $\phi(r) = g$. The interconnected community structure demonstrates that: (1) any two pairs of features r_1 and r_2, r_3 and r_4 within community g are coherently and positively correlated, i.e., $\text{cor}\left(\mathbf{y}^{(r_1)}, \mathbf{y}^{(r_2)}\right) \approx \text{cor}\left(\mathbf{y}^{(r_3)}, \mathbf{y}^{(r_4)}\right) > 0$ if $\phi(r_1) = \phi(r_2) = \phi(r_3) = \phi(r_4) = g$; (2) communities g and g' can further be *interconnected*, denoted by $g \sim g'$, if $\text{cor}\left(\mathbf{y}^{(r)}, \mathbf{y}^{(r')}\right) \neq 0$ for $\phi(r) = g$ and $\phi(r') = g'$; (3) as demonstrated in Figure 4.1(B), a subset of features (i.e., indexed by $\{R+1, \ldots, S\}$) belongs to none of the communities, for which we denote a generalized community assigning function $\phi^* : \{1, 2, \ldots, S\} \to \{0, 1, \ldots, G\}$ with $\phi^*(r) = \phi(r) = g$ if $r \in \{1, \ldots, R\}$ and $\phi^*(r) = 0$ if $r \in \{R+1, \ldots, S\}$. We further let L_g denote the cardinality of the community g for $g = 1, \ldots, G$ and L_0 refer to the number of features in no communities and $S = R + L_0$ with $R = L_1 + \cdots + L_G$. We assess the association between the multivariate outcomes $\mathbf{y}_i^*$ and covariate vector $\mathbf{x}_i \in \mathbb{R}^{p\times 1}$ for $i = 1, \ldots, n$ while accounting for the complex dependence structure by the following simultaneous autoregressive regression model:

$$y_{ir} = \boldsymbol{\beta}_r^\top \mathbf{x}_i + \underbrace{\frac{1}{L_g - 1} \sum_{r' \neq r : \phi(r') = \phi(r) = g} \rho_{gg} \left(y_{ir'} - \boldsymbol{\beta}_{r'}^\top \mathbf{x}_i\right)}_{\text{within community } g}$$
$$+ \sum_{g' : g' \sim g} \underbrace{\frac{1}{\sqrt{(L_g - 1)(L_{g'} - 1)}} \sum_{r'' : \phi(r'') = g'} \rho_{gg'} \left(y_{ir''} - \boldsymbol{\beta}_{r''}^\top \mathbf{x}_i\right)}_{\text{between communities } g \text{ and } g'} + \epsilon_{ir}, \tag{4.2.1}$$

where $y_{ir} \in \mathbb{R}$ denotes an outcome of r-th feature (e.g., a biomarker), $\boldsymbol{\beta}_r \in \mathbb{R}^{p\times 1}$ denotes the feature-specific regression coefficient vector, $\top$ denotes the transpose of a vector or matrix, $\mathbf{x}_i \in \mathbb{R}^{p\times 1}$ denotes the covariate vector across all features for the i-th participant (e.g., intercept,

age, and sex), $\rho_{gg} \in \mathbb{R}$ denotes the *autoregressive dependence parameter* within g-th community, $\rho_{gg'} = \rho_{g'g} \in \mathbb{R}$ denotes the autoregressive dependence parameter between g-th and g'-th communities, $\epsilon_{ir} \in \mathbb{R}$ denotes an independently and identically distributed (i.i.d.) error following $N(0, \omega_r)$, and $\omega_r > 0$ denotes the feature-specific variation for $i = 1, \ldots, n$ and $r = 1, \ldots, S$.

We note that model (4.2.1) reduces to a general linear model for feature r that belongs to none of the communities. Model (4.2.1) builds upon the conventional autoregressive model used in neuroimaging studies (Bowman, 2005; Derado et al., 2010; Risk et al., 2016; Lee et al., 2023) and it extends the model to a generalized form that includes community-wise dependence. If all autoregressive dependence parameters, both within and between communities, are equal to 0, then model (4.2.1) reduces to a general linear model (Worsley and Friston, 1995; Friston et al., 1995). If the autoregressive dependence parameters within communities $\rho_{gg} \neq 0$ but those between communities $\rho_{gg'} = 0$ for every $g \neq g'$, then model (4.2.1) resembles other existing models in the literature (Bowman, 2005; Derado et al., 2010; Lee et al., 2023).

4.2.2 The MAUD

Due to the complex dependence pattern, a matrix representation of (4.2.1) is required for parameter estimation. Without loss of generality, we focus on the features in communities only, because the features in no communities (i.e., $\phi^*(r) = 0$) can be modeled separately. We express the *Multivariate Autoregressive regression model with Uniform-block Dependence* (MAUD) as

follows:

$$\mathbf{y}_i = \begin{pmatrix} \boldsymbol{\beta}_1^\top \\ \vdots \\ \boldsymbol{\beta}_R^\top \end{pmatrix}_{R\times p} \mathbf{x}_i + \boldsymbol{\Upsilon} \times \left[\mathbf{y}_i - \begin{pmatrix} \boldsymbol{\beta}_1^\top \\ \vdots \\ \boldsymbol{\beta}_R^\top \end{pmatrix}_{R\times p} \mathbf{x}_i \right] + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \overset{\text{i.i.d.}}{\sim} N\left(\mathbf{0}_{R\times 1}, \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\right), \tag{4.2.2}$$

for $i = 1, \ldots, n$, where $\mathbf{y}_i = (y_{i1}, \ldots, y_{iR})^\top \in \mathbb{R}^{R\times 1}$; $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iR})^\top \in \mathbb{R}^{R\times 1}$; $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \mathrm{cov}(\boldsymbol{\epsilon}_i) = \mathrm{diag}\left(\omega_1, \ldots, \omega_R\right) \in \mathbb{R}^{R\times R}$ with the covariance matrix notation $\mathrm{cov}(\cdot)$ and the diagonal matrix notation $\mathrm{diag}(\cdot)$; the *autoregressive dependence matrix* $\boldsymbol{\Upsilon}$ is a partitioned matrix:

$$\boldsymbol{\Upsilon} = \begin{pmatrix} \boldsymbol{\Upsilon}_{11} & \boldsymbol{\Upsilon}_{12} & \ldots & \boldsymbol{\Upsilon}_{1G} \\ \boldsymbol{\Upsilon}_{21} & \boldsymbol{\Upsilon}_{22} & \ldots & \boldsymbol{\Upsilon}_{2G} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Upsilon}_{G1} & \boldsymbol{\Upsilon}_{G2} & \ldots & \boldsymbol{\Upsilon}_{GG} \end{pmatrix},$$

$$\boldsymbol{\Upsilon}_{gg} = \gamma_{gg}\left(\mathbf{J}_{L_g} - \mathbf{I}_{L_g}\right) \in \mathbb{R}^{L_g\times L_g}, \quad \gamma_{gg} = \frac{\rho_{gg}}{L_g - 1}, \qquad g' = g,$$

$$\boldsymbol{\Upsilon}_{gg'} = \gamma_{gg'}\mathbf{J}_{L_g\times L_{g'}} \in \mathbb{R}^{L_g\times L_{g'}}, \qquad \gamma_{gg'} = \gamma_{g'g} = \frac{\rho_{gg'}}{\sqrt{(L_g - 1)(L_{g'} - 1)}}, \quad g' \neq g,$$

$\gamma_{gg'}$ is the autoregressive dependence parameter $\rho_{gg'}$ scaled by a constant, $\mathbf{I}$ and $\mathbf{J}$ denote the identity matrix and matrix of ones, respectively. We can further present $\boldsymbol{\Upsilon}$ in terms of $\gamma_{gg'}$:

$$\boldsymbol{\Upsilon}\left(\mathbf{A}_{\boldsymbol{\Upsilon}}, \mathbf{B}_{\boldsymbol{\Upsilon}}, \boldsymbol{\ell}\right) = \mathbf{A}_{\boldsymbol{\Upsilon}} \circ \mathbf{I}(\boldsymbol{\ell}) + \mathbf{B}_{\boldsymbol{\Upsilon}} \circ \mathbf{J}(\boldsymbol{\ell}), \text{ with } \begin{cases} \mathbf{A}_{\boldsymbol{\Upsilon}} & = \mathrm{diag}\left(-\gamma_{11}, \ldots, -\gamma_{GG}\right) \\ \mathbf{B}_{\boldsymbol{\Upsilon}} & = (\gamma_{gg'}) \end{cases} \tag{4.2.3}$$

where $\mathbf{I}(\boldsymbol{\ell}) = \text{Bdiag}\left(\mathbf{I}_{L_1}, \ldots, \mathbf{I}_{L_G}\right)$ with the block diagonal notation $\text{Bdiag}(\cdot)$, $\mathbf{J}(\boldsymbol{\ell}) = \left(\mathbf{J}_{L_g \times L_{g'}}\right)$, and $\circ$ denotes the block Hadamard product satisfying $\mathbf{A}_{\boldsymbol{\Upsilon}} \circ \mathbf{I}(\boldsymbol{\ell}) = \text{Bdiag}\left(-\gamma_{11}\mathbf{I}_{L_1}, \ldots, -\gamma_{GG}\mathbf{I}_{L_G}\right)$ and $\mathbf{B}_{\boldsymbol{\Upsilon}} \circ \mathbf{J}(\boldsymbol{r}) = \left(\gamma_{kk'}\mathbf{J}_{L_g \times L_{g'}}\right)$. The above matrix $\boldsymbol{\Upsilon}(\mathbf{A}_{\boldsymbol{\Upsilon}}, \mathbf{B}_{\boldsymbol{\Upsilon}}, \boldsymbol{\ell})$ is also known as the *uniform-block* matrix (see Chapter 2) The autoregressive dependence parameter matrix $\boldsymbol{\Upsilon}$ is intrinsically linked with the covariance matrix $\boldsymbol{\Sigma}$ as below:

$$\mathbf{y}_i \sim N\left(\mathfrak{B}_{R\times p}\mathbf{x}_i, \boldsymbol{\Sigma}\right), \quad \boldsymbol{\Sigma} = \left(\mathbf{I}_R - \boldsymbol{\Upsilon}\right)^{-1}\left(\mathbf{I}_R - \boldsymbol{\Upsilon}\right)^{-1}, \tag{4.2.4}$$

for $i = 1, \ldots, n$, where $\mathfrak{B}^\top = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_R) \in \mathbb{R}^{p\times R}$ and we set $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \mathbf{I}_R$ (i.e., $\omega_r = 1$ for all r) in (4.2.2) for the sake of easy presentation. We present the case of $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} \neq \mathbf{I}_R$ in Chapter C. Suppose $\boldsymbol{\Sigma}$ in (4.2.4) is positive definite, denoted by $\boldsymbol{\Sigma} \succ 0$, we define $\boldsymbol{\Omega}$ as the precision matrix, i.e., $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$. We aim to reparametrize $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ by $\gamma_{gg'}$, and thus the parameters of $\boldsymbol{\Omega}$ in the likelihood function can be substituted by scaled autoregressive dependence parameters $\gamma_{gg'}$. Consequently, $\gamma_{gg'}$ can be directly estimated by maximizing the likelihood function. However, the connections, between $\gamma_{gg'}$ and $\boldsymbol{\Sigma}$, between $\gamma_{gg'}$ and $\boldsymbol{\Omega}$, are not explicit. We resort to Corollary 4.2.1 for the derivation.

Corollary 4.2.1 (express $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ by $\gamma_{gg'}$)**.** *Both $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ partitioned by $\boldsymbol{\ell}$ are uniform-block*

matrices, with the following block Hadamard product representations (see Chapter 2):

$$\boldsymbol{\Omega}\left(\boldsymbol{A}_{\boldsymbol{\Omega}}, \boldsymbol{B}_{\boldsymbol{\Omega}}, \boldsymbol{\ell}\right) = \boldsymbol{A}_{\boldsymbol{\Omega}} \circ \boldsymbol{I}(\boldsymbol{\ell}) + \boldsymbol{B}_{\boldsymbol{\Omega}} \circ \boldsymbol{J}(\boldsymbol{\ell}), \text{ with } \begin{cases} \boldsymbol{A}_{\boldsymbol{\Omega}} & = \left(\boldsymbol{I}_G - \boldsymbol{A}_{\boldsymbol{\Upsilon}}\right)^2 \\ \boldsymbol{B}_{\boldsymbol{\Omega}} & = -2\boldsymbol{B}_{\boldsymbol{\Upsilon}} + \boldsymbol{A}_{\boldsymbol{\Upsilon}}\boldsymbol{B}_{\boldsymbol{\Upsilon}} + \boldsymbol{B}_{\boldsymbol{\Upsilon}}\boldsymbol{A}_{\boldsymbol{\Upsilon}} + \boldsymbol{B}_{\boldsymbol{\Upsilon}}\boldsymbol{L}\boldsymbol{B}_{\boldsymbol{\Upsilon}} \end{cases}, \tag{4.2.5}$$

$$\boldsymbol{\Sigma}\left(\boldsymbol{A}_{\boldsymbol{\Sigma}}, \boldsymbol{B}_{\boldsymbol{\Sigma}}, \boldsymbol{\ell}\right) = \boldsymbol{A}_{\boldsymbol{\Sigma}} \circ \boldsymbol{I}(\boldsymbol{\ell}) + \boldsymbol{B}_{\boldsymbol{\Sigma}} \circ \boldsymbol{J}(\boldsymbol{\ell}), \text{ with } \begin{cases} \boldsymbol{A}_{\boldsymbol{\Sigma}} & = \boldsymbol{A}_{\boldsymbol{\Omega}}^{-1} \\ \boldsymbol{B}_{\boldsymbol{\Sigma}} & = -\boldsymbol{\Delta}_{\boldsymbol{\Omega}}^{-1}\boldsymbol{B}_{\boldsymbol{\Omega}}\boldsymbol{A}_{\boldsymbol{\Omega}}^{-1} \end{cases}, \tag{4.2.6}$$

where $\boldsymbol{A}_{\boldsymbol{\Upsilon}}$, $\boldsymbol{B}_{\boldsymbol{\Upsilon}}$ are defined in (4.2.3), $\boldsymbol{A}_{\boldsymbol{\Sigma}}$, $\boldsymbol{A}_{\boldsymbol{\Omega}}$ are diagonal matrices, $\boldsymbol{B}_{\boldsymbol{\Sigma}}$, $\boldsymbol{B}_{\boldsymbol{\Omega}}$ are symmetric matrices, $\boldsymbol{L} = \operatorname{diag}\left(L_1, \ldots, L_G\right) \in \mathbb{R}^{G\times G}$, and $\boldsymbol{\Delta}_{\boldsymbol{\Omega}} = \boldsymbol{A}_{\boldsymbol{\Omega}} + \boldsymbol{B}_{\boldsymbol{\Omega}}\boldsymbol{L}$.

Corollary 4.2.1 establishes the closed-form representation of $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ in terms of $\boldsymbol{\Upsilon}$ (or $\gamma_{gg'}$) by leveraging several derived properties of the block Hadamard product of the uniform-block matrices (see Chapter 2).

4.2.3 Estimation of the MAUD parameters

The goal of this section is to estimate the vector of scaled autoregressive dependence parameters $\boldsymbol{\gamma} = \left(\gamma_{11}, \ldots, \gamma_{1G}, \gamma_{22}, \ldots, \gamma_{2G}, \ldots, \gamma_{GG}\right)^\top \in \mathbb{R}^{G(G+1)/2\times 1}$ and the regression coefficient vector $\boldsymbol{\beta} = \operatorname{vec}\left(\mathfrak{B}^\top\right) \in \mathbb{R}^{(Rp)\times 1}$ simultaneously in the MAUD of form (4.2.4), where $\operatorname{vec}(\cdot)$ denotes the vector of columns of a matrix (Henderson and Searle, 1979). We further denote $\boldsymbol{\theta} = \left(\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top\right)^\top \in \mathbb{R}^{[(Rp)+G(G+1)/2]\times 1}$ as the vector consisting of all unknown parameters.

We consider the following multivariate normal distribution for the MAUD:

$$\mathbf{y}_{(nR)\times 1} \sim N\left(\mathbf{x}_{(nR)\times(Rp)}\boldsymbol{\beta}_{(Rp)\times 1}, \mathbf{I}_n \otimes \boldsymbol{\Sigma}\right),$$

where $\mathbf{y} = \left(\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top\right)^\top \in \mathbb{R}^{(nR)\times 1}$, $\mathbf{x} = \left(\left(\mathbf{I}_R \otimes \mathbf{x}_1^\top\right)^\top, \ldots, \left(\mathbf{I}_R \otimes \mathbf{x}_n^\top\right)^\top\right)^\top \in \mathbb{R}^{(nR)\times(Rp)}$, and $\otimes$ denotes the Kronecker product. Consequently, the likelihood function is

$$\ell_n\left(\boldsymbol{\theta}; \mathbf{x}, \mathbf{y}\right) = \frac{n}{2}\left[-R\log(2\pi) + \log\det\left(\boldsymbol{\Omega}\right) - \operatorname{tr}\left(\mathbf{S}\boldsymbol{\Omega}\right)\right], \tag{4.2.7}$$

where $\mathbf{S} = n^{-1}\sum_{i=1}^n \left(\mathbf{y}_i - \mathfrak{B}\mathbf{x}_i\right)\left(\mathbf{y}_i - \mathfrak{B}\mathbf{x}_i\right)^\top \in \mathbb{R}^{R\times R}$ is the residual matrix.

Typically, we adopt the feasible generalized least-square (FGLS) approach to estimate $\boldsymbol{\theta}$ iteratively. Specifically, at iteration $t \geq 1$, the FGLS estimator of $\boldsymbol{\beta}$ is straightforward by

$$\widehat{\boldsymbol{\beta}}^{(t)} = \left\{\mathbf{x}^\top\left[\mathbf{I}_n \otimes \widehat{\boldsymbol{\Omega}}^{(t-1)}\right]\mathbf{x}\right\}^{-1}\mathbf{x}^\top\left(\mathbf{I}_n \otimes \widehat{\boldsymbol{\Omega}}^{(t-1)}\right)\mathbf{y} \in \mathbb{R}^{(Rp)\times 1},$$

where $\widehat{\boldsymbol{\Omega}}^{(t-1)} = \widehat{\boldsymbol{\Omega}}\left(\widehat{\boldsymbol{\gamma}}^{(t-1)}\right)$ and $\widehat{\boldsymbol{\gamma}}^{(t-1)}$ are the estimators of $\boldsymbol{\Omega}$ and $\boldsymbol{\gamma}$ at the $(t-1)$-th iteration, respectively. Based on $\widehat{\boldsymbol{\beta}}^{(t)}$, we further update $\widehat{\boldsymbol{\gamma}}^{(t)}$ by maximizing the log-likelihood function (4.2.7) while plugging $\widehat{\boldsymbol{\beta}}^{(t)}$:

$$\widehat{\boldsymbol{\gamma}}^{(t)} \in \underset{\boldsymbol{\gamma}\in\Theta_{\boldsymbol{\gamma}}}{\operatorname{argmax}}\, \ell_n\left(\widehat{\boldsymbol{\beta}}^{(t)}, \boldsymbol{\gamma}; \mathbf{x}, \mathbf{y}\right),$$

where $\Theta_{\boldsymbol{\gamma}} \subset \mathbb{R}^{G(G+1)/2\times 1}$ is the parameter space of $\boldsymbol{\gamma}$.

The above estimation of $\boldsymbol{\gamma}$ by optimizing of (4.2.7) is challenging because estimating the large covariance or precision matrix is essentially difficult (Fan et al., 2016; Cai et al., 2016).

On the other hand, the advanced methods for large covariance or precision matrix estimation (e.g., shrinkage and thresholding approaches) are not applicable, because they may yield biased estimate of autoregressive dependence parameters in the model (i.e., as constraints, the diagonals of each $\boldsymbol{\Upsilon}_{gg}$ are 0). Therefore, Corollary 4.2.1 plays a critical role in the MAUD autoregressive dependence parameter estimation using the likelihood approach. Specifically, plugging $\widehat{\boldsymbol{\beta}}^{(t)}$, we rewrite the likelihood function in (4.2.7) as below:

$$\ell_n\left(\widehat{\boldsymbol{\beta}}^{(t)}, \boldsymbol{\gamma}; \mathbf{x}, \mathbf{y}\right) = \frac{n}{2}\Bigg\{ - R\log(2\pi) + \sum_{g=1}^{G}(L_g - 1)\log\left(a_{\boldsymbol{\Omega},gg}\right) + \log\det\left(\boldsymbol{\Delta}_{\boldsymbol{\Omega}}\right) \\ - \operatorname{sum}\left[\mathbf{A}_{\boldsymbol{\Omega}} \odot \operatorname{diag}\left(\operatorname{tr}\left(\mathbf{S}_{11}^{(t)}\right), \ldots, \operatorname{tr}\left(\mathbf{S}_{GG}^{(t)}\right)\right) + \mathbf{B}_{\boldsymbol{\Omega}} \odot \left(\operatorname{sum}\left(\mathbf{S}_{gg'}^{(t)}\right)\right)\right]\Bigg\}, \tag{4.2.8}$$

where $a_{\boldsymbol{\Omega},gg}$ is the (g,g)-th diagonal element of $\mathbf{A}_{\boldsymbol{\Omega}} = \mathbf{A}_{\boldsymbol{\Omega}}(\boldsymbol{\gamma}) \in \mathbb{R}^{G\times G}$, expressed in terms of $\boldsymbol{\gamma}$ in (4.2.5); $\mathbf{B}_{\boldsymbol{\Omega}} = \mathbf{B}_{\boldsymbol{\Omega}}(\boldsymbol{\gamma}) \in \mathbb{R}^{G\times G}$, $\boldsymbol{\Delta}_{\boldsymbol{\Omega}} = \boldsymbol{\Delta}_{\boldsymbol{\Omega}}(\boldsymbol{\gamma}) = \mathbf{A}_{\boldsymbol{\Omega}}(\boldsymbol{\gamma}) + \mathbf{B}_{\boldsymbol{\Omega}}(\boldsymbol{\gamma})\mathbf{L} \in \mathbb{R}^{G\times G}$ with $\mathbf{L} = \operatorname{diag}(L_1, \ldots, L_G)$, also expressed in terms of $\boldsymbol{\gamma}$ in (4.2.5); $\left(\mathbf{S}_{gg'}^{(t)}\right)$ is the G by G partitioned matrix of $\mathbf{S}^{(t)}$ by $\boldsymbol{\ell}$, $\operatorname{sum}(\cdot)$ denotes the sum of all elements of a matrix, $\left(\operatorname{sum}\left(\mathbf{S}_{gg'}^{(t)}\right)\right) = \left(\operatorname{sum}\left(\mathbf{S}_{gg'}^{(t)}\right)\right)_{g,g'=1}^{G} \in \mathbb{R}^{G\times G}$; and $\odot$ denotes the (entry-wise) Hadamard product. Compared to the log-likelihood function in (4.2.7), the alternative form in (4.2.8) largely reduces the number of parameters from $R(R+1)/2$ of $\boldsymbol{\Omega}$ to $G(G+1)/2$ of $\boldsymbol{\gamma}$ (e.g., from $14,535$ to 15 for the NMR data with $R = 170$ and $G = 5$).

In addition to simplifying the matrix calculations in the log-likelihood function, the following theorems state that we can obtain the estimator of $\boldsymbol{\theta}$ without using the iteration algorithm.

Theorem 4.2.1 (properties of $\widehat{\boldsymbol{\beta}}$)**.** *Suppose Conditions 1, 2, 3, 4, 5, and 6 hold. Then, the FGLS estimator of $\boldsymbol{\beta}$ is equal to the generalized least-square (GLS) estimator of $\boldsymbol{\beta}$, while it is equal to*

the OLS estimator, denoted by $\widehat{\boldsymbol{\beta}}$*, namely,*

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^{(t)} &= \left\{\boldsymbol{x}^\top \left(\boldsymbol{I}_n \otimes \widehat{\boldsymbol{\Omega}}^{(t-1)}\right)\boldsymbol{x}\right\}^{-1}\boldsymbol{x}^\top \left(\boldsymbol{I}_n \otimes \widehat{\boldsymbol{\Omega}}^{(t-1)}\right)\boldsymbol{y} \\ &= \left\{\boldsymbol{x}^\top \left(\boldsymbol{I}_n \otimes \boldsymbol{\Omega}\right)\boldsymbol{x}\right\}^{-1}\boldsymbol{x}^\top \left(\boldsymbol{I}_n \otimes \boldsymbol{\Omega}\right)\boldsymbol{y} \\ &= \left(\boldsymbol{x}^\top\boldsymbol{x}\right)^{-1}\boldsymbol{x}^\top\boldsymbol{y} \\ &= \widehat{\boldsymbol{\beta}},\end{aligned} \tag{4.2.9}$$

for all $t \geq 1$*. In addition,* $\widehat{\boldsymbol{\beta}}$ *is (weakly) consistent, asymptotically normally distributed, asymptotically efficient, and*

$$\widehat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{\beta}, \boldsymbol{\Sigma}\left(\boldsymbol{A}_{\boldsymbol{\Sigma}}, \boldsymbol{B}_{\boldsymbol{\Sigma}}, \boldsymbol{\ell}\right) \otimes \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top\right)^{-1}\right). \tag{4.2.10}$$

We remark here that in general, the explicit variance estimator or the finite-sample properties of an FGLS estimator is intractable (Hayashi, 2011).

Consequently, based on the OLS estimator $\widehat{\boldsymbol{\beta}}$, classical optimization algorithms (e.g., Newton–Raphson) can be straightforwardly implemented to obtain $\widehat{\boldsymbol{\gamma}}$:

$$\widehat{\boldsymbol{\gamma}} \in \underset{\boldsymbol{\gamma}\in\Theta_{\boldsymbol{\gamma}}}{\operatorname{argmax}}\, \ell_n\left(\widehat{\boldsymbol{\beta}}, \boldsymbol{\gamma}; \mathbf{x}, \mathbf{y}\right), \tag{4.2.11}$$

where the log-likelihood function is given by (4.2.8). Accordingly, we obtain $\widehat{\boldsymbol{\rho}} = \left(\widehat{\rho}_{11}, \ldots, \widehat{\rho}_{GG}\right)^\top \in \mathbb{R}^{G(G+1)/2\times 1}$ of $\rho_{gg'}$ in (4.2.1) with $\widehat{\rho}_{gg} = (L_g - 1)\widehat{\gamma}_{gg}$ for $g' = g$ and $\widehat{\rho}_{gg'} = \widehat{\rho}_{g'g} = \sqrt{(L_g - 1)(L_{g'} - 1)}\widehat{\gamma}_{gg'}$ for $g' \neq g$, which completes the MAUD parameter estimation.

Theorem 4.2.2 (properties of $\widehat{\boldsymbol{\gamma}}$)**.** *Suppose Conditions 1, 2, 3, 4, 5, and 6 hold. Then,* $\widehat{\boldsymbol{\gamma}}$ *in* (4.2.11)

is the unique ML estimator, (weakly) consistent, asymptotically normally distributed, and asymptotically efficient.

Then, $\widehat{\boldsymbol{\rho}}$ is (weakly) consistent, asymptotically normally distributed, and asymptotically efficient under the mild regularity conditions.

Theorem 4.2.3 (properties of $\widehat{\boldsymbol{\theta}}$)**.** *Suppose Conditions 1, 2, 3, 4, 5, and 6 hold. Then,* $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\beta}}^\top, \widehat{\boldsymbol{\gamma}}^\top\right)^\top$ *satisfies the score equation with respect to* $\boldsymbol{\theta}$*, i.e.,*

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ell_n\left(\boldsymbol{\theta}; \boldsymbol{x}, \boldsymbol{y}\right) \Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} = 0.$$

In addition, the Fisher information matrix of the log-likelihood function $\ell_n\left(\boldsymbol{\theta}; \boldsymbol{x}, \boldsymbol{y}\right)$ *is*

$$\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}_{\boldsymbol{\beta}} & \boldsymbol{0}_{(Rp)\times[G(G+1)/2]} \\ \boldsymbol{0}_{[G(G+1)/2]\times(Rp)} & \boldsymbol{\Psi}_{\boldsymbol{\gamma}} \end{pmatrix},$$

and it is positive definite, where

$$\boldsymbol{\Psi}_{\boldsymbol{\beta}} = \boldsymbol{x}^\top \left[\boldsymbol{I}_n \otimes \boldsymbol{\Omega}\left(\boldsymbol{A}_{\boldsymbol{\Omega}}, \boldsymbol{B}_{\boldsymbol{\Omega}}, \boldsymbol{\ell}\right)\right] \boldsymbol{x} = \boldsymbol{\Omega}\left(\boldsymbol{A}_{\boldsymbol{\Omega}}, \boldsymbol{B}_{\boldsymbol{\Omega}}, \boldsymbol{\ell}\right) \otimes \left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^\top\right) \in \mathbb{R}^{(Rp)\times(Rp)},$$

$$\boldsymbol{\Psi}_{\boldsymbol{\gamma}} = \left(\psi_{jj'}^{(\boldsymbol{\gamma})}\right), \quad \psi_{jj'}^{(\boldsymbol{\gamma})} = \frac{n}{2} \operatorname{tr}\left\{\left[\frac{\partial \boldsymbol{\Omega}\left(\boldsymbol{A}_{\boldsymbol{\Omega}}, \boldsymbol{B}_{\boldsymbol{\Omega}}, \boldsymbol{\ell}\right)}{\partial \boldsymbol{\gamma}_j}\right] \boldsymbol{\Sigma}\left(\boldsymbol{A}_{\boldsymbol{\Sigma}}, \boldsymbol{B}_{\boldsymbol{\Sigma}}, \boldsymbol{\ell}\right) \left[\frac{\partial \boldsymbol{\Omega}\left(\boldsymbol{A}_{\boldsymbol{\Omega}}, \boldsymbol{B}_{\boldsymbol{\Omega}}, \boldsymbol{\ell}\right)}{\partial \boldsymbol{\gamma}_{j'}}\right] \boldsymbol{\Sigma}\left(\boldsymbol{A}_{\boldsymbol{\Sigma}}, \boldsymbol{B}_{\boldsymbol{\Sigma}}, \boldsymbol{\ell}\right)\right\},$$

$\boldsymbol{\gamma}_j \in \{\gamma_{11}, \ldots, \gamma_{1G}, \ldots, \gamma_{2G}, \ldots, \gamma_{GG}\}$ *denotes the* j*-th component of* $\boldsymbol{\gamma}$ *for* $j = 1, \ldots, G(G+1)/2$*, and* $\psi_{jj'}^{(\boldsymbol{\gamma})}$ *is simplified to a closed-form expression in Chapter C.*

4.2.4 Inference about the MAUD parameters

In this section, we describe statistical inferences about $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$.

Inference about $\boldsymbol{\beta}$. We conduct statistical tests for $\boldsymbol{\beta}$ using the estimator presented in (4.2.9). Without loss of generality, we are simultaneously testing the following *Rp covariate-wise hypotheses*:

$$H_{0,r,q} : \beta_r^{(q)} = 0 \quad \text{versus} \quad H_{1,r,q} : \beta_r^{(q)} \neq 0, \quad q = 1, \ldots, p, r = 1, \ldots, R, \tag{4.2.12}$$

where $\beta_r^{(q)}$ is the q-th component of the r-th regression coefficient vector, i.e., $\boldsymbol{\beta}_r = \left(\beta_r^{(1)}, \ldots, \beta_r^{(p)}\right)^\top$.

An exact covariance matrix of $\widehat{\boldsymbol{\beta}}$ is available from (4.2.10):

$$\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}} = \operatorname{cov}\left(\widehat{\boldsymbol{\beta}}\right) = \boldsymbol{\Sigma}\left(\mathbf{A}_{\boldsymbol{\Sigma}}, \mathbf{B}_{\boldsymbol{\Sigma}}, \boldsymbol{\ell}\right) \otimes \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\top\right)^{-1} \in \mathbb{R}^{(Rp)\times(Rp)}. \tag{4.2.13}$$

In general, given a pre-determined matrix $\mathbf{C}^* \in \mathbb{R}^{s\times(Rp)}$ with a full rank, we can test a secondary parameter (SP) $\boldsymbol{\varrho}^* = \mathbf{C}^*\boldsymbol{\beta} \in \mathbb{R}^{s\times 1}$:

$$H_{0,\text{SP}} : \boldsymbol{\varrho}^* = \boldsymbol{\varrho}_0^* \quad \text{versus} \quad H_{1,\text{SP}} : \boldsymbol{\varrho}^* \neq \boldsymbol{\varrho}_0^*.$$

We construct a Wald-type test statistic using the estimator $\widehat{\boldsymbol{\varrho}}^* = \mathbf{C}^*\widehat{\boldsymbol{\beta}}$ that follows a multivariate normal distribution with mean $\mathbf{C}^*\boldsymbol{\beta} \in \mathbb{R}^{s\times 1}$ and covariance matrix $\mathbf{C}^*\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}}\mathbf{C}^{*,\top} \in \mathbb{R}^{s\times s}$.

Inference about $\boldsymbol{\gamma}$. By applying Theorem 4.2.3 and utilizing the log-likelihood function in (4.2.8), we can compute the observed information matrix to estimate the asymptotic standard errors of $\widehat{\boldsymbol{\gamma}}$. Subsequently, we perform $G(G+1)/2$ Wald-type tests to evaluate the null and alternative hypotheses:

$$H_{0,gg'} : \gamma_{gg'} = 0 \quad \text{versus} \quad H_{1,gg'} : \gamma_{gg'} \neq 0, \quad g \leq g' = 1, \ldots, G. \tag{4.2.14}$$

Multiple testing correction procedures (e.g., false discovery rate, FDR) can be further performed to account for the multiplicity and dependence of the simultaneous tests (Benjamini and Hochberg, 1995; Leek and Storey, 2008).

4.3 Simulation Studies

In this section, we use Monte Carlo simulation studies to (1) evaluate the performance of MAUD for parameter estimation; (2) assess the performance of statistical inference about regression coefficients of MAUD, and compare it with those of competing linear regression models, including the ordinary least-square regression model (OLS; Worsley and Friston, 1995), the autoregressive model for multivariate block-diagonal outcomes (AMBD; Lee et al., 2023), and the mixed model for repeated measures (MMRM; Bove et al., 2022); and (3) perform sensitivity analysis for the MAUD under model misspecification when the underlying covariance matrix does not have the interconnected community structure.

4.3.1 Data generation

We generate the multivariate outcomes $\mathbf{y}_i$ from $N(\mathfrak{B}\mathbf{x}_i, \mathbf{\Sigma})$, where $i = 1, \ldots, n$. We firstly specify the mean vectors by sampling $\mathbf{x}_i$ from a standard normal distribution and specifying the regression coefficients $\mathfrak{B}^\top = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_R)_{p \times R}$ with $p = 2$. We vary the regression coefficients associated with the covariates of interest (i.e., $\beta_r^{(2)}$) across communities while setting the other non-zero regression coefficients different from $\beta_r^{(2)}$. Specifically, we set $\beta_r^{(1)} \neq \beta_r^{(2)}$ and $\beta_r^{(2)}$ depending on κ in $\boldsymbol{\beta}_r = \left(\beta_r^{(1)}, \beta_r^{(2)}\right)^\top$ for the first two features within each community and set $\boldsymbol{\beta}_r$ as zero for the other features, where κ is the effect size (i.e., the mean over standard

deviation). We consider various settings with $n \in \{100, 300, 500\}$, $R \in \{10, 100, 200\}$, and $\kappa \in \{0.4, 0.6, 0.8\}$. Next, we set the covariance matrix $\mathbf{\Sigma} = \mathbf{\Sigma}(\mathbf{A}_{\mathbf{\Sigma}}, \mathbf{B}_{\mathbf{\Sigma}}, \boldsymbol{\ell})$ with

$$\mathbf{A}_{\boldsymbol{\Upsilon}} = \begin{pmatrix} -0.77 & & \\ & -1.15 & \\ & & -0.24 \end{pmatrix}, \quad \mathbf{B}_{\boldsymbol{\Upsilon}} = \begin{pmatrix} 0.77 & 0.11 & -0.57 \\ & 1.15 & 1.12 \\ & & 0.24 \end{pmatrix}.$$

This covariance structure has $G = 3$ interconnected communities. For $R \in \{10, 10, 200\}$, we set the communities sizes $\boldsymbol{\ell}^\top \in \{(3, 3, 4), (30, 30, 40), (60, 60, 80)\}$, respectively. We simulate 100 replicates for this setting.

4.3.2 Evaluation of estimation of the scaled autoregressive dependence parameters by the MAUD

For each replicated dataset, we apply the MAUD to estimate $\boldsymbol{\beta}$ using (4.2.9) and $\boldsymbol{\gamma}$ using (4.2.11). We evaluate whether the MAUD provides an accurate scaled autoregressive dependence vector $\boldsymbol{\gamma}$. To assess the performance of estimator $\widehat{\boldsymbol{\gamma}}$, we use the evaluation metrics including the average relative bias (denoted by "bias"), the Monte Carlo standard deviation (denoted by "MCSD"), the average asymptotic standard error (denoted by "ASE"), and the 95% coverage probability based on the Wald-type confidence interval (denoted by "95% CP") for each element of $\boldsymbol{\gamma}$.

Table 4.1 summarizes the estimate of $\boldsymbol{\gamma}$ (the complete results are available in Chapter C). The results demonstrate that the biases are relatively small, especially when compared with the Monte Carlo standard deviations, and the asymptotic standard errors are close to the Monte Carlo

(n, R)	parameter	bias	MCSD	ASE	95% CP
$(100, 10)$	γ_{11}	0.0077	0.0459	0.0492	0.96
	γ_{12}	-0.0022	0.0376	0.0397	0.93
	γ_{13}	-0.0072	0.0417	0.0375	0.93
	γ_{22}	0.0197	0.0908	0.0947	0.97
	γ_{23}	0.0143	0.0682	0.0622	0.94
	γ_{33}	0.0107	0.0533	0.0492	0.93
$(100, 100)$	γ_{11}	0.0195	0.0240	0.0233	0.87
	γ_{12}	-0.0126	0.0583	0.0679	0.99
	γ_{13}	0.0099	0.0781	0.0776	0.94
	γ_{22}	0.0152	0.0304	0.0280	0.92
	γ_{23}	0.0283	0.0793	0.0725	0.89
	γ_{33}	0.0125	0.0130	0.0141	0.89
$(100, 200)$	γ_{11}	0.0168	0.0148	0.0164	0.85
	γ_{12}	-0.0092	0.0648	0.0698	0.99
	γ_{13}	0.0128	0.0898	0.0795	0.91
	γ_{22}	0.0192	0.0180	0.0198	0.89
	γ_{23}	0.0257	0.0808	0.0723	0.86
	γ_{33}	0.0130	0.0097	0.0099	0.78

Table 4.1: Estimation results of $\boldsymbol{\gamma}$ under $n = 100$, where "bias" denotes the average of estimation bias, "MCSD" denotes the Monte Carlo standard deviation, "ASE" denotes the average asymptotic standard error, "95% CP" denotes the coverage probability based on a 95% Wald-type confidence interval.

standard deviations. As the number of features increases, the biases slightly increase, and therefore, the coverage probabilities based on 95% Wald-type confidence intervals are slightly smaller than the nominal level. As the sample size increases, the asymptotic standard errors decrease and are closer to the Monte Carlo standard deviations, and the coverage probabilities based on 95% Wald-type confidence intervals become closer to the nominal level.

4.3.3 Evaluation of statistical inference about the regression coefficients by the MAUD

We also evaluate the statistical inference about the regression coefficients by the MAUD and compare it with those of the OLS, AMBD, and MMRM. In contrast, we also use the true

covariance matrix $\boldsymbol{\Sigma}\left(\mathbf{A}_{\boldsymbol{\Sigma}}, \mathbf{B}_{\boldsymbol{\Sigma}}, \boldsymbol{\ell}\right)$ as an "estimator", denoted by "TRUE". Since we are interested in selecting a subset of features that are associated with the covariates of interest, we evaluate the statistical inference from the perspective of variable selection. In general, the false positive rate (FPR) and the true positive rate (TPR), defined below:

$$\begin{aligned}
\mathrm{TPR}_{\alpha} &= \frac{\#\left\{(q,r) : \beta_r^{(q)} \neq 0,\ H_{0,r,q} \text{ is rejected at } \alpha, 1 \leq q \leq p, 1 \leq r \leq R\right\}}{\#\left\{(q,r) : \beta_r^{(q)} \neq 0, 1 \leq q \leq p, 1 \leq r \leq R\right\}}, \\
\mathrm{FPR}_{\alpha} &= \frac{\#\left\{(q,r) : \beta_r^{(q)} = 0,\ H_{0,r,q} \text{ is rejected at } \alpha, 1 \leq q \leq p, 1 \leq r \leq R\right\}}{\#\left\{(q,r) : \beta_r^{(q)} = 0, 1 \leq q \leq p, 1 \leq r \leq R\right\}},
\end{aligned}$$

are commonly used criteria to evaluate variable selection with a given cutoff α ranging from 0 to 1. To avoid making an ad-hoc choice of cutoff, we utilize receiver operating characteristic (ROC) curves, as a statistical model with a higher area under the curve indicates a more accurate statistical inference.

We demonstrate the results in the first two rows of Figure 4.2 for $R = 100$ and $R = 200$. In general, the MAUD has the closest behavior to TRUE, and the ROC curves produced by the MAUD uniformly outperform the other methods which are followed by those of the AMBD, MMRM, and OLS, respectively. The difference between their TPRs becomes larger at the same FPR when the effect size increases. When $\kappa = 0.8$, the AMBD yields a slightly higher TPR than that of the MAUD at the same FPR. One possible explanation is the AMBD gains statistical power probably because it uses fewer parameters to model $\boldsymbol{\gamma}$ than the MAUD does. The OLS and MMRM have unreasonable performances due to the lowest TPRs.

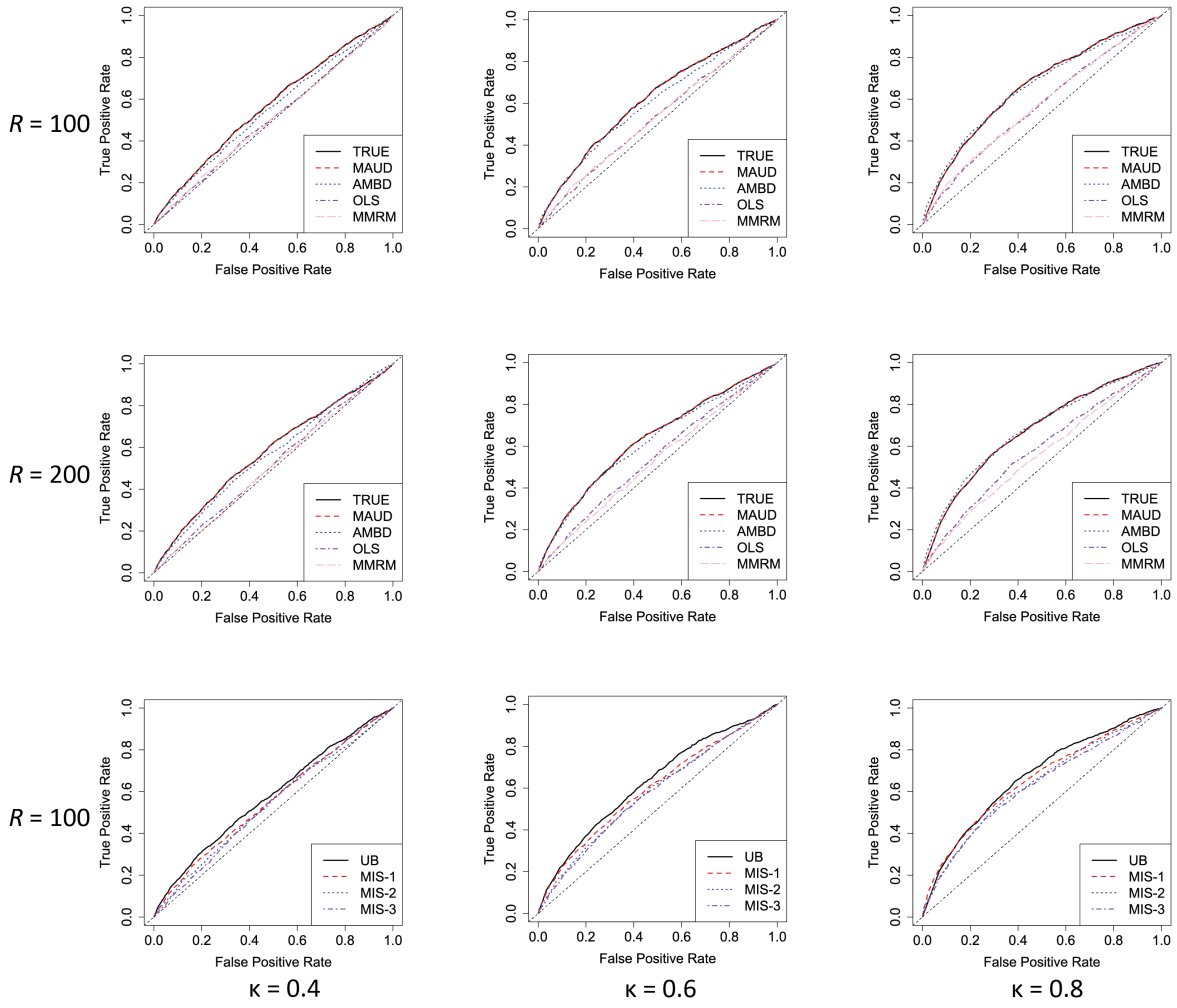


Figure 4.2: ROC curves under various levels of effect size κ for different covariance matrix estimators, different statistical models, and different noise levels of covariance matrices.

4.3.4 Misspecification analysis of the MAUD

We assumed a known interconnected community structure for the MAUD in the last two studies. However, the latent covariance structure may deviate from the interconnected community structure in practice. Here, we assess the robustness of the MAUD under model misspecification. In particular, we introduce a perturbation term $\mathbf{E}_\sigma \sim \text{Wishart}\,(R, \sigma\mathbf{I}_R)$ for $\sigma \in \{1, 2, 3\} \times 10^{-4}$ and $\mathbf{E}_\sigma = \mathbf{0}_{R\times R}$ for $\sigma = 0$. We generate the simulated datasets 100 times with covariance matrix

$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\left(\mathbf{A}_{\boldsymbol{\Sigma}}, \mathbf{B}_{\boldsymbol{\Sigma}}, \boldsymbol{\ell}\right) + \mathbf{E}_{\sigma}$ under $n = 500$ and $R = 100$.

For each simulated dataset, we calculate the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}}$. Then, we similarly test the covariate-wise hypotheses and plot the ROC curves in Figure 4.2 based on all replicates.

As expected, the ROC curves in the third row of Figure 4.2 show that the TPR increases at the same FPR when the effect size increases. As the noise level becomes larger, the TPR decreases. We note that when $\sigma = 10^{-4}$, the elements of $\mathbf{E}_{\sigma}$ nearly have the same order of magnitude as those of the true covariance matrix. It implies that the MAUD can handle slight and moderate levels of noise on the covariance structure.

4.4 Investigation of the Effect of Alcohol Intake on Plasma Metabolomics

To investigate the influence of alcohol consumption on plasma metabolomics, we apply the proposed method to a nuclear magnetic resonance (NMR) plasma metabolomics dataset, with a particular focus on the associations between the NMR metabolic biomarkers and alcohol intake. The dataset is accessible at UK Biobank and described in Ritchie et al. (2023). It comprises $S = 249$ NMR metabolic biomarker measurements as outcomes; alcohol intake frequency (912 participants daily, 1175 three or four times a week, 1017 once or twice a week, 424 one to three times a month, 300 special occasions only, and 156 never) as the exposure; and the intercept, age (when attended assessment center, 63.39 ± 7.61 in year), sex (1840 male and 2144 female), BMI (body mass index, 26.28 ± 4.11), and heavy smoking (883 participants yes, and 3101 no) as the covariates. This dataset has $p = 6$ and $n = 3984$ participants.

Detecting the interconnected community structure of NMR data. We fit the data to a multivariate linear regression model using the least-square method and obtain the estimated regression

coefficient matrix $\widetilde{\mathfrak{B}}^*_{249\times 6}$. We next calculate the residual matrix and apply a network detection algorithm proposed by Chen et al. (2023). After reordering the biomarkers, the correlation matrix of the residuals demonstrates $R = 170$ biomarkers are categorized into 5 communities with sizes of $L_1 = 77$, $L_2 = 47$, $L_3 = 19$, $L_4 = 11$, and $L_5 = 16$, and the remaining $L_0 = 79$ biomarkers are in no communities, regarded as the singletons (see the heat map in Figure 4.1(B)). We also obtain the estimated regression coefficient matrix $\widehat{\mathfrak{B}}^*_{249\times 6}$, which is calculated based on $\widetilde{\mathfrak{B}}^*_{249\times 6}$ by arranging its rows according to the reordered biomarkers, i.e., the first $R = 170$ rows of $\widehat{\mathfrak{B}}^*_{249\times 6}$ are the estimates for the biomarkers in communities and the last $L_0 = 79$ rows are the estimates for the singletons. The names of the biomarkers, their community indexes, and their source communities, defined by Ritchie et al. (2023), and $\widehat{\mathfrak{B}}^*_{249\times 6}$, are available in Chapter C.

Statistical inference about $\mathfrak{B}_{249\times 6}$. Based on the detected interconnected community structure, we aim to estimate and conduct statistical inference about the regression coefficient matrix $\mathfrak{B}^*_{249\times 6} = \begin{pmatrix} \mathfrak{B}_{170\times 6} \\ \mathfrak{B}^\dagger_{79\times 6} \end{pmatrix}$ based on the proposed MAUD and the general linear model, separately. Firstly, we partition $\widehat{\mathfrak{B}}^*_{249\times 6} = \begin{pmatrix} \widehat{\mathfrak{B}}_{170\times 6} \\ \widehat{\mathfrak{B}}^\dagger_{79\times 6} \end{pmatrix}$. Then, we estimate the standard errors for $\widehat{\mathfrak{B}}_{170\times 6}$ using the proposed MAUD and those for $\widehat{\mathfrak{B}}^\dagger_{79\times 6}$ using the general linear model (see the results in Chapter C). Finally, we select alcohol intake-related NMR biomarkers based on the inference results at FDR level 0.05. We also plot the 95% confidence intervals for all regression coefficients in Figure 4.3.

Result. Our findings reveal that at FDR level 0.05, most high-density lipoprotein (HDL) biomarkers and Apolipoprotein A1 biomarkers have significant positive associations with alcohol intake.

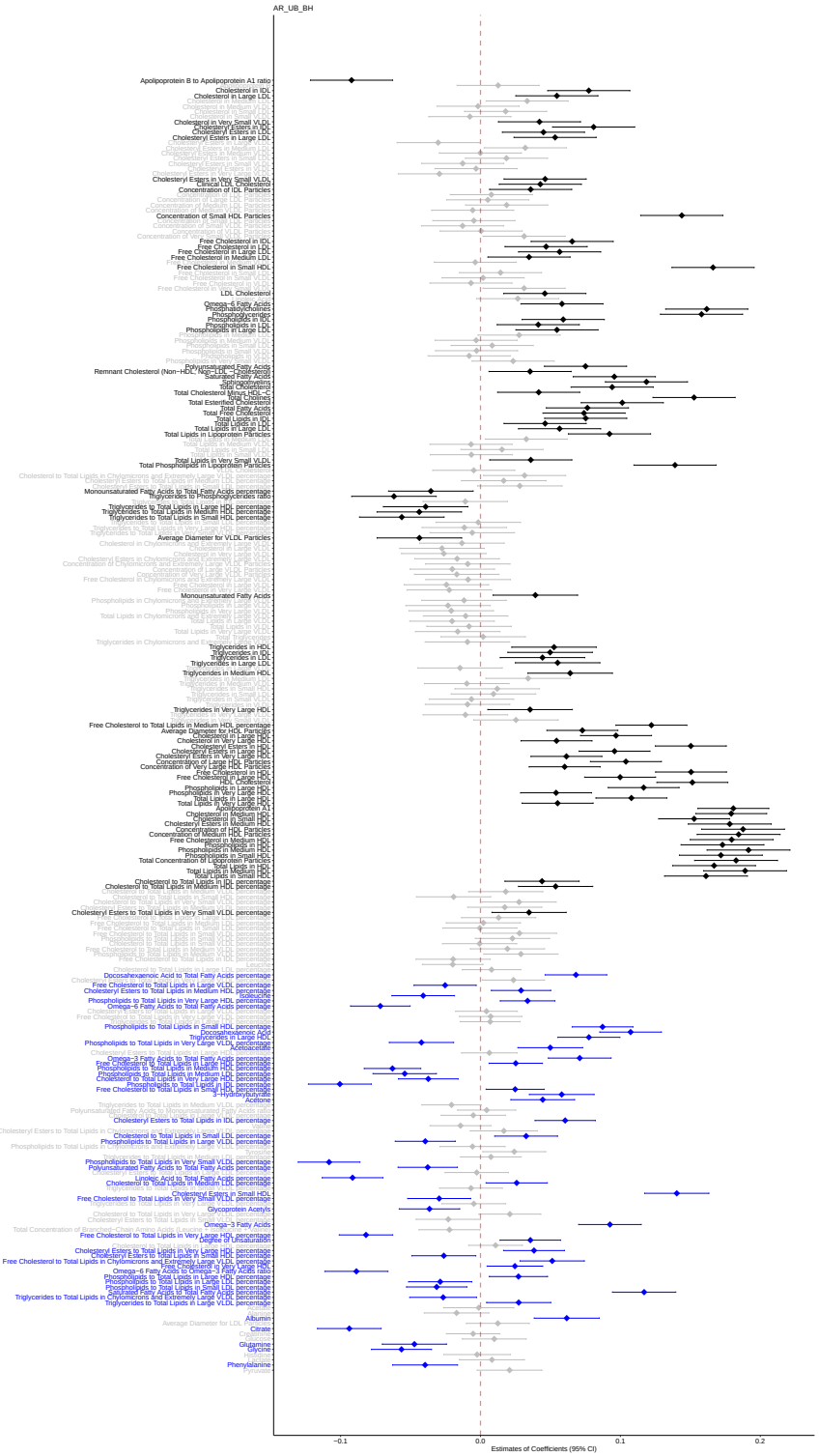


Figure 4.3: Forest plot of 95% confidence intervals for 249 biomarkers regression coefficients, biomarkers in "black" refer to the significant ones in 5 communities, biomarkers in "blue" refer to the significant singletons, biomarkers in "grey" refer to the non-significant ones.

Specifically, among all 249 NMR biomarkers, there are 12, 80, 44, and 58 NMR biomarkers associated with the intermediate-density lipoprotein (IDL), very-low-density lipoprotein (VLDL), low-density lipoprotein (LDL), and HDL, respectively. In Table 4.2, we report the numbers of biomarkers by the communities (i.e., within the 5 communities or no communities), the statistical decisions, and the signs of estimated regression coefficients.

	In 5 Communities			In No Communities			Total
	significant		non-significant	significant		non-significant	
	$+$	$-$		$+$	$-$		
IDL	8	0	2	1	1	0	12
VLDL	4	1	56	2	6	11	80
LDL	13	0	21	2	3	5	44
HDL	36	3	3	10	4	2	58

Table 4.2: Statistical results of the effect of alcohol intake frequency on the biomarkers associated with the lipoprotein at level $\alpha = 0.05$, where "IDL" refers to the intermediate-density lipoprotein, "VLDL" refers to the very-low-density lipoprotein, "LDL" refers to the low-density lipoprotein, "HDL" refers to the high-density lipoprotein, and $+$ and $-$ refer to the signs of estimated regression coefficients.

For LDL, the coefficients of 13 significant NMR biomarkers among the first 170 ones are positive, the coefficients of the rest 5 significant NMR biomarkers (i.e., phospholipids to total lipids in small/medium/large LDL percentage, and cholesterol to total lipids in small/medium LDL percentage) negative and positive, respectively.

For HDL, 36 NMR biomarkers (out of the 39 significant ones) are positively associated with, and the rest 3 (i.e., triglycerides to total lipids in small/medium/large HDL percentage), are negatively associated with the alcohol intake frequency. Among the 14 remaining significant biomarkers, the alcohol intake frequency is negatively associated with phospholipids to total lipids in medium HDL percentage, cholesterol to total lipids in very large HDL percentage, free cholesterol to total lipids in very large HDL percentage, and cholesteryl esters to total lipids in

small HDL percentage only, is positively associated with the other ones. In summary, among 58 HDL biomarkers, there are 36 ones positively associated with alcohol intake frequency. Therefore, the increasing frequency of alcohol intake significantly has a positive effect on the above 36 biomarkers. In addition to the biomarkers associated with lipoprotein, we also observe that there is a significant association between alcohol intake and Apolipoprotein B to Apolipoprotein A1 ratio, while the association between alcohol intake and Apolipoprotein A1 is significantly positive, that between alcohol intake and Apolipoprotein B is not significant at level 0.05. This is agreed with the result that apolipoprotein A1 may play an important role when increasing alcohol consumption raises HDL cholesterol levels (Silva et al., 2000).

In contrast, we also apply the AMBD and MMRM to the NMR dataset. Their inference results and the forest plots are presented in Chapter C.

4.5 Discussion

We developed a new multivariate regression technique MAUD to jointly model correlated multivariate outcomes. Compared to the classic linear regression approaches that model each outcome separately, the MAUD can effectively improve the statistical inference, i.e., the MAUD produces higher statistical powers and fewer false positive findings than the competing ones. This framework is applicable for various omics data analyses (e.g., differential expression analysis) because these datasets often exhibit latent interconnected community structures.

We build the MAUD based on the autoregressive model while accounting for the interconnected community structure. We bridge the gap between the autoregressive dependence parameters and the parameters in a large covariance model. By leveraging the interconnected community

structure, we develop efficient (i.e., closed form) estimators for the dependence parameters in the autoregressive model. By accounting for the dependence of multivariate outcomes, we obtain more accurate inference and thus can select the omics features with improved accuracy. This may also lead to potentially improved reproducibility and replicability for high-throughput data analysis. Based on the simulation studies, the MAUD is also robust to model misspecification. Although we use the interconnected community structure for the MAUD, a more general framework can be developed for the covariance structure under the autoregressive model.

In our data application, we found that alcohol intake can increase the level of "good cholesterol" HDL which may protect the cardiovascular condition. In addition, our findings on HDL and related pathways of Apolipoprotein A1 and Apolipoprotein B are biologically sensible. However, we may cautiously consider the potential cancer risks of over intake of alcohol. These jointly may provide a new perspective on alcohol intake for public health.

In conclusion, our proposed MAUD, estimation, and inference procedures are applicable to a wide range of network studies where an interconnected community structure with uniform blocks is present.

Chapter 5: Semi-Confirmatory Factor Analysis for Multivariate Data with Interconnected Community Structures

5.1 Introduction

Factor analysis is a widely utilized technique in multivariate data analysis, aimed at decomposing observed data into linear combinations of unobserved common factors and describing the relationship between observations in terms of common factors (Anderson, 2003). It is commonly categorized into two classes: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). In applications where there is no prior theory guiding the analysis, EFA is typically the preferred choice to explore the relationship between intercorrelated observations and latent common factors. On the other hand, CFA is employed when there is strong evidence supporting a specific factor structure in the data, i.e., the covariance matrix of the common factors has a particular structure rather than a diagonal one. As extensively discussed in the literature, CFA offers two primary advantages. First, it effectively reduces the dimensionality of the data by utilizing common factors. Second, unlike EFA, a CFA model allows for flexible design of the factor loading matrix, making it possible to leverage knowledge gained from preliminary studies, particularly in cases where the common factors are found to be correlated in a particular form. Detailed justifications and further references on the advantages of CFA can be found in classi-

cal works (Lawley, 1958), overviews (Schreiber et al., 2006; Jackson et al., 2009), and related literature (Basilevsky, 2009; Brown, 2015; Gana and Broc, 2019).

In statistical literature, factor analysis is inherently associated with the estimation of large covariance matrices (Fan et al., 2013; Fan and Han, 2017; Fan et al., 2019). EFA allows us to express a large population covariance matrix $\mathbf{\Sigma}_{p\times p}$ of p-dimensional observations in terms of the factor analysis parameters: $\mathbf{L}_{p\times K}$ (the factor loading matrix), $\mathbf{\Sigma}_{\boldsymbol{f}}$ (the covariance matrix of K common factors), and $\mathbf{\Sigma}_{\boldsymbol{u}}$ (the covariance matrix of error terms). The relationship is given by $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^\top + \mathbf{\Sigma}_{\boldsymbol{u}}$ (refer to Figure 5.1), under the identification conditions $\mathbf{L}^\top\mathbf{L}$ is diagonal and $\mathbf{\Sigma}_{\boldsymbol{f}} = \mathbf{I}_K$ (Bai and Li, 2012; Fan et al., 2013, 2020). However, in practice, the prior knowledge necessary for CFA is often unavailable. As a result, EFA has gained widespread popularity (Friguet et al., 2009; Bai and Li, 2012; Fan et al., 2012).

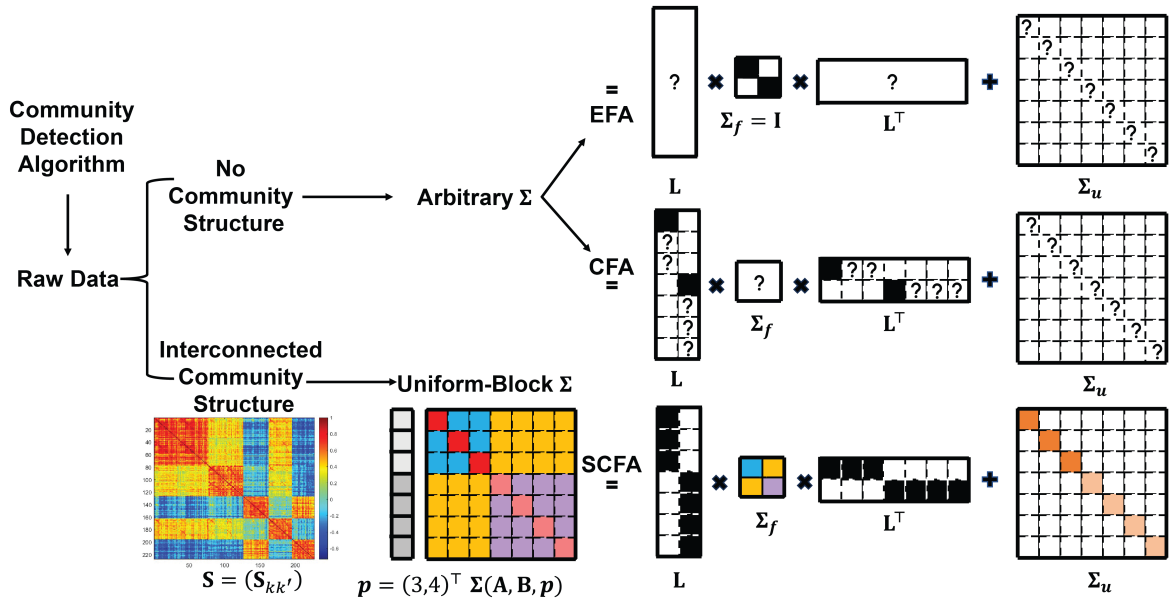


Figure 5.1: Illustration of the workflow for the SCFA, CFA, and EFA models: we first apply a community detection algorithm to the raw data; if the interconnected community structure is detected, then we use the SCFA model; otherwise, we consider the conventional EFA or CFA models.

Interestingly, in a CFA model, the large population covariance matrix $\mathbf{\Sigma}$, determined by $\mathbf{L}$, $\mathbf{\Sigma}_f$, and $\mathbf{\Sigma}_u$, exhibits a particular structure that finds extensive application in network data analysis. Specifically, we consider two cases: the orthogonal or oblique common factors, respectively, under the conditions $\mathbf{L}$ has a block-diagonal form (where non-zero column vectors lie along the diagonal) and $\mathbf{\Sigma}_f$ is arbitrary positive definite. First, we consider a CFA model with the orthogonal common factors. From a network analysis perspective, $\mathbf{\Sigma}$ possesses a *community structure*, that is, the variables between communities are independent, consequently, $\mathbf{\Sigma}$ is a block-diagonal matrix (Newman and Girvan, 2004; Fortunato, 2010). Second, when the common factors in a CFA model are dependent, we refer to the structure of $\mathbf{\Sigma}$ as an *interconnected community structure*. In this scenario, the variables between communities exhibit correlation, leading to a uniform-block structure in $\mathbf{\Sigma}$ (see real examples in Figure 5.2). Both the community structure and interconnected community structure commonly manifest in large network data. Each community represents a scientific mechanism, such as a biological pathway (Wu et al., 2021), or a social tie within social groups (Homans, 2013).

In practice, on the one hand, the interconnected community structure can be extracted from the sample covariance matrix using community detection algorithms (Chen et al., 2018; Wu et al., 2021; Perrot-Dockès et al., 2022), while the corresponding population covariance matrix has a uniform-block structure. On the other hand, a CFA model is unavailable due to a lack of the requisite knowledge, while the covariance matrix of the observations $\mathbf{\Sigma}$ has a "similar" pattern to the uniform-block structure. Motivated by this intriguing covariance structure similarity, current research focuses on exploring whether a CFA model can be informed by the detected interconnected community structure and the corresponding (uniform-block) covariance matrix. Thus, a CFA model, assisted by prior knowledge from the empirical interconnected community detec-

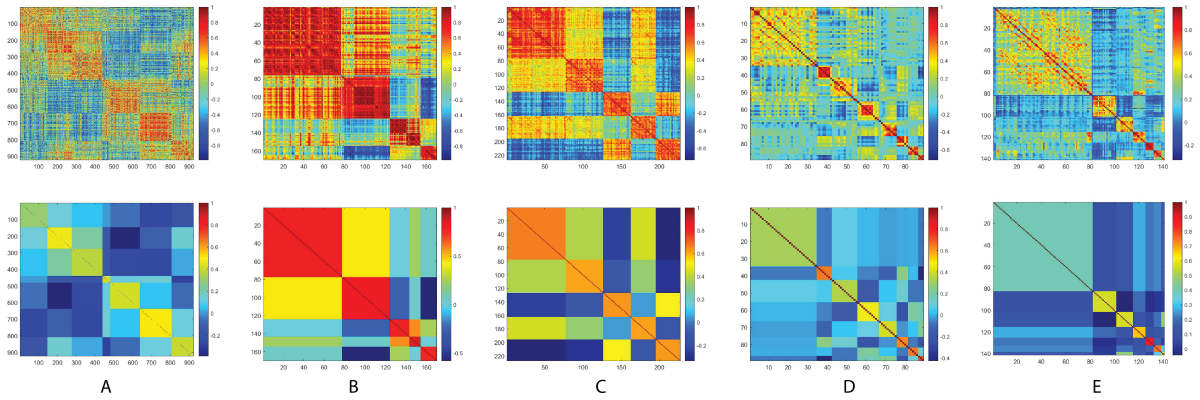


Figure 5.2: Real examples of the interconnected community structures (the first row) and the corresponding assumed population covariance matrices (the second row). A: the sample and assumed population covariance matrices for a seed quality study (Perrot-Dockès et al., 2022); B: the sample and assumed population covariance matrices for a nuclear magnetic resonance study (Ritchie et al., 2023); C: the sample and assumed population covariance matrices for an echo-planar spectroscopic imaging study (Chiappelli et al., 2019); D and E: the sample and assumed population covariance matrices for an environmental study involving exposome and metabolites (ISGlobal, 2021).

tion, may offer both scientifically and statistically explainable dimension reduction for the data. However, a statistical framework for this purpose is currently unavailable. Although numerous software and computational packages have been developed (Fox, 2006; Rosseel, 2012; Oberski, 2014; Huang, 2020), the computational burden often becomes unmanageable for many applications, particularly in high-dimensional data scenarios. To address this gap, we propose a novel semi-confirmatory factor analysis (SCFA) model.

The SCFA model first estimates the factor loading matrix $\mathbf{L}$ by leveraging the particular pattern (i.e., a uniform-block structure) and the community sizes obtained from the detected interconnected community structure. Additionally, it estimates the factor scores, as well as the covariance matrices of common factors and error terms. The presence of interconnected community structures has been observed in diverse domains, including brain imaging, biology, plant science, computer science, gene expression, and others (Girvan and Newman, 2002; Colizza et al., 2006;

Simpson et al., 2013; Levine et al., 2015; Huttlin et al., 2017; Zitnik et al., 2018; Perrot-Dockès et al., 2022). Real examples showcasing these structures can be found in Figure 5.2. Therefore, SCFA demonstrates considerable potential for wide-ranging applications.

The SCFA model offers several significant contributions to the field. First, the proposed model provides a unique and elegant decomposition representation for the covariance matrix of the observations $\mathbf{\Sigma}$ with valuable guidance by integrating the data-driven interconnected community structure. As the prior theory for the observations, we apply clustering or community detection algorithms to access the detected interconnected community structure in preliminary studies. Second, we demonstrate that the factor loading matrix $\mathbf{L}$, covariance matrix of common factors $\mathbf{\Sigma}_{\boldsymbol{f}}$, and covariance matrix of the error terms $\mathbf{\Sigma}_{\boldsymbol{u}}$ can be explicitly determined using some elements of $\mathbf{\Sigma}$ only (illustrated in Figure 5.1). The determination potentially enhances the estimation procedure and leads to more accurate and reliable results. Third, we derive explicit estimators for all unknown matrices in the SCFA model, significantly reducing the computational burden. Furthermore, we establish both finite-sample and large-sample theoretical guarantees for the proposed estimators. These offer assurance regarding the computational time and the accuracy and reliability of the estimated parameters. Fourth, the SCFA model provides a more comprehensive and nuanced understanding of the relationships between the observed variables and common factors.

The remainder of the chapter is organized as follows. Section 5.2 presents the SCFA model with specifications in the factor loading matrix, and details the estimation procedure for all unknown matrices and factor scores. The following two sections, Section 5.3 and Section 5.4, evaluate the proposed estimation procedure using simulated and real data, respectively. All technical proofs are presented in Chapter D.

5.2 Methodology

5.2.1 A confirmatory factor analysis model

Let $\boldsymbol{X} = (X_1, \ldots, X_p)^\top$ denote a p-dimensional vector of observations, $\boldsymbol{f} = (f_1, \ldots, f_K)^\top$ denote a K-dimensional vector of common factors ($1 < K < p$), and $\boldsymbol{u} = (u_1, \ldots, u_p)^\top$ denote a p-dimensional vector error terms. A (strict) *factor model* can be defined as

$$\begin{aligned} &\boldsymbol{X}_{p\times 1} = \boldsymbol{\mu}_{p\times 1} + \mathbf{L}_{p\times K}\boldsymbol{f}_{K\times 1} + \boldsymbol{u}_{p\times 1}, \\ \text{satisfying} \quad &\mathrm{E}(\boldsymbol{f}) = \mathbf{0}_{K\times 1}, \quad \mathrm{E}(\boldsymbol{u}) = \mathbf{0}_{p\times 1}, \quad \mathrm{cov}(\boldsymbol{u}) = \boldsymbol{\Sigma}_{\boldsymbol{u}}, \quad \mathrm{cov}\,(\boldsymbol{f}, \boldsymbol{u}) = 0, \end{aligned} \tag{5.2.1}$$

where $\boldsymbol{\mu}$ is a p-dimensional mean vector, without loss of generality, we assume $\boldsymbol{\mu} = \mathbf{0}_{p\times 1}$; $\mathbf{L} = (\ell_{jk})$ is a p by K factor loading matrix; and $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ is a p by p diagonal matrix. Furthermore, let $\boldsymbol{\Sigma} = \mathrm{cov}(\boldsymbol{X})$ and $\boldsymbol{\Sigma}_{\boldsymbol{f}} = \mathrm{cov}(\boldsymbol{f})$ denote the p by p covariance matrix of $\boldsymbol{X}$ and the K by K covariance matrix of $\boldsymbol{f}$, respectively. Thus, based on specifications (5.2.1), we have

$$\boldsymbol{\Sigma} = \mathbf{L}\boldsymbol{\Sigma}_{\boldsymbol{f}}\mathbf{L}^\top + \boldsymbol{\Sigma}_{\boldsymbol{u}}, \tag{5.2.2}$$

where $\top$ denotes the transpose of a vector or matrix. In particular, the common factors are said to be *orthogonal* if $\boldsymbol{\Sigma}_{\boldsymbol{f}} = \mathbf{I}_K$ or to be *oblique* if $\boldsymbol{\Sigma}_{\boldsymbol{f}}$ is not diagonal.

Given the prior theory, we may consider zero elements in specified positions of $\mathbf{L}$. For example, if we set a condition $\ell_{jk} = 0$, then it implies that the k-th factor does not affect the j-th response, where $j = 1, \ldots, p$ and $k = 1, \ldots, K$. Under these zero conditions, Jöreskog (1969) called the above factor analysis as *confirmatory factor analysis*. In the present paper, we specify

the factor loading matrix $\mathbf{L}$ as below:

$$\mathbf{L}_{p\times K} = \mathrm{Bdiag}\left(\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_K\right)$$
$$= \begin{pmatrix} \ell_{\bar{p}_0+1,1} & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{\bar{p}_1,1} & 0 & \ldots & 0 \\ 0 & \ell_{\bar{p}_1+1,2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ell_{\bar{p}_2,2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \ell_{\bar{p}_{K-1}+1,K} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \ell_{\bar{p}_K,K} \end{pmatrix}, \quad \ell_{\bar{p}_{k-1}+1,k} = \iota_k \neq 0, \quad k = 1, \ldots, K, \quad (5.2.3)$$

where $\mathrm{Bdiag}(\cdot)$ denotes a block-diagonal matrix; $p_1, \ldots, p_K$ are integers such that $p = p_1 + \cdots + p_K$; $\bar{p}_k = \sum_{k'=1}^{k} p_{k'}$ is the cumulative sum (we define $\bar{p}_0 = 0$); $\boldsymbol{\ell}_k = \left(\ell_{\bar{p}_{k-1}+1,k}, \ldots, \ell_{\bar{p}_k,k}\right)^\top$ is the p_k-dimensional *non-zero sub-vector* of columns of $\mathbf{L}$; the first element of $\boldsymbol{\ell}_k$ is a known non-zero number $\ell_{\bar{p}_{k-1}+1,k} = \iota_k \neq 0$ but the other elements of $\boldsymbol{\ell}_k$ are unknown, for $k = 1, \ldots, K$.

Typically, we estimate the $(p - K)$ non-zero elements of $\mathbf{L}$, the $K(K+1)/2$ elements of symmetric $\boldsymbol{\Sigma}_{\boldsymbol{f}}$, and the p elements of diagonal $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ in a CFA model (5.2.2) with (5.2.3) by using a likelihood-based approach. Furthermore, we also estimate the common factor $\boldsymbol{f}$ using a least-square-based approach.

5.2.2 The SCFA model

Specifications of factor loading matrix $\mathbf{L}$ in (5.2.3), i.e., the positive integer values $p_1, \ldots, p_K$ are determined by the detected interconnected community structure in the raw data. Specifically, we apply the community detection algorithm to the raw data, yielding an interconnected community structure (precisely, we refer to this interconnected community structure at the sample level as one with *well-organized blocks*, see the first row in Figure 5.2). These well-organized blocks imply the population covariance matrix has also an interconnected community structure (precisely, we refer to this interconnected community structure at the population level as one with *uniform blocks*, see the second row in Figure 5.2). Thus, our prior knowledge includes that the population covariance matrix has an interconnected community structure (with uniform blocks), the sizes of the communities $p_1, \ldots, p_K$ satisfying $p = p_1 + \cdots + p_K$, and the number of communities K.

Incorporating the interconnected community structure in the covariance matrix of observations, $\mathbf{\Sigma}$ is a uniform-block matrix (see Chapter 2):

$$\mathbf{\Sigma} = \mathbf{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p}) = \mathbf{A} \circ \mathbf{I}(\boldsymbol{p}) + \mathbf{B} \circ \mathbf{J}(\boldsymbol{p}), \tag{5.2.4}$$

where $\boldsymbol{p} = (p_1, \ldots, p_K)^\top$ is the K-dimensional *partition-size vector* satisfying $p = p_1 + \cdots + p_K$ (see Chapter 2); $\mathbf{A} = \operatorname{diag}(a_{11}, \ldots, a_{KK})$ and $\mathbf{B} = (b_{kk'})$ are K by K diagonal matrix and symmetric matrix, respectively (see Chapter 2); $\mathbf{I}(\boldsymbol{p}) = \operatorname{Bdiag}(\mathbf{I}_{p_1}, \ldots, \mathbf{I}_{p_K})$ and $\mathbf{J}(\boldsymbol{p}) = \left(\mathbf{1}_{p_k \times p_{k'}}\right)$ with identity matrix $\mathbf{I}$ and all-one matrix $\mathbf{1}$; $\circ$ is the block Hadamard product, i.e., $\mathbf{A} \circ \mathbf{I}(\boldsymbol{p}) = \operatorname{Bdiag}(a_{11}\mathbf{I}_{p_1}, \ldots, a_{KK}\mathbf{I}_{p_K})$ and $\mathbf{B} \circ \mathbf{J}(\boldsymbol{p}) = \left(b_{kk'}\mathbf{1}_{p_k \times p_{k'}}\right)$ (see Chapter 2).

Substituting covariance matrix (5.2.4) into the parametric covariance model (5.2.2), we obtain the following confirmatory covariance analysis model:

$$\boldsymbol{\Sigma}\left(\mathbf{A}, \mathbf{B}, \boldsymbol{p}\right) = \mathbf{L}\boldsymbol{\Sigma}_{\boldsymbol{f}}\mathbf{L}^{\top} + \boldsymbol{\Sigma}_{\boldsymbol{u}}. \tag{5.2.5}$$

Moreover, we refer to model (5.2.1) with (5.2.5) as a *semi-confirmatory factor analysis* (SCFA) model. The following corollary describes the relationship between $\boldsymbol{\Sigma}\left(\mathbf{A}, \mathbf{B}, \boldsymbol{p}\right)$ between $\mathbf{L}$, $\boldsymbol{\Sigma}_{\boldsymbol{f}}$, and $\boldsymbol{\Sigma}_{\boldsymbol{u}}$, respectively.

Corollary 5.2.1 (Parametrize $\mathbf{L}$, $\boldsymbol{\Sigma}_{\boldsymbol{f}}$, and $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ in terms of $\mathbf{A}$, $\mathbf{B}$, and $\boldsymbol{p}$)**.** *Suppose p-dimensional vector $\boldsymbol{X}$ with zero mean and known covariance matrix $\boldsymbol{\Sigma}\left(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{p}\right)$, satisfies a SCFA model, where $\boldsymbol{p} = \left(p_1, \ldots, p_K\right)^{\top}$ is a known with $p_k > 2$ for every k, $\boldsymbol{A} = \operatorname{diag}\left(a_{11}, \ldots, a_{KK}\right) \succ 0$ is known and positive definite, and $\boldsymbol{B} = \left(b_{kk'}\right) \succ 0$ is also known and positive definite with $b_{k'k} = b_{kk'}$ for every $k \neq k'$. Then, we have the analytic solutions:*

$$\begin{aligned}
&\boldsymbol{\ell}_k = \iota_k \boldsymbol{1}_{p_k \times 1}, \text{ i.e., } \ell_{\bar{p}_{k-1}+1,k} = \cdots = \ell_{\bar{p}_k,k} = \iota_k, \quad k = 1, \ldots, K; \\
&\boldsymbol{\Sigma}_{\boldsymbol{f}} = \operatorname{diag}\left(\iota_1^{-1}, \ldots, \iota_K^{-1}\right) \boldsymbol{B} \operatorname{diag}\left(\iota_1^{-1}, \ldots, \iota_K^{-1}\right) = \left(\frac{b_{kk'}}{\iota_k \iota_{k'}}\right); \\
&\boldsymbol{\Sigma}_{\boldsymbol{u}} = \boldsymbol{A} \circ \boldsymbol{I}(\boldsymbol{p}) = \operatorname{Bdiag}\left(a_{11}\boldsymbol{I}_{p_1}, \ldots, a_{KK}\boldsymbol{I}_{p_K}\right).
\end{aligned}$$

In particular, if we assume $\iota_1 = \cdots = \iota_K = 1$, then $\boldsymbol{\ell}_k = \boldsymbol{1}_{p_k \times 1}$ for all k, $\boldsymbol{\Sigma}_{\boldsymbol{f}} = \boldsymbol{B}$, and $\boldsymbol{\Sigma}_{\boldsymbol{u}} = \boldsymbol{A} \circ \boldsymbol{I}(\boldsymbol{p})$.

Let $\boldsymbol{\iota} = \left(\iota_1, \ldots, \iota_K\right)^{\top} \in \mathbb{R}^{K \times 1}$ denote the vector of the first elements of $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_K$. Under the condition $\boldsymbol{\iota} = \mathbf{1}_{K \times 1}$, we refer to model (5.2.1) with (5.2.5) as a *standard* SCFA model.

Corollary 5.2.1 demonstrates the following facts for a standard SCFA model. We would like

to decompose a p-dimensional vector $\boldsymbol{X}$ with zero mean and covariance matrix $\boldsymbol{\Sigma}$ into a confirmatory factor model (5.2.1). Given the knowledge from preliminary research, covariance matrix $\boldsymbol{\Sigma}$ has an interconnected community structure (with uniform blocks), i.e., $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$, and factor loading matrix $\mathbf{L}$ has preassigned 0's as (5.2.3). Furthermore, if $\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$ is known with $\mathbf{A} \succ 0$ and $\mathbf{B} \succ 0$ (therefore $\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p}) \succ 0$, see Chapter 2), and $\boldsymbol{\iota} = \mathbf{1}_{K\times 1}$, by Corollary 5.2.1, then: (1) the covariance matrix of K common factors is exactly the K by K positive definite matrix $\mathbf{B}$; (2) the non-zero sub-vectors of $\mathbf{L}$ in (5.2.3) are equal to all-one vectors; and (3) the covariance matrix of error terms or idiosyncratic components is exactly the diagonal matrix $\mathrm{Bdiag}(a_{11}\mathbf{I}_{p_1}, \ldots, a_{KK}\mathbf{I}_{p_K})$ (or expressed by $\mathbf{A} \circ \mathbf{I}(\boldsymbol{p})$).

In other words, given $\mathbf{L}$ in (5.2.3), symmetric $\boldsymbol{\Sigma}_{\boldsymbol{f}}$, and particular diagonal $\boldsymbol{\Sigma}_{\boldsymbol{u}}$, the underlying $\boldsymbol{\Sigma} = \mathbf{L}\boldsymbol{\Sigma}_{\boldsymbol{f}}\mathbf{L}^\top + \boldsymbol{\Sigma}_{\boldsymbol{u}}$ has the interconnected community structure (with uniform blocks) as $\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$. On the other hand, Corollary 5.2.1 substantiates that $\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$ in a standard SCFA can be uniquely decomposed into $\mathbf{L} = \mathrm{Bdiag}(\mathbf{1}_{p_1\times 1}, \ldots, \mathbf{1}_{p_K\times 1})$, $\boldsymbol{\Sigma}_{\boldsymbol{f}} = \mathbf{B}$, and $\boldsymbol{\Sigma}_{\boldsymbol{u}} = \mathbf{A} \circ \mathbf{I}(\boldsymbol{p})$ satisfying $\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p}) = \mathbf{L}\boldsymbol{\Sigma}_{\boldsymbol{f}}\mathbf{L}^\top + \boldsymbol{\Sigma}_{\boldsymbol{u}}$.

Therefore, for a fixed k in a standard SCFA model: (1) the k-th set of responses $\{X_{\bar{p}_{k-1}+1}, \ldots, X_{\bar{p}_k}\}$ has *equal* influence on the k-th common factor f_k only (while has no influences on the other common factors); (2) the k-th set of error terms $\{u_{\bar{p}_{k-1}+1}, \ldots, u_{\bar{p}_k}\}$ has a common variance a_{kk} ($a_{kk} > 0$ due to $\mathbf{A} \succ 0$) and are uncorrelated to the other sets of error terms; and (3) the k-th common factor f_k has a variance b_{kk} and has a covariance $b_{kk'}$ with the other k'-th common factor $f_{k'}$ for $k' \neq k$.

5.2.3 The estimation procedure

Let $\mathbf{X}_{n\times p}$ denote the data matrix. Applying a community detection algorithm to the raw data, we assume an interconnected community structure is detected, which implies the population covariance matrix has the uniform-block structure in (5.2.4). Based on the detected structure, we obtain the prior knowledge of the specific pattern in the covariance matrix (i.e., the interconnected community structure), the number of communities, i.e., $\widehat{K}$, and the community sizes, i.e., $\widehat{p} = (\widehat{p}_1, \ldots, \widehat{p}_K)^\top$ satisfying $p = \widehat{p}_1 + \cdots + \widehat{p}_K$ and $\widehat{p}_k > 2$ for $k = 1, \ldots, \widehat{K}$. Theoretical properties related to the detected communities can be found in Chen et al. (2018) and Wu et al. (2021). For ease of presentation, we suppress the notation $\widehat{\cdot}$ and use $K = \widehat{K}$ and $\boldsymbol{p} = \widehat{\boldsymbol{p}}$.

Suppose its rows $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are independently and identically distributed and satisfy the proposed standard SCFA (5.2.1) with (5.2.5):

$$\boldsymbol{X}_i = \mathbf{L}\boldsymbol{f}_i + \boldsymbol{u}_i, \quad i = 1, \ldots, n;$$

$$\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p}) = \mathbf{L}\boldsymbol{\Sigma}_{\boldsymbol{f}}\mathbf{L}^\top + \boldsymbol{\Sigma}_{\boldsymbol{u}}, \quad \text{with} \quad \boldsymbol{\iota} = \mathbf{1}_{K\times 1}.$$

Let $\mathbf{S}_{p\times p} = \mathbf{X}^\top\mathbf{X}/n$ denote the sample covariance matrix. In this section, we would like to estimate the factor loading matrix $\mathbf{L}$, the covariance matrix of common factors $\boldsymbol{\Sigma}_{\boldsymbol{f}}$, the covariance matrix of error terms $\boldsymbol{\Sigma}_{\boldsymbol{u}}$, and the factor scores $\boldsymbol{f}_i$.

Theorem 5.2.1 (The proposed matrix estimators)**.** *Let $(\boldsymbol{S}_{kk'})$ is partitioned matrix of $\boldsymbol{S}$ by $\boldsymbol{p}$ (see Chapter 2). The maximum likelihood estimators of $\boldsymbol{L}$, $\boldsymbol{\Sigma}_{\boldsymbol{f}}$, and $\boldsymbol{\Sigma}_{\boldsymbol{u}}$ are*

$$\widehat{\boldsymbol{L}} = \mathrm{Bdiag}\left(\boldsymbol{I}_{p_1\times 1}, \ldots, \boldsymbol{I}_{p_K\times 1}\right), \quad \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}} = \widehat{\boldsymbol{B}}, \quad \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{u}} = \widehat{\boldsymbol{A}} \circ \boldsymbol{I}(\boldsymbol{p}) = \mathrm{Bdiag}\left(\widehat{a}_{11}\boldsymbol{I}_{p_1}, \ldots, \widehat{a}_{KK}\boldsymbol{I}_{p_K}\right), \tag{5.2.6}$$

where $\widehat{\boldsymbol{A}} = \operatorname{diag}(\widehat{a}_{11}, \ldots, \widehat{a}_{KK})$, $\widehat{\boldsymbol{B}} = \left(\widehat{b}_{kk'}\right)$ *with* $\widehat{b}_{kk'} = \widehat{b}_{k'k}$, *and*

$$\widehat{a}_{kk} = \frac{p_k \operatorname{tr}(\boldsymbol{S}_{kk}) - \operatorname{sum}(\boldsymbol{S}_{kk})}{p_k(p_k - 1)}, \quad \widehat{b}_{kk'} = \begin{cases} \dfrac{\operatorname{sum}(\boldsymbol{S}_{kk'})}{p_k \times p_{k'}}, & k \neq k' \\ \dfrac{\operatorname{sum}(\boldsymbol{S}_{kk}) - \operatorname{tr}(\boldsymbol{S}_{kk})}{p_k(p_k - 1)}, & k = k' \end{cases},$$

for all k *and* k', $\operatorname{tr}(\cdot)$ *and* $\operatorname{sum}(\cdot)$ *denote the trace and the sum of all elements of a matrix, respectively.*

We remark here that the proof of Theorem 5.2.1 is straightforward by Corollary 5.2.1. In addition, the derivation of maximum likelihood estimators $\widehat{a}_{kk}$ and $\widehat{b}_{kk'}$, and their exact closed-form variance estimators, i.e., $\operatorname{var}(\widehat{a}_{kk})$ and $\operatorname{var}\left(\widehat{b}_{kk'}\right)$, are presented in Chapter 3.

Theorem 5.2.2 (Properties of the proposed matrix estimators)**.** *The proposed matrix estimators* $\widehat{\boldsymbol{L}}$, $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{f}}$, *and* $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{u}}$ *are uniformly minimum-variance unbiased estimators (UMVUE).*

The proof of Theorem 5.2.2 depends on the fact that all $\widehat{a}_{kk}$ and $\widehat{b}_{kk'}$ are the UMVUE of a_{kk} and $b_{kk'}$. See the details in Chapter 3.

Theorem 5.2.3 (The proposed factor score estimator and its properties)**.** *The proposed factor score estimator of* $\boldsymbol{f}_i$ *is*

$$\widehat{\boldsymbol{f}}_i = \left(\widehat{\boldsymbol{L}}^\top \widehat{\boldsymbol{L}}\right)^{-1} \widehat{\boldsymbol{L}}^\top \boldsymbol{X}_i = \left(\widehat{\boldsymbol{L}}^\top \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1} \widehat{\boldsymbol{L}}\right)^{-1} \widehat{\boldsymbol{L}}^\top \boldsymbol{\Sigma}_{\boldsymbol{u}}^{-1} \boldsymbol{X}_i = \left(\widehat{\boldsymbol{L}}^\top \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{u}}^{-1} \widehat{\boldsymbol{L}}\right)^{-1} \widehat{\boldsymbol{L}}^\top \widehat{\boldsymbol{\Sigma}}_{\boldsymbol{u}}^{-1} \boldsymbol{X}_i, \tag{5.2.7}$$

for $i = 1, \ldots, n$, *where* $\boldsymbol{\Sigma}_{\boldsymbol{u}} = \boldsymbol{A} \circ \boldsymbol{I}(\boldsymbol{p})$ *and* $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{u}} = \widehat{\boldsymbol{A}} \circ \boldsymbol{I}(\boldsymbol{p})$ *are diagonal matrices. So, the proposed OLS estimator is identical to the generalized least-square (GLS) estimator and the feasible generalized least-square (FGLS) estimator. In addition, under mild regularity conditions,*

$\widehat{\boldsymbol{f}}_i$ is (weakly) consistent, asymptotically normally distributed, asymptotically efficient as $p \to \infty$.

The proof of Theorem 5.2.3 follows the arguments in Chapter 4.

5.3 Simulation Studies

We first assess the finite-sample performance of the proposed estimators in (5.2.6) for a standard SCFA model and compare it with a standard CFA model numerical solved by function "cfa()" in the R package **lavaan** (Rosseel et al., 2023). Then, we evaluate the performance of estimation of the factor scores in (5.2.7). In contrast, we also compare it with a standard CFA model numerically solved by the "cfa()" function, and with a standard EFA model numerically solved by the "fa()" function in the package **psych** (Revelle, 2023). Lastly, we perform a misspecification analysis for the SCFA under the scenario where the underlying interconnected community structure is violated.

5.3.1 Data generation procedure

We randomly generate the vectors of observations $\boldsymbol{X}_i = \mathbf{L}_{p\times K}\boldsymbol{f}_i + \boldsymbol{u}_i$, where the factor loading matrix $\mathbf{L} = \mathrm{Bdiag}\left(\mathbf{1}_{p_1\times 1}, \ldots, \mathbf{1}_{p_K\times 1}\right)$, the common factor $\boldsymbol{f}_i \sim N\left(\mathbf{0}_{K\times 1}, \mathbf{B}_{K\times K}\right)$, and the error term $\boldsymbol{u}_i \sim N\left(\mathbf{0}_{p\times 1}, \mathbf{A}_{K\times K} \circ \mathbf{I}(\boldsymbol{p})\right)$ for $i = 1, \ldots, n$. Specifically, we set $\boldsymbol{p}^\top \in \{(15, 15, 20), (30, 30, 40)\}$ with $K = 3$ and $p \in \{50, 100\}$, $n \in \{200, 500, 800\}$, and

$$\mathbf{A} = \begin{pmatrix} 0.1 & & \\ & 0.2 & \\ & & 0.5 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 2.02 & 0.79 & 1.15 \\ & 3.13 & 1.63 \\ & & 3.69 \end{pmatrix}.$$

We repeat the above data generation procedure 100 times.

5.3.2 Study 1: performance of parameter estimation

We fit the standard SCFA model (i.e., $\boldsymbol{\iota} = (1, 1, 1)^\top$) to each simulated dataset using the estimators in (5.2.6).

In contrast, we also fit a standard CFA model to each simulated data using the function "cfa()" in R package **lavaan**. Specifically, we use the function "cfa()" to estimate all $\boldsymbol{\ell}_k$ in the factor loading matrix $\mathbf{L}$, all diagonal elements of $\mathbf{A} \circ \mathbf{I}(\boldsymbol{p})$, all elements of $\mathbf{B}$.

To evaluate the estimation performance for $\mathbf{L}$, $\boldsymbol{\Sigma}_{\boldsymbol{f}}$, and $\boldsymbol{\Sigma}_{\boldsymbol{u}}$, it is equivalent to access the estimation performance for all a_{kk} and $b_{kk'}$. So, we consider the following metrics: the average bias, Monte Carlo standard deviation, average standard error, and 95% Wald-type empirical coverage probability, using the proposed estimates, for each a_{kk} and $b_{kk'}$ in Table 5.1. We also compute the average bias and asymptotic standard error using the results produced by **lavaan**, for all elements of $\boldsymbol{\ell}_k$, $\mathbf{A} \circ \mathbf{I}(\boldsymbol{p})$ (in Chapter D), and each $b_{kk'}$ in Table 5.2. The execution time for both methods is calculated.

The estimation results, as presented in Tables 5.1 and 5.2, demonstrate that our estimation procedure produces the expected estimates. Specifically, for each parameter, the bias is relatively small when compared to the Monte Carlo standard deviation, while the average standard error is close to the Monte Carlo standard deviation. Furthermore, both the bias and average standard error decrease as the sample size increases, and the 95% coverage probability approaches the nominal level as the sample size grows. The proposed estimators are comparable with the estimates produced by **lavaan** and are computationally less expensive.

(n,p)	bias	MCSD	ASE	95% CP	(n,p)	bias	MCSD	ASE	95% CP
$(200,50)$					$(200,100)$				
a_{11}	0.0002	0.0027	0.0027	0.95	a_{11}	0.0002	0.0018	0.0019	0.95
a_{22}	0.0009	0.0051	0.0054	0.95	a_{22}	0.0003	0.0035	0.0037	0.94
a_{33}	0.0000	0.0107	0.0115	0.98	a_{33}	-0.0001	0.0082	0.0080	0.94
b_{11}	0.0045	0.2133	0.2037	0.95	b_{11}	0.0265	0.2193	0.2056	0.97
b_{12}	-0.0224	0.1746	0.1855	0.97	b_{12}	-0.0071	0.1822	0.1868	0.97
b_{13}	-0.0048	0.2155	0.2108	0.97	b_{13}	0.0126	0.2117	0.2112	0.95
b_{22}	-0.0123	0.3570	0.3140	0.91	b_{22}	-0.0025	0.3440	0.3143	0.93
b_{23}	-0.0195	0.2983	0.2671	0.92	b_{23}	0.0078	0.2598	0.2670	0.96
b_{33}	-0.0128	0.3785	0.3715	0.95	b_{33}	-0.0319	0.3512	0.3683	0.94
$(500,50)$					$(500,100)$				
a_{11}	0.0002	0.0015	0.0017	0.96	a_{11}	0.0000	0.0011	0.0012	0.96
a_{22}	-0.0001	0.0035	0.0034	0.94	a_{22}	-0.0004	0.0024	0.0023	0.92
a_{33}	0.0002	0.0067	0.0073	0.97	a_{33}	0.0004	0.0054	0.0051	0.95
b_{11}	-0.0006	0.1280	0.1283	0.96	b_{11}	-0.0104	0.1175	0.1275	0.97
b_{12}	-0.0280	0.1123	0.1171	0.97	b_{12}	-0.0135	0.1071	0.1171	0.99
b_{13}	-0.0122	0.1475	0.1323	0.93	b_{13}	-0.0060	0.1350	0.1327	0.92
b_{22}	-0.0135	0.1878	0.1982	0.97	b_{22}	0.0030	0.1936	0.1988	0.95
b_{23}	-0.0517	0.1775	0.1674	0.94	b_{23}	-0.0023	0.1501	0.1693	0.98
b_{33}	-0.0490	0.2359	0.2323	0.95	b_{33}	0.0034	0.2153	0.2348	0.96
$(800,50)$					$(800,100)$				
a_{11}	-0.0003	0.0012	0.0013	0.97	a_{11}	0.0000	0.0009	0.0009	0.94
a_{22}	-0.0002	0.0029	0.0027	0.94	a_{22}	-0.0001	0.0017	0.0019	0.96
a_{33}	0.0008	0.0053	0.0057	0.96	a_{33}	-0.0004	0.0043	0.0040	0.92
b_{11}	-0.0102	0.1086	0.1009	0.90	b_{11}	0.0122	0.1024	0.1019	0.95
b_{12}	-0.0148	0.0858	0.0926	0.95	b_{12}	-0.0042	0.0931	0.0934	0.93
b_{13}	-0.0150	0.1052	0.1050	0.94	b_{13}	0.0200	0.1010	0.1059	0.97
b_{22}	-0.0020	0.1423	0.1572	0.96	b_{22}	0.0294	0.1561	0.1585	0.96
b_{23}	0.0061	0.1308	0.1343	0.97	b_{23}	0.0396	0.1201	0.1352	0.97
b_{33}	0.0109	0.1732	0.1866	0.97	b_{33}	0.0204	0.1489	0.1864	0.98

Table 5.1: Estimation results of $\mathbf{A} = \mathrm{diag}\,(a_{11}, a_{22}, a_{33})$ and $\mathbf{B} = (b_{kk'})$ with $b_{k'k} = b_{kk'}$ for $k \neq k'$ in Study 1 by using the proposed method under various n and p, where "bias" denotes the average of estimation bias, "MCSD" denotes the Monte Carlo standard deviation, "ASE" denotes the average asymptotic standard error, "95% CP" denotes the coverage probability based on a 95% Wald-type confidence interval. The computational times of 100 replicates are around 0.05 seconds and 0.07 seconds for $p = 50$ and $p = 100$, respectively.

5.3.3 Study 2: performance of factor scores estimation

We also examine the proposed estimation procedure for the factor scores in (5.2.7) in a standard SCFA model. As the competing methods, we fit the standard CFA and EFA models using

(n,p)	bias	MCSD	ASE	(n,p)	bias	MCSD	ASE
$(200,50)$				$(200,100)$			
b_{11}	-0.0105	0.2255	0.2107	b_{11}	0.0055	0.2331	0.2124
b_{12}	-0.0274	0.1734	0.1846	b_{12}	-0.0132	0.1819	0.1856
b_{13}	-0.0138	0.2161	0.2114	b_{13}	0.0071	0.2114	0.2126
b_{22}	-0.0266	0.3711	0.3300	b_{22}	-0.0206	0.3728	0.3302
b_{23}	-0.0300	0.2976	0.2694	b_{23}	0.0032	0.2581	0.2705
b_{33}	-0.0399	0.3922	0.4122	b_{33}	-0.0256	0.3617	0.4126
$(500,50)$				$(500,100)$			
b_{11}	-0.0078	0.1254	0.1335	b_{11}	-0.0134	0.1221	0.1331
b_{12}	-0.0303	0.1128	0.1170	b_{12}	-0.0124	0.1089	0.1176
b_{13}	-0.0137	0.1486	0.1337	b_{13}	-0.0078	0.1382	0.1341
b_{22}	-0.0236	0.1951	0.2088	b_{22}	0.0166	0.2167	0.2113
b_{23}	-0.0532	0.1824	0.1699	b_{23}	-0.0007	0.1540	0.1724
b_{33}	-0.0445	0.2697	0.2605	b_{33}	-0.0026	0.2515	0.2633
$(800,50)$				$(800,100)$			
b_{11}	-0.0151	0.1103	0.1051	b_{11}	0.0103	0.1097	0.1064
b_{12}	-0.0156	0.0842	0.0928	b_{12}	-0.0058	0.0938	0.0936
b_{13}	-0.0155	0.1032	0.1063	b_{13}	0.0190	0.1038	0.1072
b_{22}	0.0001	0.1496	0.1663	b_{22}	0.0200	0.1772	0.1673
b_{23}	0.0082	0.1316	0.1367	b_{23}	0.0364	0.1249	0.1373
b_{33}	0.0193	0.1987	0.2092	b_{33}	0.0180	0.1718	0.2092

Table 5.2: Estimation results of $\mathbf{A} = \text{diag}\,(a_{11}, a_{22}, a_{33})$ and $\mathbf{B} = (b_{kk'})$ with $b_{k'k} = b_{kk'}$ for $k \neq k'$ in Study 1 by using the **lavaan** package under various n and p, where "bias" denotes the average of estimation bias, "MCSD" denotes the Monte Carlo standard deviation, "ASE" denotes the average asymptotic standard error. The computational times of 100 replicates are around 97 seconds and 1776 seconds for $p = 50$ and $p = 100$, respectively.

the function "cfa()" in R package **lavaan** and the function "fa()" in R package **psych**, respectively. We compute the average and standard deviation of the Euclidean loss $\sum_{i=1}^{n} \left\| \widehat{\boldsymbol{f}}_i - \boldsymbol{f}_i \right\|$ based on 100 replicates for each method in Tabel 5.3.

The results demonstrate that the proposed method outperforms the standard CFA and EFA numerical approaches due to the smallest loss. Compared with the standard EFA model, both the proposed SCFA and standard CFA model take advantage of the dependent structure of the common factors. We also note here the conventional numerical approaches may not handle the case of a large dimension p.

(n, p)	SCFA		CFA		EFA	
	mean	standard devation	mean	standard devation	mean	standard devation
$(200, 50)$	38.52	1.20	56.44	14.82	502.35	24.94
$(200, 100)$	27.31	0.82	48.80	13.60	504.62	19.48
$(500, 50)$	96.43	2.15	114.17	14.09	1267.38	21.55
$(500, 100)$	68.21	1.36	93.59	19.34	1266.60	23.55
$(800, 50)$	154.59	2.41	173.20	14.24	2026.37	31.64
$(800, 100)$	109.31	1.83	134.56	21.73	2024.54	30.97

Table 5.3: Euclidean losses of $\sum_{i=1}^{n} \left\| \widehat{\boldsymbol{f}}_i - \boldsymbol{f}_i \right\|$ in Study 2 by using a standard SCFA model, CFA model, and EFA model, respectively, under various n and p, based on 100 replicates.

5.3.4 Study 3: misspecification analysis

We assess the robustness of the proposed SCFA model because the interconnected community structure may not hold in real applications. Therefore, we generate $\boldsymbol{X}_i \sim N\left(\mathbf{0}_{p\times 1}, \boldsymbol{\Sigma}_\sigma\right)$ with $\boldsymbol{\Sigma}_\sigma = \boldsymbol{\Sigma}\left(\mathbf{A}, \mathbf{B}, \boldsymbol{p}\right) + \mathbf{E}_\sigma$, where $n = 500$ is fixed, $\sigma \in \{0.1, 0.3, 0.5\}$ is the noise level and $\mathbf{E}_\sigma \sim$ Wishart $\left(p, \sigma \mathbf{I}_p\right)$ We also repeat this data generation procedure 100 times.

For each simulated dataset, we compute the average bias, Monte Carlo standard deviation, average standard error, and 95% coverage probability for each parameter using the proposed method, similarly to Study 1, in Table 5.4. Table 5.4 shows that when the assumption of a uniform-block structure is violated, the proposed SCFA is still robust for the estimates of $b_{kk'}$.

5.4 Real Application

We fit the standard SCFA model to an echo-planar spectroscopic imaging (EPSI) dataset. A description of the complete dataset and basic statistical analysis can be found in Chiappelli et al. (2019). There are 78 participants (average age is 42.1 with standard deviation 18.8) with 227 measurements of combinations of neurometabolites and brain regions, including 39 male

noise level	(n, p)	bias	MCSD	ASE	95% CP	(n, p)	bias	MCSD	ASE	95% CP
	$(500, 50)$					$(500, 100)$				
	a_{11}	4.9805	0.2932	0.0860	0	a_{11}	10.0059	0.2839	0.1188	0
	a_{22}	5.0391	0.3023	0.0886	0	a_{22}	9.9546	0.2984	0.1194	0
	a_{33}	5.0162	0.2659	0.0801	0	a_{33}	9.9940	0.2492	0.1064	0
$\sigma = 0.01$	b_{11}	-0.0261	0.1755	0.1478	0.91	b_{11}	0.0031	0.1589	0.1495	0.93
	b_{12}	-0.0052	0.1373	0.1318	0.96	b_{12}	-0.0060	0.1433	0.1322	0.94
	b_{13}	0.0241	0.1618	0.1463	0.92	b_{13}	-0.0148	0.1377	0.1454	0.95
	b_{22}	0.0097	0.2276	0.2210	0.97	b_{22}	0.0027	0.2398	0.2199	0.96
	b_{23}	-0.0064	0.1872	0.1821	0.94	b_{23}	-0.0114	0.1842	0.1805	0.93
	b_{33}	0.0164	0.2629	0.2523	0.93	b_{33}	-0.0299	0.2445	0.2485	0.96
	$(500, 50)$					$(500, 100)$				
	a_{11}	14.9683	0.8653	0.2550	0	a_{11}	30.0094	0.9440	0.3540	0
	a_{22}	14.7400	0.7478	0.2528	0	a_{22}	30.0733	0.8624	0.3559	0
	a_{33}	14.9031	0.7004	0.2237	0	a_{33}	29.9873	0.7588	0.3091	0
$\sigma = 0.03$	b_{11}	-0.0093	0.2663	0.1917	0.83	b_{11}	-0.0043	0.2296	0.1916	0.89
	b_{12}	0.0273	0.2271	0.1623	0.82	b_{12}	-0.0288	0.1998	0.1615	0.90
	b_{13}	0.0042	0.2061	0.1728	0.94	b_{13}	0.0102	0.1957	0.1725	0.92
	b_{22}	0.0363	0.3749	0.2641	0.83	b_{22}	0.0043	0.3045	0.2626	0.88
	b_{23}	0.0267	0.2335	0.2073	0.93	b_{23}	0.0169	0.2201	0.2063	0.91
	b_{33}	0.0322	0.3338	0.2848	0.92	b_{33}	0.0121	0.2709	0.2829	0.96
	$(500, 50)$					$(500, 100)$				
	a_{11}	25.1933	1.4343	0.4280	0	a_{11}	50.0152	1.6098	0.5892	0
	a_{22}	24.9853	1.2179	0.4261	0	a_{22}	50.0171	1.5973	0.5904	0
	a_{33}	25.0542	1.2210	0.3712	0	a_{33}	49.9621	1.3609	0.5116	0
$\sigma = 0.05$	b_{11}	-0.0384	0.4052	0.2340	0.75	b_{11}	-0.0077	0.3684	0.2340	0.79
	b_{12}	0.0306	0.2730	0.1920	0.84	b_{12}	-0.0229	0.2394	0.1908	0.88
	b_{13}	0.0186	0.2879	0.1982	0.82	b_{13}	0.0013	0.2263	0.1985	0.94
	b_{22}	0.0499	0.4656	0.3090	0.78	b_{22}	-0.0036	0.4390	0.3046	0.84
	b_{23}	0.0507	0.3120	0.2325	0.86	b_{23}	-0.0096	0.2986	0.2307	0.90
	b_{33}	-0.0018	0.3781	0.3151	0.85	b_{33}	0.0301	0.3836	0.3158	0.88

Table 5.4: Misspecification analysis results of $\mathbf{A} = \text{diag}\,(a_{11}, \ldots, a_{33})$ and $\mathbf{B} = (b_{kk'})$ with $b_{k'k} = b_{kk'}$ for $k \neq k'$ in Study 3 by using the proposed method under various σ and p, where "bias" denotes the average of estimation bias, "MCSD" denotes the Monte Carlo standard deviation, "ASE" denotes the average asymptotic standard error, "95% CP" denotes the coverage probability based on a 95% Wald-type confidence interval.

participants and 39 female participants. We adopt Chen et al. (2018)'s community detection algorithm to the dataset. These 227 measurements of combinations are categorized into $K = 5$ interconnected communities with sizes of 77, 49, 36, 33, and 32, illustrated as Figure 5.2(C). Next, we fit the dataset to a standard SCFA model with $\boldsymbol{\iota} = \mathbf{1}_{5\times 1}$. The estimation results are summarized in Table 5.5.

The results demonstrate that 227 combinations of neurometabolites and brain regions can be decomposed into 5 unobserved common factors: (1) the first common factor influences choline (74 out of 77 are measurements of choline across brain regions); (2) the second common factor influences myo-inositol (49 out of 49 are measurements of myo-inositol across brain regions); (3) the third common factor influences "mixed" neurometabolites (26 out of 36 and 10 out of 36 are measurements of N-acetylaspartate and glutamate–glutamine, respectively, across brain

regions); (4) the fourth common factor influences creatine-containing compounds (33 out of 33 are measurements of creatine-containing compounds across brain regions); and the fifth common factor influences N-acetylaspartate (32 out of 32 measurements of N-acetylaspartate across brain regions). In addition to the classification of combination measurements, these 5 common factors can be furthermore classed into 2 groups: the first group contains the first, second, and fourth common factors, and the second group contains the third and fifth common factors. Among each group, the common factors are significant positive associated with each other. Between the two groups, the common factors are significantly negative associated.

	estimate (SE)	95% CI		estimate (SE)	95% CI
a_{11}	0.3386 (0.0063)	(0.3264, 0.3509)	b_{22}	0.6092 (0.0995)	(0.4142, 0.8041)
a_{22}	0.3908 (0.0091)	(0.3730, 0.4086)	b_{23}	-0.1159 (0.0730)	(-0.2591, 0.0272) †
a_{33}	0.3666 (0.0100)	(0.3470, 0.3862)	b_{24}	0.3290 (0.0804)	(0.1715, 0.4865)
a_{44}	0.3812 (0.0109)	(0.3599, 0.4025)	b_{25}	-0.2863 (0.0786)	(-0.4404, -0.1322)
a_{55}	0.3736 (0.0108)	(0.3524, 0.3948)	b_{33}	0.6334 (0.1037)	(0.4301, 0.8367)
b_{11}	0.6614 (0.1073)	(0.4511, 0.8717)	b_{34}	0.0426 (0.0727)	(-0.1000, 0.1851) †
b_{12}	0.3823 (0.0851)	(0.2156, 0.5490)	b_{35}	0.5172 (0.0938)	(0.3333, 0.7011)
b_{13}	-0.2494 (0.0798)	(-0.4059, -0.0930)	b_{44}	0.6188 (0.1016)	(0.4197, 0.8179)
b_{14}	0.4389 (0.0892)	(0.2641, 0.6136)	b_{45}	-0.1256 (0.0737)	(-0.2701, 0.0188) †
b_{15}	-0.3354 (0.0835)	(-0.4992, -0.1717)	b_{55}	0.6264 (0.1028)	(0.4248, 0.8279)

Table 5.5: Estimation of the correlation matrix of 227 combinations of neurometabolites and brain regions, where "SE" denotes the estimated standard error, "95% CI" denotes the 95% Wald-type confidence interval, and † denotes the 95% CI containing 0.

5.5 Discussion

We propose a semi-confirmatory factor analysis model by incorporating a widely observed interconnected community structure in real applications. It is more computationally efficient than the conventional numerical methods by leveraging closed-form solutions and exact variance estimators. Our proposed model is available for diverse data applications.

The proposed model provides an elegant decomposition representation for the covariance matrix of the observations, which bridges the gap in the conventional CFA models. Under the guidance of the detected interconnected community structure, we derive the uniformly minimum-variance unbiased matrix estimators for the standard SCFA model. In addition, we develop the factor score estimators by a least-square method, which is identical to the GLS and FGLS estimators with expected optimal properties. We examine the performance of estimation and the robustness of the proposed estimators and compare the proposed model with the traditional competing ones. Using a real EPSI dataset, we validate the SCFA model and find the observations can be clustered into five unobserved common factors, which can be furthermore categorized into two significant negative associated classes.

Lastly, the proposed SCFA model can be applied to a rich variety of studies with latent interconnected community structures.

Chapter 6: Conclusions

Recent advancements in community detection and structure characterization have provided valuable insights into the complex interactive relations among features in various large-scale network data, including biology, biomedicine, computer science, finance, plant science, and many others. Incorporating the detected community structures into statistical models can lead to a better understanding of the underlying scientific mechanisms. However, subsequent statistical analyses are challenging due to the high dimensionality of features, the dependence structure of the detected communities, and the heavy-computational burden. Our focus is on an interconnected community structure with uniform blocks that allows non-null connections among features at the community level and is frequently observed in high-dimensional network data.

In Chapter 2, we proposed a block Hadamard product model to represent a covariance matrix with a uniform-block structure. This representation is critical for unraveling the uniform-block matrix and simplifying operations with matrices with unknown parameters, e.g., the product, the inverse, and the eigendecomposition of a covariance matrix.

In Chapter 3, we have developed a computationally efficient method for estimating large covariance and precision matrices with a uniform-block structure. Building on the algebraic properties of uniform-block matrices, we formulate closed-form solutions to the unknown covariance (and precision) parameters in the blocks. Then, we established exact variance estimators and the

asymptotic properties of the covariance parameters, including strong consistency, convergence rate, asymptotic efficiency, and normality. We also have extended this approach to a large number of diagonal blocks and obtained a consistent covariance-matrix estimator under some regularity conditions. Finally, by capitalizing on the uniform-block structure, our estimating approach outperforms the conventional large covariance- and precision-matrix estimation methods.

In Chapter 4, we proposed a novel multivariate regression model that accounts for the dependency parameters among features within and between groups, leveraging the preliminary community analysis. We also derived efficient estimators of regression coefficients in close-form and dependence parameters simultaneously and present a statistical inference procedure. Extensive simulation studies and an analysis of nuclear magnetic resonance (NMR) data validated our proposed method with accurate inferences.

In Chapter 5, we proposed an adaptive confirmatory factor analysis model for interconnected data with a uniform-block structure. Under mild conditions, we provided an elegant statistical interpretation and derived the best-unbiased estimators for the confirmatory factor analysis model with uniform blocks in closed form. We compared the proposed estimation procedure to conventional numerical methods and validated it through various Monte Carlo simulations and real applications.

Appendix A: Supplementary Materials for Chapter 2

A.1 Technical Proofs

A.1.1 Proofs in Section 2.2

Proof of Lemma 2.2.1. The proof of the block Hadamard product representation is straightforward. To show the uniqueness, consider two equal UB matrices $\mathbf{N}_1[\mathbf{A}_1, \mathbf{B}_1, \boldsymbol{p}] = (\mathbf{N}_{1,kk'})$ and $\mathbf{N}_2[\mathbf{A}_2, \mathbf{B}_2, \boldsymbol{p}] = (\mathbf{N}_{2,kk'})$ with a common pre-determined partition-size vector $\boldsymbol{p} = (p_1, \ldots, p_K)^\top$ satisfying that $p_k > 1$ for every k and $p = p_1 + \cdots + p_K$, where $\mathbf{A}_i = \text{diag}(a_{i,11}, \ldots, a_{i,KK})$ is a diagonal matrix and $\mathbf{B}_i = (b_{i,kk'})$ is a symmetric matrix with $b_{i,k'k} = b_{i,kk'}$ for $i = 1, 2$. By the equality, $\mathbf{N}_{1,kk'} = \mathbf{N}_{2,kk'}$ for every k and k'. If $k \neq k'$, then $b_{1,kk'}\mathbf{1}_{p_k \times p_{k'}} = b_{2,kk'}\mathbf{1}_{p_k \times p_{k'}}$ and therefore $b_{1,kk'} = b_{2,kk'}$. If $k = k'$, then $a_{1,kk}\mathbf{I}_{p_k} + b_{1,kk}\mathbf{J}_{p_k} = a_{2,kk}\mathbf{I}_{p_k} + b_{2,kk}\mathbf{J}_{p_k}$, equivalently, $[(a_{1,kk} - a_{2,kk}) + (b_{1,kk} - b_{2,kk})]\mathbf{I}_{p_k} + (b_{1,kk} - b_{2,kk})\mathbf{J}_{p_k} = \mathbf{0}_{p_k \times p_k}$. Due to $p_k > 1$, off-diagonally, $b_{1,kk} = b_{2,kk}$; diagonally, $a_{1,kk} = a_{2,kk}$. Eventually, $\mathbf{A}_1 = \mathbf{A}_2$ and $\mathbf{B}_1 = \mathbf{B}_2$. ■

Proof of Corollary 2.2.1. (1) holds and the proof is straightforward. (2), (2-1), and (2-2) hold because of the following equalities

$$(\mathbf{A}_1 \circ \mathbf{I}[\boldsymbol{p}]) \times (\mathbf{A}_2 \circ \mathbf{I}[\boldsymbol{p}]) = (\mathbf{A}_1 \times \mathbf{A}_2) \circ \mathbf{I}[\boldsymbol{p}], \quad (\mathbf{B}_1 \circ \mathbf{J}[\boldsymbol{p}]) \times (\mathbf{B}_2 \circ \mathbf{J}[\boldsymbol{p}]) = (\mathbf{B}_1 \times \mathbf{P} \times \mathbf{B}_2) \circ \mathbf{J}[\boldsymbol{p}],$$

$$(\mathbf{B}_1 \circ \mathbf{J}[\boldsymbol{p}]) \times (\mathbf{A}_2 \circ \mathbf{I}[\boldsymbol{p}]) = (\mathbf{B}_1 \times \mathbf{A}_2) \circ \mathbf{J}[\boldsymbol{p}], \quad (\mathbf{A}_1 \circ \mathbf{I}[\boldsymbol{p}]) \times (\mathbf{B}_2 \circ \mathbf{J}[\boldsymbol{p}]) = (\mathbf{A}_1 \times \mathbf{B}_2) \circ \mathbf{J}[\boldsymbol{p}],$$

which are easy to verify since $\mathbf{A}_i$ and $\mathbf{P}$ are diagonal matrices for $i = 1, 2$ and $\mathbf{B}_i$ is a symmetric matrix for $i = 1, 2$.

(3) and (4) are proved by using induction with respect to the number of diagonal blocks K. Specifically, we assume that both (3) and (4) hold for the case of K and check whether they hold for the case of $K + 1$. Let $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_K)^\top \in \mathbb{R}^K$, $a, b \in \mathbb{R}$, $q \in \mathbb{N}$ and $q > 1$. Denote $\mathbf{A}^* = \mathrm{Bdiag}(\mathbf{A}, a) \in \mathbb{R}^{(K+1)\times(K+1)}$, $\mathbf{P}^* = \mathrm{Bdiag}(\mathbf{P}, q) \in \mathbb{R}^{(K+1)\times(K+1)}$, and $\mathbf{B}^* = \left(\mathbf{B}, \boldsymbol{\eta}; \boldsymbol{\eta}^\top, b\right) \in \mathbb{R}^{(K+1)\times(K+1)}$. We would like to obtain the eigenvalues of the $(p+q)$ by $(p+q)$ UB matrix $\mathbf{N}^*[\mathbf{A}^*, \mathbf{B}^*, \boldsymbol{p}^*]$ with $(K+1)$ diagonal blocks, where $\boldsymbol{p}^* = \left(\boldsymbol{p}^\top, q\right)^\top$. By definition, the eigenvalues of $\mathbf{N}^*[\mathbf{A}^*, \mathbf{B}^*, \boldsymbol{p}^*]$ are the solutions to the characteristic equation $\det\left(\mathbf{N}^*[\mathbf{A}^*, \mathbf{B}^*, \boldsymbol{p}^*] - \lambda \mathbf{I}_{p+q}\right) = 0$. Equivalently,

$$0 = \det\left(\mathbf{N}^*[\mathbf{A}^*, \mathbf{B}^*, \boldsymbol{p}^*] - \lambda \mathbf{I}_{p+q}\right) = \det \begin{pmatrix} (\mathbf{A} - \lambda \mathbf{I}_K) \circ \mathbf{I}[\boldsymbol{p}] + \mathbf{B} \circ \mathbf{J}[\boldsymbol{p}] & (\eta_k \mathbf{1}_{p_k \times q}) \\ (\eta_k \mathbf{1}_{q \times p_k}) & (a - \lambda)\mathbf{I}_q + b\mathbf{J}_q \end{pmatrix},$$

where $(\eta_k \mathbf{1}_{p_k \times q}) \in \mathbb{R}^{p \times q}$ and $(b_k \mathbf{1}_{q \times p_k}) \in \mathbb{R}^{q \times p}$.

Without loss of generality, assume that $((a - \lambda)\mathbf{I}_q + b\mathbf{J}_q)$ is invertible, or we can consider $\epsilon > 0$ satisfying that $((a - \lambda)\mathbf{I}_q + b\mathbf{J}_q) + \epsilon \mathbf{I}_q$ is positive definite and let $\epsilon \to 0^+$. Then, the characteristic equation can be written as

$$0 = \det\left[(a - \lambda)\mathbf{I}_q + b\mathbf{J}_q\right] \tag{A.1.1}$$

$$\times \det\left\{\left[(\mathbf{A} - \lambda \mathbf{I}_K) \circ \mathbf{I}[\boldsymbol{p}] + \mathbf{B} \circ \mathbf{J}[\boldsymbol{p}]\right] - (\eta_k \mathbf{1}_{p_k \times q})\left[(a - \lambda)\mathbf{I}_q + b\mathbf{J}_q\right]^{-1}(\eta_k \mathbf{1}_{q \times p_k})\right\}. \tag{A.1.2}$$

In other words, the eigenvalues of $\mathbf{N}^*[\mathbf{A}^*, \mathbf{B}^*, \boldsymbol{p}^*]$ consist of some eigenvalues of $(a\mathbf{I}_q + b\mathbf{J}_q)$ and some roots of the rational equation (A.1.2) $= 0$ (not a polynomial equation, i.e., not a character-

istic equation).

First, it is easy to observe that the eigenvalues of $(a\mathbf{I}_q + b\mathbf{J}_q)$ are exactly a with multiplicity $(q-1)$ and $(a+bq)$. We will discard $(a+bq)$ because it is not the eigenvalue of $\mathbf{N}^*[\mathbf{A}^*, \mathbf{B}^*, \boldsymbol{p}^*]$. Second, using the fact $[(a-\lambda)\mathbf{I}_q + b\mathbf{J}_q]^{-1} = (a-\lambda)^{-1}\mathbf{I}_q - b(a-\lambda)^{-1}(a-\lambda+bq)^{-1}\mathbf{J}_q$, we simplify (A.1.2) to (A.1.3) as below,

$$\det\left\{(\mathbf{A}-\lambda\mathbf{I}_K)\circ\mathbf{I}[\boldsymbol{p}] + \left(\mathbf{B} - \frac{q}{a-\lambda+bq}\boldsymbol{\eta}\boldsymbol{\eta}^\top\right)\circ\mathbf{J}[\boldsymbol{p}]\right\}, \tag{A.1.3}$$

which is the determinant of a UB matrix with K diagonal blocks. Thus, use the induction assumption and $\mathbf{A} = \operatorname{diag}(a_{11}, \ldots, a_{KK})$, this determinant equals to the product of $\prod_{k=1}^{K}(a_{kk}-\lambda)^{p_k-1}$ and (A.1.4), which is

$$\det\left\{(\mathbf{A}-\lambda\mathbf{I}_K) + \left(\mathbf{B} - \frac{q}{a-\lambda+bq}\boldsymbol{\eta}\boldsymbol{\eta}^\top\right)\mathbf{P}\right\} = \det\left\{(\boldsymbol{\Delta}-\lambda\mathbf{I}_K) - \frac{q}{a+bq-\lambda}\boldsymbol{\eta}\boldsymbol{\eta}^\top\mathbf{P}\right\}. \tag{A.1.4}$$

Thus, (A.1.2) $= 0$ yields the solutions consist of a_{kk} with multiplicity (p_k-1) for $k = 1, \ldots, K$ and the roots of (A.1.4) $= 0$, which are the eigenvalues of $\boldsymbol{\Delta}^* = \mathbf{A}^* + \mathbf{B}^* \times \mathbf{P}^*$. It is because,

$$\boldsymbol{\Delta}^* = \begin{pmatrix} \mathbf{A} & \mathbf{0}_{K\times 1} \\ \mathbf{0}_{1\times K} & a \end{pmatrix} + \begin{pmatrix} \mathbf{B} & \boldsymbol{\eta} \\ \boldsymbol{\eta}^\top & b \end{pmatrix} \times \begin{pmatrix} \mathbf{P} & \mathbf{0}_{K\times 1} \\ \mathbf{0}_{1\times K} & q \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Delta} & q\boldsymbol{\eta} \\ \boldsymbol{\eta}^\top\mathbf{P} & a+bq \end{pmatrix}.$$

Assuming that $(a + bq - \lambda)$ is invertible, we obtain

$$\begin{aligned} 0 &= \det\left(\boldsymbol{\Delta}^* - \lambda \mathbf{I}_{K+1}\right) \\ &= \det\left(a + bq - \lambda\right) \det\left\{\left(\boldsymbol{\Delta} - \lambda \mathbf{I}_K\right) - \left(q\boldsymbol{\eta}\right)\left(a + bq - \lambda\right)^{-1}\left(\boldsymbol{\eta}^\top \mathbf{P}\right)\right\} \\ &= \det\left(a + bq - \lambda\right) \det\left\{\left(\boldsymbol{\Delta} - \lambda \mathbf{I}_K\right) - \frac{q}{a + bq - \lambda}\boldsymbol{\eta}\boldsymbol{\eta}^\top \mathbf{P}\right\}. \end{aligned}$$

Since $(a + bq)$ cannot be the eigenvalue of $\boldsymbol{\Delta}^*$, all eigenvalues of $\boldsymbol{\Delta}^*$ are the roots of the rational equation (A.1.4) $= 0$. In summary, the eigenvalues of $\mathbf{N}^*\left[\mathbf{A}^*, \mathbf{B}^*, \boldsymbol{p}^*\right]$ consist of a with multiplicity $(q - 1)$, a_{kk} with multiplicity $(p_k - 1)$ for $k = 1, \ldots, K$, and all eigenvalues of $\boldsymbol{\Delta}^* = \mathbf{A}^* + \mathbf{B}^* \times \mathbf{P}^*$.

(5) Let $\mathbf{A}^* = \mathbf{A}^{-1}$ and $\mathbf{B}^* = -\boldsymbol{\Delta}^{-1} \times \mathbf{B} \times \mathbf{A}^{-1}$, where the inverses exist by the assumption. Thus, $\mathbf{A}^*$ is a diagonal matrix and $\mathbf{B}^*$ is a symmetric matrix. It is because both $\mathbf{A}$ and $\mathbf{P}$ are diagonal matrices, so they are commute, and the following equalities are equivalent

$$\mathbf{B} \times \mathbf{P} \times \mathbf{A}^{-1} \times \mathbf{B} = \mathbf{B} \times \mathbf{A}^{-1} \times \mathbf{P} \times \mathbf{B},$$

$$\left(\mathbf{A} + \mathbf{B} \times \mathbf{P}\right) \times \mathbf{A}^{-1} \times \mathbf{B} = \mathbf{B} \times \mathbf{A}^{-1} \times \left(\mathbf{A} + \mathbf{P} \times \mathbf{B}\right),$$

$$-\mathbf{A}^{-1} \times \mathbf{B} \times \left(\mathbf{A} + \mathbf{P} \times \mathbf{B}\right)^{-1} = -\left(\mathbf{A} + \mathbf{B} \times \mathbf{P}\right)^{-1} \times \mathbf{B} \times \mathbf{A}^{-1},$$

$$\mathbf{B}^{*,\top} = \mathbf{B}^*.$$

Then, follow the analogous lines of arguments for the square formula,

$$\mathbf{A} \times \mathbf{A}^* = \mathbf{I}_K, \quad \mathbf{A} \times \mathbf{B}^* + \mathbf{B} \times \mathbf{A}^* + \mathbf{B} \times \mathbf{P} \times \mathbf{B}^* = \mathbf{0}_{K \times K},$$

thus, $\left(\mathbf{A} \circ \mathbf{I}\left[\boldsymbol{p}\right] + \mathbf{B} \circ \mathbf{J}\left[\boldsymbol{p}\right]\right) \times \left(\mathbf{A}^* \circ \mathbf{I}\left[\boldsymbol{p}\right] + \mathbf{B}^* \circ \mathbf{J}\left[\boldsymbol{p}\right]\right) = \mathbf{I}_K \circ \mathbf{I}\left[\boldsymbol{p}\right] + \mathbf{0}_{K \times K} \circ \mathbf{J}\left[\boldsymbol{p}\right] = \mathbf{I}_p$.

Finally, the proof of (6) is straightforward, noting that $\tilde{\mathbf{H}}_k\tilde{\mathbf{H}}_k^\top = \mathbf{I}_{p_k-1}$, $\tilde{\mathbf{H}}_k\mathbf{1}_{1\times p_k}^\top = \mathbf{0}_{(p_k-1)\times 1}$ for every k, and suppose λ is the common eigenvalue of $\boldsymbol{\Delta}$ and $\mathbf{N}[\mathbf{A},\mathbf{B},\boldsymbol{p}]$, if $\boldsymbol{\xi} = (\xi_1,\ldots,\xi_K)^\top$ is the corresponding eigenvector of $\boldsymbol{\Delta}$, then the corresponding eigenvector of $\mathbf{N}[\mathbf{A},\mathbf{B},\boldsymbol{p}]$ is $(\xi_1\mathbf{1}_{1\times p_1},\ldots,\xi_K\mathbf{1}_{1\times p_K})^\top$. ■

A.1.2 Proofs in Section 2.3 and Section 2.4

Before proving the main theorems, we present two lemmas below.

Lemma A.1.1 (Craig's Theorem)**.** *Suppose $\boldsymbol{X} \in \mathbb{R}^p$ is a normal vector with mean $\boldsymbol{\mu} \in \mathbb{R}^p$ and positive definite covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$. Let $\boldsymbol{Y}_1, \boldsymbol{Y}_2 \in \mathbb{R}^{p\times p}$ be two real symmetric matrices. Then, $\boldsymbol{X}^\top\boldsymbol{Y}_1\boldsymbol{X}$ and $\boldsymbol{X}^\top\boldsymbol{Y}_2\boldsymbol{X}$ are independently distributed if and only if $\boldsymbol{Y}_1\boldsymbol{\Sigma}\boldsymbol{Y}_2 = \boldsymbol{0}_{p\times p}$.*

Proof of Lemma A.1.1. This result was originally proposed by Craig (1943). Please refer the corrected proof and discussion in Ravishanker and Dey (2002, page 175, proof of Result 5.4.4), Mathai and Provost (1992, page 209, proof of Theorem 5.2.1), Driscoll and Gundberg (1986) and Ogawa and Olkin (2008). ■

Lemma A.1.2. *Suppose $\boldsymbol{X} \in \mathbb{R}^p$ is a normal vector with mean $\boldsymbol{\mu} \in \mathbb{R}^p$ and positive definite covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$. Let $\boldsymbol{Y} \in \mathbb{R}^{p\times p}$ be a real symmetric matrix. Then, $\boldsymbol{X}^\top\boldsymbol{Y}\boldsymbol{X}$ follows a non-central chi-squared distribution with degree of freedom m and noncentrality parameter $\delta = \frac{1}{2}\boldsymbol{\mu}^\top\boldsymbol{Y}\boldsymbol{\mu}$ if and only if $\boldsymbol{Y}\boldsymbol{\Sigma}$ is idempotent with rank m.*

Proof of Lemma A.1.2. Please refer the proof and discussion in Muirhead (2005, page 31, proof of Theorem 1.4.5), Mathai and Provost (1992, page 199, proof of Theorem 5.1.3) and Zhang (2018). ■

Proof of Theorem 2.3.1. The proof can be divided into three steps. First, we need to decompose U into a sum of several components. Second, we examine the independence of these components. Third, we determine the distribution of each component.

Step 1: decomposition of U. The following equality demonstrates an elegant way to decompose U:

$$\left(\widehat{\mathbf{A}} \times \mathbf{P}\right)^{-1} - \widehat{\mathbf{\Delta}}^{-1} \times \widehat{\mathbf{B}} \times \widehat{\mathbf{A}}^{-1} = \left(\mathbf{P} \times \widehat{\mathbf{\Delta}}\right)^{-1}. \tag{A.1.5}$$

Equality (A.1.5) holds if and only if $\mathbf{P}^{-1} \times \widehat{\mathbf{A}}^{-1} - \widehat{\mathbf{\Delta}}^{-1} \times \widehat{\mathbf{B}} \times \widehat{\mathbf{A}}^{-1} = \widehat{\mathbf{\Delta}}^{-1} \times \mathbf{P}^{-1}$. First, right-multiple $\widehat{\mathbf{A}}$ on both sides of (A.1.5) and switch $\mathbf{P}^{-1}$ and $\widehat{\mathbf{A}}$ on the right hand side since both $\widehat{\mathbf{A}}$ and $\mathbf{P}$ are diagonal matrices, yielding that (A.1.5) holds if and only if $\mathbf{P}^{-1} - \widehat{\mathbf{\Delta}}^{-1} \times \widehat{\mathbf{B}} = \widehat{\mathbf{\Delta}}^{-1} \times \widehat{\mathbf{A}} \times \mathbf{P}^{-1}$. Then, right-multiple $\mathbf{P}$ on the both sides, (A.1.5) holds if and only if $\mathbf{I}_K - \widehat{\mathbf{\Delta}}^{-1} \times \widehat{\mathbf{B}} = \widehat{\mathbf{\Delta}}^{-1} \times \widehat{\mathbf{A}}$. Next, left-multiple $\widehat{\mathbf{\Delta}}$ on the both sides, yielding that (A.1.5) holds if and only if $\widehat{\mathbf{\Delta}} - \widehat{\mathbf{B}} \times \mathbf{P} = \widehat{\mathbf{A}}$, which holds by the definition of $\widehat{\mathbf{\Delta}}$.

Using (A.1.5) and substituting $\widehat{\mathbf{A}}_{\mathbf{\Theta}}$ and $\widehat{\mathbf{B}}_{\mathbf{\Theta}}$, we have

$$\widehat{\mathbf{\Theta}}\left[\widehat{\mathbf{A}}_{\mathbf{\Theta}}, \widehat{\mathbf{B}}_{\mathbf{\Theta}}, \boldsymbol{p}\right] = \widehat{\mathbf{A}}^{-1} \circ \mathbf{I}[\boldsymbol{p}] - \left(\widehat{\mathbf{A}} \times \mathbf{P}\right)^{-1} \circ \mathbf{J}[\boldsymbol{p}] + \left(\mathbf{P} \times \widehat{\mathbf{\Delta}}\right)^{-1} \circ \mathbf{J}[\boldsymbol{p}].$$

Since both $\widehat{\mathbf{A}}$ and $\mathbf{P}$ are diagonal matrices, $\widehat{\mathbf{A}}^{-1} \circ \mathbf{I}[\boldsymbol{p}] - \left(\widehat{\mathbf{A}} \times \mathbf{P}\right)^{-1} \circ \mathbf{J}[\boldsymbol{p}]$ is a block diagonal matrix $\mathrm{Bdiag}\left(\widehat{a}_{11}^{-1}\mathbf{I}_{p_1} - \widehat{a}_{11}^{-1}p_1^{-1}\mathbf{J}_{p_1}, \ldots, \widehat{a}_{KK}^{-1}\mathbf{I}_{p_K} - \widehat{a}_{KK}^{-1}p_K^{-1}\mathbf{J}_{p_K}\right)$.

Let $\mathbf{W}_k = \mathrm{Bdiag}\left(\mathbf{0}_{p_1 \times p_1}, \ldots, \widehat{a}_{kk}^{-1}\mathbf{I}_{p_k} - \widehat{a}_{kk}^{-1}p_k^{-1}\mathbf{J}_{p_k}, \ldots, \mathbf{0}_{p_K \times p_K}\right)$ for every k, which is a UB matrix, expressed by $\mathbf{W}_k = \mathbf{W}_k\left[\mathbf{A}_k, \mathbf{B}_k, \boldsymbol{p}\right]$ for every k, where both $\mathbf{A}_k = \mathrm{diag}\left(0, \ldots, \widehat{a}_{kk}^{-1}, \ldots, 0\right)$ and $\mathbf{B}_k = \mathrm{diag}\left(0, \ldots, -\widehat{a}_{kk}p_k^{-1}, \ldots, 0\right)$ have the non-zero values on the (k,k)-th elements.

Herein, U can be written as a sum of $(K+1)$ components:

$$\begin{aligned} U &= n\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)^{\top}\left(\sum_{k=1}^{K}\mathbf{W}_k\right)\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)+n\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)^{\top}\left[\left(\mathbf{P}\times\widehat{\boldsymbol{\Delta}}\right)^{-1}\circ\mathbf{J}[\boldsymbol{p}]\right]\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right) \\ &= \sum_{k=1}^{K} n\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)^{\top}\mathbf{W}_k\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)+n\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)^{\top}\left[\left(\mathbf{P}\times\widehat{\boldsymbol{\Delta}}\right)^{-1}\circ\mathbf{J}[\boldsymbol{p}]\right]\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right) \\ &\equiv F_1+F_2+\cdots+F_K+F_{K+1}, \end{aligned}$$

where F_k denotes $n\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)^{\top}\mathbf{W}_k\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)$, for $k=1,\ldots,K$, and F_{K+1} denotes the last term $n\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)^{\top}\left[\left(\mathbf{P}\times\widehat{\boldsymbol{\Delta}}\right)^{-1}\circ\mathbf{J}[\boldsymbol{p}]\right]\left(\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0\right)$. Next, we will show these F_k are mutually independent in Step 2 and each follows a F-distribution in Step 3.

Step 2: independence. Recall $\bar{\boldsymbol{X}}-\boldsymbol{\mu}_0$ follows $N\left(\boldsymbol{\mu}-\boldsymbol{\mu}_0, n^{-1}\boldsymbol{\Sigma}\left[\mathbf{A},\mathbf{B},\boldsymbol{p}\right]\right)$. To show $F_1,\ldots,F_{K+1}$ are mutually independent, by Lemma A.1.1, we need to check $\mathbf{W}_k\times\left(n^{-1}\boldsymbol{\Sigma}\left[\mathbf{A},\mathbf{B},\boldsymbol{p}\right]\right)\times\mathbf{W}_{k'}=\mathbf{0}_{p\times p}$ for every $k\neq k'$ and $\mathbf{W}_k\times\left(n^{-1}\boldsymbol{\Sigma}\left[\mathbf{A},\mathbf{B},\boldsymbol{p}\right]\right)\times\left[\left(\mathbf{P}\times\widehat{\boldsymbol{\Delta}}\right)^{-1}\circ\mathbf{J}[\boldsymbol{p}]\right]=\mathbf{0}_{p\times p}$ for every k. It is easy to check the former holds by using the representation of $\mathbf{W}_k=\mathbf{W}_k\left[\mathbf{A}_k,\mathbf{B}_k,\boldsymbol{p}\right]$:

$$\begin{aligned} &\mathbf{W}_k\times\boldsymbol{\Sigma}\left[\mathbf{A},\mathbf{B},\boldsymbol{p}\right]\times\mathbf{W}_{k'} \\ &\quad=\left(\mathbf{A}_k\circ\mathbf{I}[\boldsymbol{p}]+\mathbf{B}_k\circ\mathbf{J}[\boldsymbol{p}]\right)\times\left(\mathbf{A}\circ\mathbf{I}[\boldsymbol{p}]+\mathbf{B}\circ\mathbf{J}[\boldsymbol{p}]\right)\times\left(\mathbf{A}_{k'}\circ\mathbf{I}[\boldsymbol{p}]+\mathbf{B}_{k'}\circ\mathbf{J}[\boldsymbol{p}]\right) \\ &\quad=\left[\left(\mathbf{A}_k\mathbf{A}\right)\circ\mathbf{I}[\boldsymbol{p}]+\left(\mathbf{A}_k\mathbf{B}+\mathbf{B}_k\mathbf{A}+\mathbf{B}_k\mathbf{P}\mathbf{B}\right)\circ\mathbf{J}[\boldsymbol{p}]\right]\times\left(\mathbf{A}_{k'}\circ\mathbf{I}[\boldsymbol{p}]+\mathbf{B}_{k'}\circ\mathbf{J}[\boldsymbol{p}]\right) \\ &\quad=\left(\mathbf{A}_k\mathbf{A}\mathbf{A}_{k'}\right)\circ\mathbf{I}[\boldsymbol{p}] \\ &\quad\quad+\left[\left(\mathbf{A}_k\mathbf{B}+\mathbf{B}_k\mathbf{A}+\mathbf{B}_k\mathbf{P}\mathbf{B}\right)\mathbf{A}_{k'}+\mathbf{A}_k\mathbf{A}\mathbf{B}_{k'}+\left(\mathbf{A}_k\mathbf{B}+\mathbf{B}_k\mathbf{A}+\mathbf{B}_k\mathbf{P}\mathbf{B}\right)\mathbf{P}\mathbf{B}_{k'}\right]\circ\mathbf{J}[\boldsymbol{p}] \\ &\quad=\mathbf{0}_{K\times K}\circ\mathbf{I}[\boldsymbol{p}]+\mathbf{0}_{K\times K}\circ\mathbf{J}[\boldsymbol{p}] \\ &\quad=\mathbf{0}_{p\times p}. \end{aligned}$$

To check if the latter one holds,

$$\begin{aligned}
&\mathbf{W}_k \times \mathbf{\Sigma}\left[\mathbf{A}, \mathbf{B}, \boldsymbol{p}\right] \times \left[\left(\mathbf{P} \times \widehat{\mathbf{\Delta}}\right)^{-1} \circ \mathbf{J}[\boldsymbol{p}]\right] \\
&= \left(\mathbf{A}_k \circ \mathbf{I}[\boldsymbol{p}] + \mathbf{B}_k \circ \mathbf{J}[\boldsymbol{p}]\right) \times \left(\mathbf{A} \circ \mathbf{I}[\boldsymbol{p}] + \mathbf{B} \circ \mathbf{J}[\boldsymbol{p}]\right) \times \left[\left(\mathbf{P} \times \widehat{\mathbf{\Delta}}\right)^{-1} \circ \mathbf{J}[\boldsymbol{p}]\right] \\
&= \left[\left(\mathbf{A}_k\mathbf{A}\right) \circ \mathbf{I}[\boldsymbol{p}] + \left(\mathbf{A}_k\mathbf{B} + \mathbf{B}_k\mathbf{A} + \mathbf{B}_k\mathbf{P}\mathbf{B}\right) \circ \mathbf{J}[\boldsymbol{p}]\right] \times \left[\left(\mathbf{P}\widehat{\mathbf{\Delta}}\right)^{-1} \circ \mathbf{J}[\boldsymbol{p}]\right] \\
&= \left[\mathbf{A}_k\mathbf{A}\left(\mathbf{P}\widehat{\mathbf{\Delta}}\right)^{-1} + \left(\mathbf{A}_k\mathbf{B} + \mathbf{B}_k\mathbf{A} + \mathbf{B}_k\mathbf{P}\mathbf{B}\right)\mathbf{P}\left(\mathbf{P}\widehat{\mathbf{\Delta}}\right)^{-1}\right] \circ \mathbf{J}[\boldsymbol{p}] \\
&= \left[\left(\left(\mathbf{A}_k + \mathbf{B}_k\mathbf{P}\right)\mathbf{A} + \left(\mathbf{A}_k + \mathbf{B}_k\mathbf{P}\right)\mathbf{B}\mathbf{P}\right)\left(\mathbf{P}\widehat{\mathbf{\Delta}}\right)^{-1}\right] \circ \mathbf{J}[\boldsymbol{p}] \\
&= \mathbf{0}_{K\times K} \circ \mathbf{J}[\boldsymbol{p}] \\
&= \mathbf{0}_{p\times p},
\end{aligned}$$

using $\mathbf{A}_k + \mathbf{B}_k\mathbf{P} = \mathbf{0}_{K\times K}$.

Step 3: distribution. Now, we specify the distributions for $F_1, \ldots, F_{K+1}$ respectively. Furthermore, let $\bar{\boldsymbol{X}} = \left(\bar{\boldsymbol{X}}^{(1),\top}, \ldots, \bar{\boldsymbol{X}}^{(K),\top}\right)^\top$ and $\boldsymbol{\mu}_0 = \left(\boldsymbol{\mu}_0^{(1),\top}, \ldots, \boldsymbol{\mu}_0^{(K),\top}\right)^\top$, where $\bar{\boldsymbol{X}}^{(k)}, \boldsymbol{\mu}_0^{(k)} \in \mathbb{R}^{p_k}$ for every k. Focus on the first K component, i.e., F_k for $k = 1, \ldots, K$. Substituting the block diagonal matrix $\mathbf{W}_k$, we have

$$F_k = n\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right)^\top \mathbf{W}_k \left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right) = \left(\bar{\boldsymbol{X}}^{(k)} - \boldsymbol{\mu}_0^{(k)}\right)^\top \left(\frac{n}{\widehat{a}_{kk}}\mathbf{I}_{p_k} - \frac{n}{\widehat{a}_{kk}p_k}\mathbf{J}_{p_k}\right)\left(\bar{\boldsymbol{X}}^{(k)} - \boldsymbol{\mu}_0^{(k)}\right).$$

Let $\mathbf{U}_k = n\left(\bar{\boldsymbol{X}}^{(k)} - \boldsymbol{\mu}_0^{(k)}\right)^\top \left[(1/a_{kk})\mathbf{I}_{p_k} - 1/(a_{kk}p_k)\mathbf{J}_{p_k}\right]\left(\bar{\boldsymbol{X}}^{(k)} - \boldsymbol{\mu}_0^{(k)}\right)$ and $\mathbf{V}_k = (p_k - 1)(n - 1)(\widehat{a}_{kk}/a_{kk})$ for every k. Thus, $\mathbf{U}_k/\mathbf{V}_k = F_k/((p_k - 1)(n-1))$ for every k. Let $\mathbf{M}_k = \mathbf{I}_{p_k} - \mathbf{J}_{p_k}/p_k$ for every k. It is clear that $\mathbf{M}_k^2 = \left(\mathbf{I}_{p_k} - \mathbf{J}_{p_k}/p_k\right)^2 = \mathbf{I}_{p_k} + \mathbf{J}_{p_k}/p_k - 2\mathbf{J}_{p_k}/p_k = \mathbf{M}_k$ therefore $\mathbf{M}_k$

is idempotent. We can observe that

$$\begin{aligned}\operatorname{tr}\left(\mathbf{M}_k\mathbf{S}_{kk}\mathbf{M}_k\right) &= \operatorname{tr}\left(\mathbf{M}_k\mathbf{S}_{kk}\right) = \operatorname{tr}\left(\mathbf{S}_{kk} - \mathbf{1}_{p_k\times 1}\mathbf{1}_{p_k\times 1}^\top\mathbf{S}_{kk}/p_k\right) = \operatorname{tr}\left(\mathbf{S}_{kk}\right) - \operatorname{tr}\left(\mathbf{1}_{p_k\times 1}^\top\mathbf{S}_{kk}\mathbf{1}_{p_k\times 1}\right)/p_k \\ &= \operatorname{tr}\left(\mathbf{S}_{kk}\right) - \operatorname{sum}\left(\mathbf{S}_{kk}\right)/p_k = (p_k-1)\widehat{a}_{kk}.\end{aligned}$$

Thus, by the fact that $(n-1)\mathbf{S}_{kk} \sim \text{Wishart}\,(n-1,\boldsymbol{\Sigma}_{kk})$, where $\boldsymbol{\Sigma}_{kk}$ is positive definite for every k because $\left(\mathbf{0}_{1\times p_1},\ldots,\boldsymbol{\alpha}^\top,\ldots,\mathbf{0}_{1\times p_K}\right)\boldsymbol{\Sigma}\left(\mathbf{0}_{1\times p_1},\ldots,\boldsymbol{\alpha}^\top,\ldots,\mathbf{0}_{1\times p_K}\right)^\top = \boldsymbol{\alpha}^\top\boldsymbol{\Sigma}_{kk}\boldsymbol{\alpha} > 0$ for any $\boldsymbol{\alpha}\in\mathbb{R}^{p_k}$. So, $(n-1)\mathbf{S}_{kk}$ can be expressed as $\sum_{j=1}^{n-1}\boldsymbol{Z}_j^{(k)}\boldsymbol{Z}_j^{(k),\top}$ where $\boldsymbol{Z}_1^{(k)},\ldots,\boldsymbol{Z}_{n-1}^{(k)}$ are mutually independently distributed as $N\left(\mathbf{0}_{p_k\times 1},\boldsymbol{\Sigma}_{kk}\right)$ for every k. We then observe that

$$\begin{aligned}a_{kk}\mathbf{V}_k = (n-1)(p_k-1)\widehat{a}_{kk} &= \operatorname{tr}\left(\mathbf{M}_k\left[(n-1)\mathbf{S}_{kk}\right]\mathbf{M}_k\right) = \operatorname{tr}\left(\mathbf{M}_k\left(\sum_{j=1}^{n-1}\boldsymbol{Z}_j^{(k)}\boldsymbol{Z}_j^{(k),\top}\right)\mathbf{M}_k\right) \\ &= \sum_{j=1}^{n-1}\boldsymbol{Z}_j^{(k),\top}\mathbf{M}_k\boldsymbol{Z}_j^{(k)}.\end{aligned}$$

On the one hand, $\bar{\boldsymbol{X}}$ and $\mathbf{S}$ are independent, so $\mathbf{U}_k$ and $\mathbf{V}_k$ for every k are mutually independent. On the other hand, $\sqrt{n}\left(\bar{\boldsymbol{X}}^{(k)} - \boldsymbol{\mu}_0^{(k)}\right) \sim N\left(\boldsymbol{\mu}^{(k)} - \boldsymbol{\mu}_0^{(k)},\boldsymbol{\Sigma}_{kk}\right)$. By Lemma A.1.2 and

$$\left(\frac{1}{a_{kk}}\mathbf{I}_{p_k} - \frac{1}{a_{kk}p_k}\mathbf{J}_{p_k}\right)\boldsymbol{\Sigma}_{kk} = \left(\frac{1}{a_{kk}}\mathbf{I}_{p_k} - \frac{1}{a_{kk}p_k}\mathbf{J}_{p_k}\right)\left(a_{kk}\mathbf{I}_{p_k} + b_{kk}\mathbf{J}_{p_k}\right) = \mathbf{I}_{p_K} - p_k^{-1}\mathbf{J}_{p_k},$$

which is idempotent with rank (p_k-1). Therefore, $\mathbf{U}_k$ follows a $\chi^2(\lambda_k)$ distribution with degree of freedom (p_k-1), and the noncentrality parameter is given by

$$\delta_k = \frac{1}{2}\left(\boldsymbol{\mu}^{(k)} - \boldsymbol{\mu}_0^{(k)}\right)^\top\left(a_{kk}^{-1}\mathbf{I}_{p_k} - a_{kk}^{-1}p_k^{-1}\mathbf{J}_{p_k}\right)\left(\boldsymbol{\mu}^{(k)} - \boldsymbol{\mu}_0^{(k)}\right), \quad k=1,\ldots,K.$$

Since $\mathbf{M}_k = \mathbf{I}_{p_k} - p_k^{-1}\mathbf{J}_{p_k}$ is also idempotent with rank $(p_k - 1)$. Therefore, $a_{kk}\mathbf{V}_k$ follows a central χ^2-distribution with degree of freedom $(n-1)(p_k-1)$. Furthermore,

$$F_k = (p_k - 1)\frac{\mathbf{U}_k/(p_k-1)}{\mathbf{V}_k/((p_k-1)(n-1))} \sim (p_k-1)F^{(k)}_{(p_k-1),(n-1)(p_k-1)}(\delta_k), \quad k = 1, \ldots, K.$$

For $F_{K+1} = n\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right)^\top \left[\left(\mathbf{P}\widehat{\boldsymbol{\Delta}}\right)^{-1} \circ \mathbf{J}[\boldsymbol{p}]\right]\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right)$, consider a transformation $\boldsymbol{Y} = \mathbf{C}\boldsymbol{X}$, where $\mathbf{C} = \mathrm{Bdiag}\left(\mathbf{1}_{1\times p_1}/p_1, \ldots, \mathbf{1}_{1\times p_K}/p_K\right) \in \mathbb{R}^{K\times p}$. As $\boldsymbol{X} \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}[\mathbf{A}, \mathbf{B}, \boldsymbol{p}]\right)$, $\boldsymbol{Y} \sim N\left(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y\right)$, where

$$\boldsymbol{\mu}_y = \mathbf{C}\boldsymbol{\mu} = \begin{pmatrix} \mathbf{1}_{1\times p_1}\boldsymbol{\mu}^{(1)}/p_1 \\ \vdots \\ \mathbf{1}_{1\times p_K}\boldsymbol{\mu}^{(K)}/p_K \end{pmatrix}, \quad \boldsymbol{\Sigma}_y = \mathbf{C} \times \boldsymbol{\Sigma}[\mathbf{A}, \mathbf{B}, \boldsymbol{p}] \times \mathbf{C}^\top = \mathbf{A} \times \mathbf{P}^{-1} + \mathbf{B}.$$

Furthermore, let $\boldsymbol{\nu}_0 = \mathbf{C} \times \boldsymbol{\mu}_0 = \left(\mathbf{1}_{1\times p_1}\boldsymbol{\mu}_0^{(1)}/p_1, \ldots, \mathbf{1}_{1\times p_K}\boldsymbol{\mu}_0^{(K)}/p_K\right)^\top$. By noting that $(\mathbf{PC})^\top \times \boldsymbol{\Gamma} \times (\mathbf{PC}) = \boldsymbol{\Gamma} \circ \mathbf{J}[\boldsymbol{p}]$ for any $\boldsymbol{\Gamma} \in \mathbb{R}^{K\times K}$, then,

$$\begin{aligned}
F_{K+1} &= n\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right)^\top \left[\left(\mathbf{P}\widehat{\boldsymbol{\Delta}}\right)^{-1} \circ \mathbf{J}[\boldsymbol{p}]\right]\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right) \\
&= n\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right)^\top \left[(\mathbf{PC})^\top \left(\mathbf{P}\widehat{\boldsymbol{\Delta}}\right)^{-1} (\mathbf{PC})\right]\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right) \\
&= n\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right)^\top \left[\mathbf{C}^\top\mathbf{P}\widehat{\boldsymbol{\Delta}}^{-1}\mathbf{P}^{-1}\mathbf{PC}\right]\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0\right) \\
&= n\left(\bar{\boldsymbol{Y}} - \boldsymbol{\nu}_0\right)^\top \left(\mathbf{P}\widehat{\boldsymbol{\Delta}}^{-1}\right)\left(\bar{\boldsymbol{Y}} - \boldsymbol{\nu}_0\right) \\
&= n\left(\bar{\boldsymbol{Y}} - \boldsymbol{\nu}_0\right)^\top \left(\widehat{\boldsymbol{\Sigma}}_y^{-1}\right)\left(\bar{\boldsymbol{Y}} - \boldsymbol{\nu}_0\right),
\end{aligned}$$

where $\mathbf{P}\widehat{\boldsymbol{\Delta}}^{-1} = \mathbf{P}\left(\widehat{\mathbf{A}}\mathbf{P}^{-1}\mathbf{P} + \widehat{\mathbf{B}}\mathbf{P}\right)^{-1} = \mathbf{P}\left[\left(\widehat{\mathbf{A}}\mathbf{P}^{-1} + \widehat{\mathbf{B}}\right)\mathbf{P}\right]^{-1} = \mathbf{P}\mathbf{P}^{-1}\left(\widehat{\mathbf{A}}\mathbf{P}^{-1} + \widehat{\mathbf{B}}\right)^{-1} = \widehat{\boldsymbol{\Sigma}}_y^{-1}$.

By the definition of the Hotelling's T^2-statistic, $F_{K+1} \sim T^2 = \frac{K(n-1)}{n-K} F^{(K+1)}_{K,n-K}(\delta_{K+1})$, where $\delta_{K+1} = \frac{1}{2} n \left(\boldsymbol{\mu}_y - \boldsymbol{\nu}_0\right)^\top \boldsymbol{\Sigma}_y^{-1} \left(\boldsymbol{\mu}_y - \boldsymbol{\nu}_0\right)$.

Finally, by the mutual independence of $F_1, \ldots, F_K, F_{K+1}$, U is decomposed as a linear combination of mutually independent F-variates, distributed as $\sum_{k=1}^{K}(p_k - 1) F^{(k)}_{(p_k-1),(n-1)(p_k-1)}(\delta_k) + \frac{K(n-1)}{n-K} F^{(K+1)}_{K,n-K}(\delta_{K+1})$. Under H_0, $\delta_k = 0$ for $K = 1, \ldots, K+1$. ■

Proof of Theorem 2.4.1. We also divide the proof into three steps.

Step 1: decomposition of U_M. Let $\boldsymbol{Z} = \left(\bar{\boldsymbol{X}}^{(1),\top}, \ldots, \bar{\boldsymbol{X}}^{(M),\top}\right)^\top$ denote a (pM)-dimensional normal vector with mean $\boldsymbol{\mu}_{\boldsymbol{Z}} = \left(\boldsymbol{\mu}^{(1),\top}, \ldots, \boldsymbol{\mu}^{(M),\top}\right)^\top$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{Z}} = \mathbf{N}^{-1} \otimes \boldsymbol{\Sigma}\left[\mathbf{A}, \mathbf{B}, \boldsymbol{p}\right]$, where $\mathbf{N} = \operatorname{diag}\left(n_1, \ldots, n_M\right) \in \mathbb{R}^{M \times M}$ and $\otimes$ denotes the Kronecker product. Thus, there exits $\mathbf{C}_{\boldsymbol{Z}} = \left[\mathbf{N}^{1/2} \times \left(\mathbf{I}_M - n^{-1}\mathbf{J}_M \times \mathbf{N}\right)\right] \otimes \mathbf{I}_p \in \mathbb{R}^{(pM)\times(pM)}$ such that

$$\left(\sqrt{n_1}\left(\bar{\boldsymbol{X}}^{(1)} - \bar{\boldsymbol{X}}\right)^\top, \ldots, \sqrt{n_M}\left(\bar{\boldsymbol{X}}^{(M)} - \bar{\boldsymbol{X}}\right)^\top\right)^\top = \mathbf{C}_Z \times \boldsymbol{Z}.$$

Therefore, U_M can be rewritten as

$$U_M = \left(\mathbf{C}_{\boldsymbol{Z}} \times \boldsymbol{Z}\right)^\top \left\{\mathbf{I}_M \otimes \widehat{\boldsymbol{\Theta}}\left[\widehat{\mathbf{A}}_{\boldsymbol{\Theta}}, \widehat{\mathbf{B}}_{\boldsymbol{\Theta}}, \boldsymbol{p}\right]\right\}\left(\mathbf{C}_{\boldsymbol{Z}} \times \boldsymbol{Z}\right).$$

Using the same decomposition, $\widehat{\boldsymbol{\Theta}}\left[\widehat{\mathbf{A}}_{\boldsymbol{\Theta}}, \widehat{\mathbf{B}}_{\boldsymbol{\Theta}}, \boldsymbol{p}\right] = \widehat{\mathbf{A}}^{-1} \circ \mathbf{I}[\boldsymbol{p}] - \left(\widehat{\mathbf{A}} \times \mathbf{P}\right)^{-1} \circ \mathbf{J}[\boldsymbol{p}] + \left(\mathbf{P} \times \widehat{\boldsymbol{\Delta}}\right)^{-1} \circ$

$\mathbf{J}[\boldsymbol{p}]$, U_M can be expressed by

$$\begin{aligned} U_M &= (\mathbf{C}_{\boldsymbol{Z}} \times \boldsymbol{Z})^\top \left[\mathbf{I}_M \otimes \left(\sum_{k=1}^{K} \mathbf{W}_k\right)\right] (\mathbf{C}_{\boldsymbol{Z}} \times \boldsymbol{Z}) + (\mathbf{C}_{\boldsymbol{Z}} \times \boldsymbol{Z})^\top (\mathbf{I}_M \otimes \mathbf{W}_{K+1}) (\mathbf{C}_{\boldsymbol{Z}} \times \boldsymbol{Z}) \\ &= \sum_{k=1}^{K} (\mathbf{C}_{\boldsymbol{Z}} \times \boldsymbol{Z})^\top (\mathbf{I}_M \otimes \mathbf{W}_k) (\mathbf{C}_{\boldsymbol{Z}} \times \boldsymbol{Z}) + (\mathbf{C}_{\boldsymbol{Z}} \times \boldsymbol{Z})^\top (\mathbf{I}_M \otimes \mathbf{W}_{K+1}) (\mathbf{C}_{\boldsymbol{Z}} \times \boldsymbol{Z}) \\ &\equiv F_1 + \cdots + F_K + F_{K+1}, \end{aligned}$$

where $\mathbf{W}_k = \text{Bdiag}\left(\mathbf{0}_{p_1 \times p_1}, \ldots, \widehat{a}_{kk}^{-1} - \widehat{a}_{kk}^{-1} p_k^{-1} \mathbf{J}_{p_k}, \ldots, \mathbf{0}_{p_K \times p_K}\right) = \mathbf{W}_k[\mathbf{A}_k, \mathbf{B}_k, \boldsymbol{p}]$ with $\mathbf{A}_k = \text{diag}\left(0, \ldots, \widehat{a}_{kk}^{-1}, \ldots, 0\right) \in \mathbb{R}^{K \times K}$ and $\mathbf{B}_k = \text{diag}\left(0, \ldots, -\widehat{a}_{kk}^{-1} p_k^{-1}, \ldots, 0\right) \in \mathbb{R}^{K \times K}$ having the non-zero values on the (k,k)-th elements, and $\mathbf{W}_{K+1} = \left(\mathbf{P} \times \widehat{\boldsymbol{\Delta}}\right)^{-1} \circ \mathbf{J}[\boldsymbol{p}]$. Next, we need to prove the mutual independence between F_1, ..., F_{K+1} and specify the distribution for each of them.

Step 2: independence. By Lemma A.1.1, we need to check

$$\left[\mathbf{C}_{\boldsymbol{Z}}^\top (\mathbf{I}_M \otimes \mathbf{W}_k) \mathbf{C}_{\boldsymbol{Z}}\right] \left(\mathbf{N}^{-1} \otimes [\mathbf{A}, \mathbf{B}, \boldsymbol{p}]\right) \left[\mathbf{C}_{\boldsymbol{Z}}^\top (\mathbf{I}_M \otimes \mathbf{W}_{k'}) \mathbf{C}_{\boldsymbol{Z}}\right] = \mathbf{0}_{(pM)\times(pM)}, \quad k \neq k', k, k' = 1, \ldots, K$$

$$\left[\mathbf{C}_{\boldsymbol{Z}}^\top (\mathbf{I}_M \otimes \mathbf{W}_k) \mathbf{C}_{\boldsymbol{Z}}\right] \left(\mathbf{N}^{-1} \otimes \boldsymbol{\Sigma}[\mathbf{A}, \mathbf{B}, \boldsymbol{p}]\right) \left[\mathbf{C}_{\boldsymbol{Z}}^\top (\mathbf{I}_M \otimes \mathbf{W}_{K+1}) \mathbf{C}_{\boldsymbol{Z}}\right] = \mathbf{0}_{(pM)\times(pM)}, \quad k = 1, \ldots, K.$$

Let $\mathbf{M} = \mathbf{N}^{1/2} \times (\mathbf{I}_M - n^{-1}\mathbf{J}_M \times \mathbf{N})$. Given $k \neq k'$ and $\mathbf{C}_{\boldsymbol{Z}} = \mathbf{M} \otimes \mathbf{I}_p$ and the fact $(A \otimes B)^\top = A^\top \otimes B^\top$, we can observe that

$$\begin{aligned} \mathbf{C}_{\boldsymbol{Z}}^\top (\mathbf{I}_M \otimes \mathbf{W}_k) \mathbf{C}_{\boldsymbol{Z}} &= \left(\mathbf{M}^\top \mathbf{I}_M \mathbf{M}\right) \otimes (\mathbf{I}_p \mathbf{W}_k \mathbf{I}_p) = \left(\mathbf{M}^\top \mathbf{M}\right) \otimes \mathbf{W}_k, \\ \mathbf{C}_{\boldsymbol{Z}}^\top (\mathbf{I}_M \otimes \mathbf{W}_{k'}) \mathbf{C}_{\boldsymbol{Z}} &= \left(\mathbf{M}^\top \mathbf{M}\right) \otimes \mathbf{W}_{k'}. \end{aligned}$$

On the one hand, we can calculate the following result:

$$\left[\mathbf{C}_{\boldsymbol{Z}}^{\top}(\mathbf{I}_M \otimes \mathbf{W}_k)\mathbf{C}_{\boldsymbol{Z}}\right]\left(\mathbf{N}^{-1} \otimes \boldsymbol{\Sigma}[\mathbf{A}, \mathbf{B}, \boldsymbol{p}]\right)\left[\mathbf{C}_{\boldsymbol{Z}}^{\top}(\mathbf{I}_M \otimes \mathbf{W}_{k'})\mathbf{C}_{\boldsymbol{Z}}\right]$$
$$= \left[\left(\mathbf{M}^{\top}\mathbf{M}\right)\mathbf{N}^{-1}\left(\mathbf{M}^{\top}\mathbf{M}\right)\right] \otimes \left[\mathbf{W}_k \boldsymbol{\Sigma}\left[\mathbf{A}, \mathbf{B}, \boldsymbol{p}\right]\mathbf{W}_{k'}\right].$$

Noting that $\mathbf{W}_k = \mathbf{W}_k\left[\mathbf{A}_k, \mathbf{B}_k, \boldsymbol{p}\right]$ and $\mathbf{W}_{k'} = \mathbf{W}_{k'}\left[\mathbf{A}_{k'}, \mathbf{B}_{k'}, \boldsymbol{p}\right]$, we have the result $\mathbf{W}_k \times \boldsymbol{\Sigma}\left[\mathbf{A}, \mathbf{B}, \boldsymbol{p}\right] \times \mathbf{W}_{k'} = \mathbf{0}_{p\times p}$. Therefore, $\left[\mathbf{C}_{\boldsymbol{Z}}^{\top}(\mathbf{I}_M \otimes \mathbf{W}_k)\mathbf{C}_{\boldsymbol{Z}}\right]\left(\mathbf{N}^{-1} \otimes \boldsymbol{\Sigma}[\mathbf{A}, \mathbf{B}, \boldsymbol{p}]\right)\left[\mathbf{C}_{\boldsymbol{Z}}^{\top}(\mathbf{I}_M \otimes \mathbf{W}_{k'})\mathbf{C}_{\boldsymbol{Z}}\right] = \mathbf{0}_{(pM)\times(pM)}$. On the other hand, we can calculate that

$$\left[\mathbf{C}_{\boldsymbol{Z}}^{\top}(\mathbf{I}_M \otimes \mathbf{W}_k)\mathbf{C}_{\boldsymbol{Z}}\right]\left(\mathbf{N}^{-1} \otimes \boldsymbol{\Sigma}[\mathbf{A}, \mathbf{B}, \boldsymbol{p}]\right)\left[\mathbf{C}_{\boldsymbol{Z}}^{\top}(\mathbf{I}_M \otimes \mathbf{W}_{K+1})\mathbf{C}_{\boldsymbol{Z}}\right]$$
$$= \left[\left(\mathbf{M}^{\top}\mathbf{M}\right)\mathbf{N}^{-1}\left(\mathbf{M}^{\top}\mathbf{M}\right)\right] \otimes \left[\mathbf{W}_k \boldsymbol{\Sigma}\left[\mathbf{A}, \mathbf{B}, \boldsymbol{p}\right]\mathbf{W}_{K+1}\right]$$
$$= \mathbf{0}_{(pM)\times(pM)}$$

using the result $\mathbf{W}_k \boldsymbol{\Sigma}\left[\mathbf{A}, \mathbf{B}, \boldsymbol{p}\right]\mathbf{W}_{K+1} = \mathbf{0}_{p\times p}$.

Step 3: distribution. By definition, for $k = 1, \ldots, K$

$$F_k = \boldsymbol{Z}^{\top} \times \left[\mathbf{C}_{\boldsymbol{Z}}^{\top}(\mathbf{I}_M \otimes \mathbf{W}_k)\mathbf{C}_{\boldsymbol{Z}}\right] \times \boldsymbol{Z} = \boldsymbol{Z}^{\top}\left[\left(\mathbf{M}^{\top}\mathbf{M}\right) \otimes \mathbf{W}_k\right]\boldsymbol{Z},$$

Let $\widetilde{\mathbf{W}}_k = \mathrm{Bdiag}\left(\mathbf{0}_{p_1\times p_1}, a_{kk}^{-1}\mathbf{I}_{p_k} - a_{kk}^{-1}p_k^{-1}\mathbf{J}_{p_k}, \ldots, \mathbf{0}_{p_K\times p_k}\right)$ have non-zero values on the (k,k)-th element for every k. Let $\mathbf{U}_k = \boldsymbol{Z}^{\top}\left(\left(\mathbf{M}^{\top}\mathbf{M}\right) \otimes \widetilde{\mathbf{W}}_k\right)\boldsymbol{Z}$ and $\mathbf{V}_k = (n-M)(p_k-1)\widehat{a}_{kk}/a_{kk}$, and therefore, $\mathbf{U}_k/\mathbf{V}_k = F_k/\left[(n-M)(p_k-1)\right]$ for $k = 1, \ldots, K$.

Since $\mathbf{S}$ is independent from $\bar{\boldsymbol{X}}^{(1)}, \ldots, \bar{\boldsymbol{X}}^{(M)}$, then $\mathbf{S}$ is independent from $\boldsymbol{Z}$, and therefore $\mathbf{U}_k$ and $\mathbf{V}_k$ are independent for every k. By Lemma A.1.2 and $\boldsymbol{Z} \sim N\left(\boldsymbol{\mu}_{\boldsymbol{Z}}, \mathbf{N}^{-1} \otimes \boldsymbol{\Sigma}[\mathbf{A}, \mathbf{B}, \boldsymbol{p}]\right)$,

we obtain that

$$\left[\left(\mathbf{M}^\top\mathbf{M}\right)\otimes\widetilde{\mathbf{W}}_k\right]\left[\mathbf{N}^{-1}\otimes\boldsymbol{\Sigma}\left[\mathbf{A},\mathbf{B},\boldsymbol{p}\right]\right]=\left(\mathbf{M}^\top\mathbf{M}\mathbf{N}^{-1}\right)\otimes\left(\widetilde{\mathbf{W}}_k\times\boldsymbol{\Sigma}[\mathbf{A},\mathbf{B},\boldsymbol{p}]\right),$$

whose square equals $\left(\mathbf{M}^\top\mathbf{M}\mathbf{N}^{-1}\right)^2\otimes\left(\widetilde{\mathbf{W}}_k\times\boldsymbol{\Sigma}[\mathbf{A},\mathbf{B},\boldsymbol{p}]\right)^2=\left(\mathbf{M}^\top\mathbf{M}\mathbf{N}^{-1}\right)\otimes\left(\widetilde{\mathbf{W}}_k\times\boldsymbol{\Sigma}[\mathbf{A},\mathbf{B},\boldsymbol{p}]\right)$ since

$$\begin{aligned}
\mathbf{M}^\top\mathbf{M}\mathbf{N}^{-1}&=\left(\mathbf{N}^{1/2}\left(\mathbf{I}_M-\frac{1}{n}\mathbf{J}_M\mathbf{N}\right)\right)^\top\times\left(\mathbf{N}^{1/2}\left(\mathbf{I}_M-\frac{1}{n}\mathbf{J}_M\mathbf{N}\right)\right)\times\mathbf{N}^{-1}\\
&=\left(\mathbf{I}_M-\frac{1}{n}\mathbf{N}\mathbf{J}_M\right)\times\mathbf{N}^{\frac{1}{2}}\times\mathbf{N}^{\frac{1}{2}}\times\left(\mathbf{I}_M-\frac{1}{n}\mathbf{J}_M\mathbf{N}\right)\times\mathbf{N}^{-1}\\
&=\left(\mathbf{N}-\frac{1}{n}\mathbf{N}\mathbf{J}_M\mathbf{N}\right)\times\left(\mathbf{N}^{-1}-\frac{1}{n}\mathbf{J}_M\right)\\
&=\mathbf{I}_M-\frac{1}{n}\mathbf{N}\mathbf{J}_M-\frac{1}{n}\mathbf{N}\mathbf{J}_M+\frac{1}{n^2}\mathbf{N}\mathbf{J}_M\mathbf{N}\mathbf{J}_M\\
&=\mathbf{I}_M-n^{-1}\mathbf{N}\mathbf{J}_M,
\end{aligned}$$

and using $\mathbf{J}_M\mathbf{N}\mathbf{J}_M=n\mathbf{J}_M$. It is clear that $\mathbf{I}_M-n^{-1}\mathbf{N}\mathbf{J}_M$ is an idempotent matrix since its square equals $\mathbf{I}_M+n^{-2}\mathbf{N}\mathbf{J}_M\mathbf{N}\mathbf{J}_M-2n^{-1}\mathbf{N}\mathbf{J}_M=\mathbf{I}_M-n^{-1}\mathbf{N}\mathbf{J}_M$. Also, the result shows that $\widetilde{\mathbf{W}}_k\times\boldsymbol{\Sigma}[\mathbf{A},\mathbf{B},\boldsymbol{p}]$ is an idempotent matrix.

Therefore, $\mathbf{U}_k$ follows a noncentral χ^2-distribution with the noncentrality parameter $\delta_k=\frac{1}{2}\boldsymbol{\mu}_{\boldsymbol{Z}}^\top\left[\left(\mathbf{M}^\top\mathbf{M}\right)\otimes\widetilde{\mathbf{W}}_k\right]\boldsymbol{\mu}_{\boldsymbol{Z}}$ and a degree of freedom $\operatorname{rank}\left(\left(\mathbf{M}^\top\mathbf{M}\right)\otimes\widetilde{\mathbf{W}}_k\right)=\operatorname{rank}\left(\mathbf{M}^\top\mathbf{M}\right)\times\operatorname{rank}\left(\widetilde{\mathbf{W}}_k\right)=(M-1)(p_k-1)$ by using the fact that $\operatorname{rank}\left(A\otimes B\right)=\operatorname{rank}(A)\operatorname{rank}(B)$.

Given $a_{kk}\mathbf{V}_k=(n-M)(p_k-1)\widehat{a}_{kk}$, the result that $(p_k-1)\widehat{a}_{kk}=\operatorname{tr}\left(\mathbf{M}_k\mathbf{S}_{kk}\mathbf{M}_{kk}\right)$, and the fact that $(n-M)\mathbf{S}_{kk}\sim\text{Wishart}(n-M,\boldsymbol{\Sigma}_{kk})$, we obtain that $a_{kk}\mathbf{V}_k$ follows a central χ^2-

distribution with a degree of freedom $(n-M)(p_k-1)$. Therefore, for $k=1,\ldots,K$,

$$\begin{aligned}
F_k &= (n-M)(p_k-1)\frac{\mathbf{U}_k}{\mathbf{V}_k} \\
&= (n-M)(p_k-1)\frac{\chi^2_{(M-1)(p_k-1)}(\delta_k)}{\chi^2_{(n-M)(p_k-1)}} \\
&= \frac{\chi^2_{(M-1)(p_k-1)}(\delta_k)/[(M-1)(p_k-1)]}{\chi^2_{(n-M)(p_k-1)}/[(n-M)(p_k-1)]}(n-M)(p_k-1)\times\frac{(M-1)(p_k-1)}{(n-M)(p_k-1)} \\
&= (M-1)(p_k-1)F^{(k)}_{(M-1)(p_k-1),(n-M)(p_k-1)}(\delta_k).
\end{aligned}$$

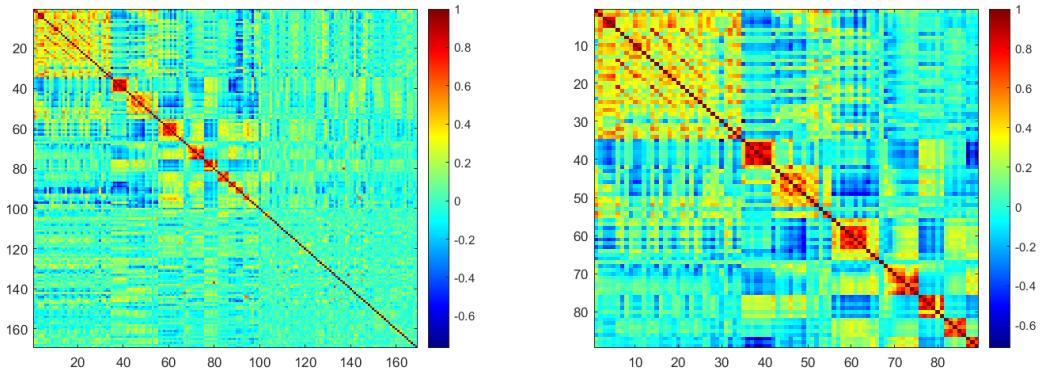
Let $\mathbf{C}_{\boldsymbol{Y}} = \mathrm{Bdiag}\left(\mathbf{1}_{1\times p_1}/p_1,\ldots,\mathbf{1}_{1\times p_K}/p_K\right)$, let $\boldsymbol{\nu}^{(m)} = \mathbf{C}_{\boldsymbol{Y}}\boldsymbol{\mu}^{(m)}$, and let $\boldsymbol{Y}_j^{(m)} = \mathbf{C}_{\boldsymbol{Y}}\boldsymbol{X}_j^{(m)}$ for every j and m with mean $\boldsymbol{\nu}^{(m)} = \mathbf{C}_{\boldsymbol{Y}}\times\boldsymbol{\mu}^{(m)}$ and $\boldsymbol{\Sigma}_{\boldsymbol{Y}} = \mathbf{C}_{\boldsymbol{Y}}\times\boldsymbol{\Sigma}[\mathbf{A},\mathbf{B},\boldsymbol{p}]\times\mathbf{C}_{\boldsymbol{Y}}^\top = \mathbf{A}\mathbf{P}^{-1}+\mathbf{B} = (\mathbf{A}+\mathbf{B}\mathbf{P})\mathbf{P}^{-1} = \boldsymbol{\Delta}\mathbf{P}^{-1} = (\mathbf{P}\boldsymbol{\Delta}^{-1})^{-1}$. Thus, $\bar{\boldsymbol{Y}}^{(m)} = \mathbf{C}_{\boldsymbol{Y}}\bar{\boldsymbol{X}}^{(m)}$ for every m and $\bar{\boldsymbol{Y}} = \mathbf{C}_{\boldsymbol{Y}}\bar{\boldsymbol{X}}$. Finally, F_{K+1} can be expressed by

$$\begin{aligned}
F_{K+1} &= \sum_{m=1}^{M} n_m\left(\bar{\boldsymbol{X}}^{(m)}-\bar{\boldsymbol{X}}\right)^\top\mathbf{W}_{K+1}\left(\bar{\boldsymbol{X}}^{(m)}-\bar{\boldsymbol{X}}\right) \\
&= \sum_{m=1}^{M} n_m\left(\bar{\boldsymbol{X}}^{(m)}-\bar{\boldsymbol{X}}\right)^\top\left[\mathbf{C}_{\boldsymbol{Y}}^\top\mathbf{P}\widehat{\boldsymbol{\Delta}}^{-1}\mathbf{P}^{-1}\mathbf{P}\mathbf{C}_{\boldsymbol{Y}}\right]\left(\bar{\boldsymbol{X}}^{(m)}-\bar{\boldsymbol{X}}\right) \\
&= \sum_{m=1}^{M} n_m\left(\bar{\boldsymbol{Y}}^{(m)}-\bar{\boldsymbol{Y}}\right)^\top\left(\mathbf{P}\widehat{\boldsymbol{\Delta}}^{-1}\right)\left(\bar{\boldsymbol{Y}}^{(m)}-\bar{\boldsymbol{Y}}\right) \\
&= \mathrm{tr}\left[\left(\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}}\right)^{-1}\sum_{m=1}^{M} n_m\left(\bar{\boldsymbol{Y}}^{(m)}-\bar{\boldsymbol{Y}}\right)\left(\bar{\boldsymbol{Y}}^{(m)}-\bar{\boldsymbol{Y}}\right)^\top\right],
\end{aligned}$$

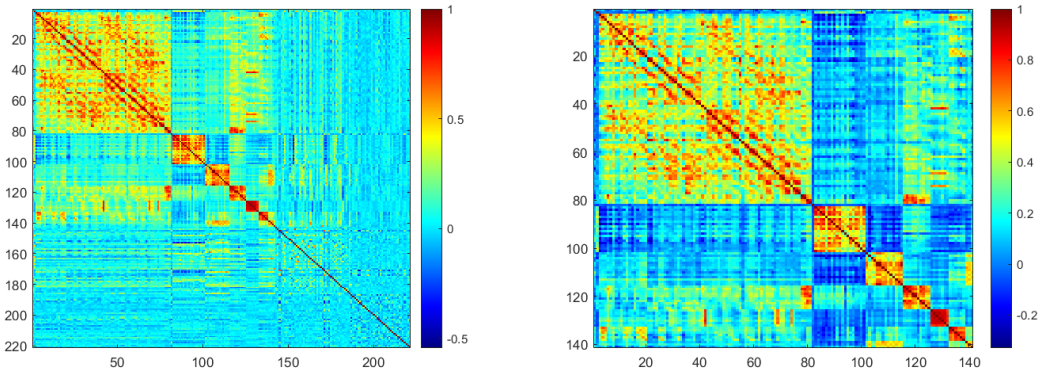
which is Hotelling's T_0^2-statistic. ■

Appendix B: Supplementary Materials for Chapter 3

B.1 Extra Data Examples



(a) correlations for exposome data



(b) correlations for metabolite data

Figure B.1: Heat maps for exposome and metabolite data in an environmental research

We present two extra real-data examples for the pattern of well-organized blocks in sample correlation matrices. In this study, there are 169 exposome variables and 221 metabolite variables

for 1192 subjects. First, 89 out of 169 exposome variables are identified to form a 9 by 9 well-organized blocks. Second, 141 out of 221 metabolite variables are identified to form a 7 by 7 well-organized blocks. In particular, Figure B.1a contains the heat maps of the sample correlation matrix calculated by the clustering method (Wu et al., 2021) for exposome variables from the ISGobal dataset (ISGlobal, 2021). Figure B.1b contains the heat maps of the sample correlation matrix calculated by the clustering method for metabolite variables from the ISGobal dataset. All data examples can be found at `https://github.com/yiorfun/UBCovEst`.

B.2 Exact Covariance Estimators for $\widetilde{\boldsymbol{\theta}}$

Corollary B.2.1. *Given a partition-size vector* $\boldsymbol{p} = (p_1, \ldots, p_K)^\top$ *and a UB matrix* $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_{kk'})$ *partitioned by* $\boldsymbol{p}$*, let* $\alpha_{kk} = \mathrm{tr}(\boldsymbol{\Sigma}_{kk})/p_k = a_{kk} + b_{kk}$ *denote the average of the diagonal entries in* $\boldsymbol{\Sigma}_{kk}$*, let* $\beta_{kk} = \mathrm{sum}(\boldsymbol{\Sigma}_{kk})/(p_k^2) = a_{kk}/p_k + b_{kk}$ *denote the average of all entries in* $\boldsymbol{\Sigma}_{kk}$ *for every* k*, and let* $\beta_{kk'} = \mathrm{sum}(\boldsymbol{\Sigma}_{kk'})/(p_k p_{k'}) = b_{kk'}$ *denote the average of all entries in* $\boldsymbol{\Sigma}_{kk'}$ *for every* $k \neq k'$*. Let* $\widetilde{\alpha}_{kk} = \mathrm{tr}(\boldsymbol{S}_{kk})/p_k$*,* $\widetilde{\beta}_{kk} = \mathrm{sum}(\boldsymbol{S}_{kk})/p_k^2$ *for every* k*, and* $\widetilde{\beta}_{kk'} = \mathrm{sum}(\boldsymbol{S}_{kk'})/(p_k p_{k'})$ *for every* $k \neq k'$*.*

(1) $\widetilde{\alpha}_{kk}$*,* $\widetilde{\beta}_{kk'}$ *are the best unbiased estimators of* α_{kk}*,* $\beta_{kk'}$ *for every* k *and* k'*; the variances of* $\widetilde{\alpha}_{kk}$ *and* $\widetilde{\beta}_{kk'}$ *are*

$$\mathrm{var}\left(\widetilde{\alpha}_{kk}\right) = \frac{2}{(n-1)p_k}\left(a_{kk}^2 + 2a_{kk}b_{kk} + p_k b_{kk}^2\right),$$

$$\mathrm{var}\left(\widetilde{\beta}_{kk'}\right) = \begin{cases} \dfrac{2}{(n-1)p_k^2}\left(a_{kk} + p_k b_{kk}\right)^2, & k = k' \\ \dfrac{1}{2(n-1)p_k p_{k'}}\left\{p_k p_{k'}\left(b_{kk'}^2 + b_{k'k}^2\right) + 2\left(a_{kk} + p_k b_{kk}\right)\left(a_{k'k'} + p_{k'} b_{k'k'}\right)\right\}, & k \neq k' \end{cases},$$

respectively, for every k and k'; the covariance between $\widetilde{\alpha}_{kk}$, $\widetilde{\alpha}_{k'k'}$, $\widetilde{\beta}_{kk}$, and $\widetilde{\beta}_{k'k'}$ are

$$\operatorname{cov}\left(\widetilde{\alpha}_{kk}, \widetilde{\alpha}_{k'k'}\right) = \frac{2}{n-1} b_{kk'} b_{k'k}, \quad \operatorname{cov}\left(\widetilde{\beta}_{kk}, \widetilde{\beta}_{k'k'}\right) = \frac{2}{n-1} b_{kk'} b_{k'k}, \quad k \neq k'$$

$$\operatorname{cov}\left(\widetilde{\alpha}_{kk}, \widetilde{\beta}_{k'k'}\right) = \begin{cases} \dfrac{2}{(n-1)p_k^2} \left(a_{kk} + p_k b_{kk}\right)^2, & k = k' \\ \dfrac{2}{n-1} b_{kk'} b_{k'k}, & k \neq k' \end{cases},$$

respectively, for every k and k'; the covariance between $\widetilde{\beta}_{kk'}$ and the other estimators are

$$\operatorname{cov}\left(\widetilde{\alpha}_{kk}, \widetilde{\beta}_{k'k''}\right) = \frac{1}{(n-1)p_{k'}p_{k''}} \begin{cases} p_{k'}p_{k''}\left(b_{k'k}b_{kk''} + b_{k''k}b_{kk'}\right), & k \neq k', k \neq k'' \\ p_{k''}\left(b_{k'k''} + b_{k''k'}\right)\left(a_{kk} + p_k b_{kk}\right), & k = k' \\ p_{k'}\left(b_{k'k''} + b_{k''k'}\right)\left(a_{kk} + p_k b_{kk}\right), & k = k'' \end{cases},$$

$$\operatorname{cov}\left(\widetilde{\beta}_{kk}, \widetilde{\beta}_{k'k''}\right) = \frac{1}{(n-1)p_k} \begin{cases} p_k\left(b_{k'k}b_{kk''} + b_{k''k}b_{kk'}\right), & k \neq k', k \neq k'' \\ \left(a_{kk} + p_k b_{kk}\right)\left(b_{k'k''} + b_{k''k'}\right), & k = k' \\ \left(a_{kk} + p_k b_{kk}\right)\left(b_{k'k''} + b_{k''k'}\right), & k = k'' \end{cases},$$

respectively, for every k, $k' \neq k''$; and

$$\operatorname{cov}\left(\widetilde{\beta}_{k_1k_2}, \widetilde{\beta}_{l_1,l_2}\right) = \frac{1}{2(n-1)}$$

$$\begin{cases}
b_{l_1k_1}b_{k_2l_2} + b_{l_2k_1}b_{k_2l_1} + b_{l_1k_2}b_{k_1l_2} + b_{l_2k_2}b_{k_1l_1}, & \\
& \textit{(1-1)}\ k_1 \neq l_1, l_2; k_2 \neq l_1, l_2 \\
\frac{1}{p_{l_1}}\left(a_{k_2k_2}b_{l_2k_1} + a_{k_2k_2}b_{k_1l_2}\right) + b_{l_1k_1}b_{k_2l_2} + b_{l_2k_1}b_{k_2l_1} + b_{l_1k_2}b_{k_1l_2} + b_{l_2k_2}b_{k_1l_1}, & \\
& \textit{(1-2)}\ k_1 \neq l_1, l_2; k_2 = l_1 \\
\frac{1}{p_{l_2}}\left(a_{k_2k_2}b_{l_1k_1} + a_{k_2k_2}b_{k_1l_1}\right) + b_{l_1k_1}b_{k_2l_2} + b_{l_2k_1}b_{k_2l_1} + b_{l_1k_2}b_{k_1l_2} + b_{l_2k_2}b_{k_1l_1}, & \\
& \textit{(1-3)}\ k_1 \neq l_1, l_2; k_2 = l_2 \\
\frac{1}{p_{l_1}}\left(a_{k_1k_1}b_{l_2k_2} + a_{k_1k_1}b_{k_2l_2}\right) + b_{l_1k_2}b_{k_1l_2} + b_{l_2k_2}b_{k_1l_1} + b_{l_1k_1}b_{k_2l_2} + b_{l_2k_1}b_{k_2l_1}, & \\
& \textit{(2-1)}\ k_1 = l_1; k_2 \neq l_1, l_2; \\
& \textit{switch}\ k_1, k_2\ \textit{in (1-2)} \\
\left(b_{l_1l_2}^2 + b_{l_2l_1}^2\right) + \frac{2}{p_{l_1}p_{l_2}}\left(a_{l_1l_1} + p_{l_1}b_{l_1l_1}\right)\left(a_{l_2l_2} + p_{l_2}b_{l_2l_2}\right), & \\
& \textit{(2-2)}\ k_1 = l_1; k_2 = l_2; \\
& i.e., \operatorname{var}\left(\widetilde{\beta}_{l_1,l_2}\right) \\
\frac{1}{p_{l_2}}\left(a_{k_1k_1}b_{l_1k_2} + a_{k_1k_1}b_{k_2l_1}\right) + b_{l_1k_2}b_{k_1l_2} + b_{l_2k_2}b_{k_1l_1} + b_{l_1k_1}b_{k_2l_2} + b_{l_2k_1}b_{k_2l_1}, & \\
& \textit{(3-1)}\ k_1 = l_2; k_2 \neq l_1, l_2; \\
& \textit{switch}\ k_1, k_2\ \textit{in (1-3)} \\
\left(b_{l_2l_1}^2 + b_{l_1l_2}^2\right) + \frac{2}{p_{l_2}p_{l_1}}\left(a_{l_2l_2} + p_{l_2}b_{l_2l_2}\right)\left(a_{l_1l_1} + p_{l_1}b_{l_1l_1}\right), & \\
& \textit{(3-2)}\ k_1 = l_2; k_2 = l_1; \\
& i.e., \operatorname{var}\left(\widetilde{\beta}_{l_2,l_1}\right).
\end{cases}$$

for every $k_1 \neq k_2$ and $l_1 \neq l_2$.

(2) Furthermore, using the below transformations

$$\begin{pmatrix} \widetilde{a}_{kk} \\ \widetilde{b}_{kk} \end{pmatrix} = \frac{1}{p_k - 1} \begin{pmatrix} p_k & -p_k \\ -1 & p_k \end{pmatrix} \begin{pmatrix} \widetilde{\alpha}_{kk} \\ \widetilde{\beta}_{kk} \end{pmatrix} \quad \textit{for every } k, \quad \widetilde{b}_{kk'} = \widetilde{\beta}_{kk'} \quad \textit{for every } k \neq k',$$

the q by q covariance matrix of $\widetilde{\boldsymbol{\theta}}$ consisting the variance and covariance estimators for $\widetilde{\boldsymbol{\theta}}$, can be obtained by $\mathrm{var}(\widetilde{\boldsymbol{\theta}}) = \boldsymbol{\Phi}_{\boldsymbol{p}} \times \mathrm{var}\{(\widetilde{\alpha}_{11}, \ldots, \widetilde{\alpha}_{KK}, \widetilde{\beta}_{11}, \ldots, \widetilde{\beta}_{1K}, \widetilde{\beta}_{22}, \widetilde{\beta}_{KK})^\top\} \times \boldsymbol{\Phi}_{\boldsymbol{p}}^\top$, where $\boldsymbol{\Phi}_{\boldsymbol{p}} \in \mathbb{R}^{q \times q}$ is a matrix containing the elements of $\boldsymbol{p}$ only. In particular, by rearranging the order of the elements of $\widetilde{\boldsymbol{\theta}}$, we can obtain the results in Corollary 3.2.3.

B.3 Technical Proofs

B.3.1 Proof of Lemma 3.2.1

Proof of Lemma 3.2.1. See the proofs in Chapter 2. ■

B.3.2 Proof of Corollary 3.2.1

Proof of Corollary 3.2.1. The result can be obtained immediately from the properties of UB matrices in Chapter 2. ■

B.3.3 Derivations of the maximum likelihood estimator and its property

Starting from

$$\frac{\partial}{\partial \boldsymbol{\theta}_j}\ell_n(\boldsymbol{\theta};\mathbf{X}) = \frac{n}{2}\,\mathrm{tr}\left[\{\boldsymbol{\Sigma}(\mathbf{A},\mathbf{B},\boldsymbol{p}) - (\mathbf{S}^*_{kk'})\} \times \left\{\frac{\partial \mathbf{A}_{\boldsymbol{\Theta}}}{\partial \boldsymbol{\theta}_j} \circ \mathbf{I}(\boldsymbol{p}) + \frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial \boldsymbol{\theta}_j} \circ \mathbf{J}(\boldsymbol{p})\right\}\right],$$

let $(\mathbf{M}_{kk'})$ denote $\mathbf{M} = \boldsymbol{\Sigma}(\mathbf{A},\mathbf{B},\boldsymbol{p}) - (\mathbf{S}^*_{kk'})$ partitioned by $\boldsymbol{p}$. Then the system of the score equations is

$$\mathbf{S}_n(\boldsymbol{\theta};\mathbf{X}) = \frac{n}{2}\begin{pmatrix} \mathrm{tr}\left\{(\mathbf{M}_{kk'}) \times \frac{\partial \boldsymbol{\Theta}(\mathbf{A}_{\boldsymbol{\Theta}},\mathbf{B}_{\boldsymbol{\Theta}},\boldsymbol{p})}{\partial a_{11}}\right\} \\ \mathrm{tr}\left\{(\mathbf{M}_{kk'}) \times \frac{\partial \boldsymbol{\Theta}(\mathbf{A}_{\boldsymbol{\Theta}},\mathbf{B}_{\boldsymbol{\Theta}},\boldsymbol{p})}{\partial a_{22}}\right\} \\ \vdots \\ \mathrm{tr}\left\{(\mathbf{M}_{kk'}) \times \frac{\partial \boldsymbol{\Theta}(\mathbf{A}_{\boldsymbol{\Theta}},\mathbf{B}_{\boldsymbol{\Theta}},\boldsymbol{p})}{\partial a_{KK}}\right\} \\ \mathrm{tr}\left\{(\mathbf{M}_{kk'}) \times \frac{\partial \boldsymbol{\Theta}(\mathbf{A}_{\boldsymbol{\Theta}},\mathbf{B}_{\boldsymbol{\Theta}},\boldsymbol{p})}{\partial b_{11}}\right\} \\ \vdots \\ \mathrm{tr}\left\{(\mathbf{M}_{kk'}) \times \frac{\partial \boldsymbol{\Theta}(\mathbf{A}_{\boldsymbol{\Theta}},\mathbf{B}_{\boldsymbol{\Theta}},\boldsymbol{p})}{\partial b_{KK}}\right\} \end{pmatrix} = \mathbf{0}_{q\times 1}. \tag{B.3.1}$$

Recall the notations $\mathbf{P} = \mathrm{diag}(p_1,\ldots,p_K)$ and $\boldsymbol{\Delta} = \mathbf{A} + \mathbf{B}\times\mathbf{P}$. Using a fact that $(\mathbf{A}\times\mathbf{P})^{-1} - \boldsymbol{\Delta}\times\mathbf{B}\times\mathbf{A}^{-1} = (\mathbf{P}\times\boldsymbol{\Delta})^{-1}$, and doing some algebra, we obtain the following derivatives,

$$\frac{\partial \mathbf{A}_{\boldsymbol{\Theta}}}{\partial a_{kk}} = -a_{kk}^{-2}\mathbf{E}_{kk}, \quad \frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial a_{kk}} = \boldsymbol{\Delta}^{-1}\mathbf{E}_{kk}\boldsymbol{\Delta}^{-1}\mathbf{B}\mathbf{A}^{-1} + a_{kk}^{-2}\boldsymbol{\Delta}^{-1}\mathbf{B}\mathbf{E}_{kk}, \quad \text{for every } k,$$

$$\frac{\partial \mathbf{A}_{\boldsymbol{\Theta}}}{\partial b_{kk'}} = \mathbf{0}_{K\times K}, \quad \frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}} = \begin{cases} -\boldsymbol{\Delta}^{-1}\mathbf{E}_{kk}\mathbf{P}\boldsymbol{\Delta}^{-1}\mathbf{P}^{-1}, & k = k' \\ -\boldsymbol{\Delta}^{-1}(\mathbf{E}_{kk'} + \mathbf{E}_{k'k})\mathbf{P}\boldsymbol{\Delta}^{-1}\mathbf{P}^{-1}, & k \neq k' \end{cases}, \quad \text{for every } k \text{ and } k'.$$

Then, the individual equations (B.3.1) can be simplified as

$$\operatorname{tr}\left\{(\mathbf{M}_{kk'}) \times \frac{\partial \boldsymbol{\Theta}(\mathbf{A}_{\boldsymbol{\Theta}}, \mathbf{B}_{\boldsymbol{\Theta}}, \boldsymbol{p})}{\partial a_{kk}}\right\} = \left(-a_{kk}^{-2}\right) \operatorname{tr}\left(\mathbf{M}_{kk}\right) + \sum_{\ell=1}^{K}\sum_{\ell'=1}^{K} \operatorname{sum}(\mathbf{M}_{\ell\ell'}) \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial a_{kk}}\right)_{\ell',\ell} = 0,$$

$$\operatorname{tr}\left\{(\mathbf{M}_{kk'}) \times \frac{\partial \boldsymbol{\Theta}(\mathbf{A}_{\boldsymbol{\Theta}}, \mathbf{B}_{\boldsymbol{\Theta}}, \boldsymbol{p})}{\partial b_{kk'}}\right\} = \sum_{\ell=1}^{K}\sum_{\ell'=1}^{K} \operatorname{sum}(\mathbf{M}_{\ell\ell'}) \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}}\right)_{\ell',\ell} = 0,$$

for every k and k', where $(\partial \mathbf{B}_{\boldsymbol{\Theta}}/\partial a_{kk}), (\partial \mathbf{B}_{\boldsymbol{\Theta}}/\partial b_{kk'}) \in \mathbb{R}^{K\times K}$ and the subscript (ℓ', ℓ) denotes the (ℓ', ℓ) element of $(\partial \mathbf{B}_{\boldsymbol{\Theta}}/\partial a_{kk})$ or $(\partial \mathbf{B}_{\boldsymbol{\Theta}}/\partial b_{kk'})$.

Now, we claim the system of score equations (B.3.1) has a unique solution. Since $\mathbf{A} = \operatorname{diag}(a_{11}, \ldots, a_{KK})$, then $-\mathbf{A}^{-2} = \operatorname{diag}\left(-a_{11}^{-2}, \ldots, -a_{KK}^{-2}\right)$, denoted by $\mathbf{A}^{\star}$. Let $\boldsymbol{\beta}^{\star} \in \mathbb{R}^{q\times 1}$ denote a vector as below,

$$\boldsymbol{\beta}^{\star} = \{\operatorname{tr}(\mathbf{M}_{11}), \ldots, \operatorname{tr}(\mathbf{M}_{KK}), \operatorname{sum}(\mathbf{M}_{11}), \ldots, \operatorname{sum}(\mathbf{M}_{1K}), \operatorname{sum}(\mathbf{M}_{22}), \ldots, \operatorname{sum}(\mathbf{M}_{KK})\}^{\top}.$$

Let $\mathbf{B}^{(1),\star} \in \mathbb{R}^{K\times(q-K)}$ denote a matrix with k th row

$$\left\{\left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial a_{kk}}\right)_{1,1}, \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial a_{kk}}\right)_{1,2} + \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial a_{kk}}\right)_{2,1}, \ldots, \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial a_{kk}}\right)_{1,K} + \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial a_{kk}}\right)_{K,1}, \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial a_{kk}}\right)_{2,2}, \ldots, \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial a_{kk}}\right)_{K-1,K} + \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial a_{kk}}\right)_{K,K-1}, \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial a_{kk}}\right)_{K,K}\right\},$$

for $k = 1, \ldots, K$, and let $\mathbf{B}^{(2),\star} \in \mathbb{R}^{(q-K)\times(q-K)}$ denote a matrix with rows

$$\left\{\left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}}\right)_{1,1}, \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}}\right)_{1,2} + \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}}\right)_{2,1}, \ldots, \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}}\right)_{1,K} + \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}}\right)_{K,1}, \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}}\right)_{2,2}, \ldots, \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}}\right)_{K-1,K} + \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}}\right)_{K,K-1}, \left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}}\right)_{K,K}\right\}$$

for every $k \leq k'$. Thus, the system of the score equations (B.3.1) can be rewritten as

$$\mathbf{S}_n\left(\boldsymbol{\theta}; \mathbf{X}\right) = \begin{pmatrix} \mathbf{A}^\star & \mathbf{B}^{(1),\star} \\ \mathbf{0}_{(q-K)\times K} & \mathbf{B}^{(2),\star} \end{pmatrix} \boldsymbol{\beta}^\star = \mathbf{0}_{q\times 1},$$

and $\operatorname{rank}\begin{pmatrix} \mathbf{A}^\star & \mathbf{B}^{(1),\star} \\ \mathbf{0}_{(q-K)\times K} & \mathbf{B}^{(2),\star} \end{pmatrix} = \operatorname{rank}(\mathbf{A}^\star) + \operatorname{rank}(\mathbf{B}^{(2),\star})$ due to $\mathbf{I}_K - \mathbf{A}^\star(\mathbf{A}^\star)^{-1} = \mathbf{0}_{K\times K}$, which satisfies the condition provided in Buaphim et al. (2018, Theorem 3.10, p.334).

On the one hand, $\operatorname{rank}(\mathbf{A}^\star) = K$ by the positive definiteness that $a_{kk} > 0$ for every k. On the other hand, to compute $\operatorname{rank}(\mathbf{B}^{(2),\star})$, use the fact $\operatorname{vech}(AXB) = (B^\top \otimes A) \times \operatorname{vech}(X)$, where the matrices A, B and X with suitable sizes, and $\otimes$ denotes the Kronecker product. Since

$$\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}} = \begin{cases} -\boldsymbol{\Delta}^{-1}\mathbf{E}_{kk}\mathbf{P}\boldsymbol{\Delta}^{-1}\mathbf{P}^{-1}, & k = k' \\ -\boldsymbol{\Delta}^{-1}(\mathbf{E}_{kk'} + \mathbf{E}_{k'k})\mathbf{P}\boldsymbol{\Delta}^{-1}\mathbf{P}^{-1}, & k \neq k' \end{cases},$$

for every k and k', where $\mathbf{E}_{kk}$ and $\mathbf{E}_{kk'} + \mathbf{E}_{k'k}$ span the entire matrix space of symmetric matrices with the size of K by K. Thus, $\operatorname{rank}(\mathbf{B}^{(2),\star}) = (K+1)K/2$ due to the equivalence of the matrix space and the vector space spanned by the matrices. Therefore, the coefficients matrix of the given homogeneous system of linear equations has full rank and has a unique solution $\boldsymbol{\beta}^\star = \mathbf{0}_{q\times 1}$. Finally, the solution to equation (B.3.1), denoted by $\widetilde{\boldsymbol{\theta}}^*$, must be the maximum likelihood estimator due to the uniqueness.

Following equation (B.3.1), the first-order partial derivative of the score function with re-

spect to $\boldsymbol{\theta}$ is

$$\frac{\partial \mathbf{S}_n\left(\boldsymbol{\theta};\mathbf{X}\right)}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \dfrac{\partial^2 \ell_n(\boldsymbol{\theta};\mathbf{X})}{\partial \boldsymbol{\theta}^{(\mathrm{A})} \partial \boldsymbol{\theta}^{(\mathrm{A}),\top}} & \dfrac{\partial^2 \ell_n(\boldsymbol{\theta};\mathbf{X})}{\partial \boldsymbol{\theta}^{(\mathrm{A})} \partial \boldsymbol{\theta}^{(\mathrm{B}),\top}} \\ \dfrac{\partial^2 \ell_n(\boldsymbol{\theta};\mathbf{X})}{\partial \boldsymbol{\theta}^{(\mathrm{B})} \partial \boldsymbol{\theta}^{(\mathrm{A}),\top}} & \dfrac{\partial^2 \ell_n(\boldsymbol{\theta};\mathbf{X})}{\partial \boldsymbol{\theta}^{(\mathrm{B})} \partial \boldsymbol{\theta}^{(\mathrm{B}),\top}} \end{pmatrix} = \frac{n}{2} \begin{pmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{H}_2^\top & \mathbf{H}_3 \end{pmatrix}, \tag{B.3.2}$$

where $\boldsymbol{\theta}^{(\mathrm{A})} = (a_{11}, \ldots, a_{KK})^\top$, $\boldsymbol{\theta}^{(\mathrm{B})} = (b_{11}, \ldots, b_{1K}, b_{22}, \ldots, b_{KK})^\top$, and therefore $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(\mathrm{A}),\top}, \boldsymbol{\theta}^{(\mathrm{B}),\top})^\top$, the blocks $\mathbf{H}_1 \in \mathbb{R}^{K\times K}$, $\mathbf{H}_2 \in \mathbb{R}^{K\times(q-K)}$, and $\mathbf{H}_3 \in \mathbb{R}^{(q-K)\times(q-K)}$. In particular, given k and k',

$$\mathbf{H}_1 = \begin{cases} 2a_{kk}^{-3}\operatorname{tr}(\mathbf{M}_{kk}) - a_{kk}^{-2}\left\{\dfrac{\partial}{\partial a_{mm}}\operatorname{tr}(\mathbf{M}_{kk})\right\} \\ + \sum_{\ell=1}^K \sum_{\ell'=1}^K \left[\left\{\dfrac{\partial}{\partial a_{mm}}\operatorname{sum}(\mathbf{M}_{\ell\ell'})\right\}\left(\dfrac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial a_{kk}}\right)_{\ell',\ell} + \operatorname{sum}(\mathbf{M}_{\ell\ell'})\left\{\dfrac{\partial}{\partial a_{mm}}\left(\dfrac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial a_{kk}}\right)\right\}_{\ell',\ell}\right], & m = k \\ \sum_{\ell=1}^K \sum_{\ell'=1}^K \left[\left\{\dfrac{\partial}{\partial a_{mm}}\operatorname{sum}(\mathbf{M}_{\ell\ell'})\right\}\left(\dfrac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial a_{kk}}\right)_{\ell',\ell} + \operatorname{sum}(\mathbf{M}_{\ell\ell'})\left\{\dfrac{\partial}{\partial a_{mm}}\left(\dfrac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial a_{kk}}\right)\right\}_{\ell',\ell}\right], & m \neq k \end{cases}$$

for $m = 1, 2, \ldots, K$; and

$$\mathbf{H}_2 = \sum_{\ell=1}^K \sum_{\ell'=1}^K \left[\left\{\frac{\partial}{\partial a_{mm}}\operatorname{sum}(\mathbf{M}_{\ell\ell'})\right\}\left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}}\right)_{\ell',\ell} + \operatorname{sum}(\mathbf{M}_{\ell\ell'})\left\{\frac{\partial}{\partial a_{mm}}\left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}}\right)\right\}_{\ell',\ell}\right],$$
$$\mathbf{H}_3 = \sum_{\ell=1}^K \sum_{\ell'=1}^K \left[\left\{\frac{\partial}{\partial b_{mm'}}\operatorname{sum}(\mathbf{M}_{\ell\ell'})\right\}\left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}}\right)_{\ell',\ell} + \operatorname{sum}(\mathbf{M}_{\ell\ell'})\left\{\frac{\partial}{\partial b_{mm'}}\left(\frac{\partial \mathbf{B}_{\boldsymbol{\Theta}}}{\partial b_{kk'}}\right)\right\}_{\ell',\ell}\right],$$

for $m, m' = 1, 2, \ldots, K$. The first-order partial derivatives $\partial \operatorname{sum}(\mathbf{M}_{\ell\ell'})/\partial a_{mm}$ and $\partial \operatorname{sum}(\mathbf{M}_{\ell\ell'})/\partial b_{mm'}$ are easily computed, where $\mathbf{M}_{\ell\ell'} = \boldsymbol{\Sigma}_{\ell\ell'} - \mathbf{S}^*_{\ell\ell'}$ for $m, m', \ell, \ell' = 1, 2, \ldots, K$. The second-order partial derivatives $\partial(\partial \mathbf{B}_{\boldsymbol{\Theta}}/\partial a_{kk})/\partial a_{mm}$, $\partial(\partial \mathbf{B}_{\boldsymbol{\Theta}}/\partial b_{kk'})/\partial a_{mm}$, and $\partial(\partial \mathbf{B}_{\boldsymbol{\Theta}}/\partial b_{kk'})/\partial b_{mm'}$

are matrices for $m, m', k, k' = 1, 2, \ldots, K$, shown respectively as below.

$$\frac{\partial}{\partial a_{mm}}\left(\frac{\partial \mathbf{B}_{\mathbf{\Theta}}}{\partial a_{kk}}\right) = \begin{cases} -\mathbf{\Delta}^{-1}\mathbf{E}_{kk}\mathbf{\Delta}^{-1}\left(2\mathbf{E}_{kk}\mathbf{\Delta}^{-1}\mathbf{B}\mathbf{A}^{-1} + \mathbf{B}\mathbf{A}^{-1}\mathbf{E}_{kk}\mathbf{A}^{-1}\right) \\ \quad -\left(2a_{kk}^{-3}\mathbf{I}_K + a_{kk}^{-2}\mathbf{\Delta}^{-1}\mathbf{E}_{kk}\right)\mathbf{\Delta}^{-1}\mathbf{B}\mathbf{E}_{kk}, & k = m \\ -\mathbf{\Delta}^{-1}\left(\mathbf{E}_{mm}\mathbf{\Delta}^{-1}\mathbf{E}_{kk} + \mathbf{E}_{kk}\mathbf{\Delta}^{-1}\mathbf{E}_{mm}\right)\mathbf{\Delta}^{-1}\mathbf{B}\mathbf{A}^{-1} \\ \quad -\mathbf{\Delta}^{-1}\mathbf{E}_{kk}\mathbf{\Delta}^{-1}\mathbf{B}\mathbf{A}^{-1}\mathbf{E}_{mm}\mathbf{A}^{-1} - a_{kk}^{-2}\mathbf{\Delta}^{-1}\mathbf{E}_{mm}\mathbf{\Delta}^{-1}\mathbf{B}\mathbf{E}_{kk}, & k \neq m \end{cases},$$

and,

$$\frac{\partial}{\partial a_{mm}}\left(\frac{\partial \mathbf{B}_{\mathbf{\Theta}}}{\partial b_{kk'}}\right) = \begin{cases} \mathbf{\Delta}^{-1}\mathbf{E}_{mm}\mathbf{\Delta}^{-1}\mathbf{E}_{kk}\mathbf{P}\mathbf{\Delta}^{-1}\mathbf{P} \\ \quad -\mathbf{\Delta}^{-1}\mathbf{E}_{kk}\mathbf{P}\mathbf{\Delta}^{-1}(\mathbf{E}_{mm}\mathbf{\Delta}^{-1}\mathbf{B} - \mathbf{P}\mathbf{E}_{mm})\mathbf{A}^{-1}, & k = k' \\ \mathbf{\Delta}^{-1}\mathbf{E}_{mm}\mathbf{\Delta}^{-1}\left(\mathbf{E}_{kk'} + \mathbf{E}_{k'k}\right)\mathbf{P}\mathbf{\Delta}^{-1}\mathbf{P} \\ \quad -\mathbf{\Delta}^{-1}\left(\mathbf{E}_{kk'} + \mathbf{E}_{k'k}\right)\mathbf{P}\mathbf{\Delta}^{-1}(\mathbf{E}_{mm}\mathbf{\Delta}^{-1}\mathbf{B} - \mathbf{P}\mathbf{E}_{mm})\mathbf{A}^{-1}, & k \neq k' \end{cases}$$

and,

$$\frac{\partial}{\partial b_{mm'}}\left(\frac{\partial \mathbf{B}_{\mathbf{\Theta}}}{\partial b_{kk'}}\right)$$
$$= \begin{cases} \begin{cases} \mathbf{\Delta}^{-1}(\mathbf{E}_{mm} + \mathbf{E}_{kk})\mathbf{P}\mathbf{\Delta}^{-1}(\mathbf{E}_{kk} + \mathbf{E}_{mm})\mathbf{P}\mathbf{\Delta}^{-1}\mathbf{P}, & m = m' \\ \mathbf{\Delta}^{-1}(\mathbf{E}_{mm'} + \mathbf{E}_{m'm} + \mathbf{E}_{kk})\mathbf{P}\mathbf{\Delta}^{-1}(\mathbf{E}_{kk} + \mathbf{E}_{mm'} + \mathbf{E}_{m'm})\mathbf{P}\mathbf{\Delta}^{-1}\mathbf{P}, & m \neq m' \end{cases}, & k = k' \\ \begin{cases} \mathbf{\Delta}^{-1}(\mathbf{E}_{mm} + \mathbf{E}_{kk'} + \mathbf{E}_{k'k})\mathbf{P}\mathbf{\Delta}^{-1}(\mathbf{E}_{kk'} + \mathbf{E}_{k'k} + \mathbf{E}_{mm})\mathbf{P}\mathbf{\Delta}^{-1}\mathbf{P}, & m = m' \\ \mathbf{\Delta}^{-1}(\mathbf{E}_{mm'} + \mathbf{E}_{m'm} + \mathbf{E}_{kk'} + \mathbf{E}_{k'k})\mathbf{P}\mathbf{\Delta}^{-1}(\mathbf{E}_{kk'} + \mathbf{E}_{k'k} + \mathbf{E}_{mm'} + \mathbf{E}_{m'm})\mathbf{P}\mathbf{\Delta}^{-1}\mathbf{P}, & m \neq m' \end{cases}, & k \neq k' \end{cases}.$$

By the unbiasedness from Theorem 3.2.1, the expectations of $\mathbf{H}_1$, $\mathbf{H}_2$, and $\mathbf{H}_3$ depend

on the terms $\mathrm{E}(\mathrm{sum}(\mathbf{M}_{\ell\ell})) = (1-c_n)p_\ell(a_{\ell\ell}+b_{\ell\ell}p_\ell)$, $\mathrm{E}(\mathrm{sum}(\mathbf{M}_{\ell\ell'})) = (1-c_n)p_\ell p_{\ell'}b_{\ell\ell'}$ and $\mathrm{E}(\mathrm{tr}(\mathbf{M}_{\ell\ell})) = (1-c_n)p_\ell(a_\ell+b_\ell)$ only, where $c_n = (n-1)/n$. Then Fisher's information matrices for n observations and one observation can be calculated by $\mathbb{I}_n(\boldsymbol{\theta}) = -\mathrm{E}(\partial \mathbf{S}_n(\boldsymbol{\theta};\mathbf{X})/\partial\boldsymbol{\theta})$ and $\mathbb{I}_1(\boldsymbol{\theta}) = \mathbb{I}_n(\boldsymbol{\theta})/n$, respectively. Since both a determinant operator and a trace operator are continuous, the log-likelihood function $\ell_n(\boldsymbol{\theta};\mathbf{X})$ is continuous with respect to $\boldsymbol{\theta}$. In addition, there is a unique solution to the likelihood equation for every n. Thus, $\widetilde{\boldsymbol{\theta}}^*$ is strongly consistent, asymptotically normal, and asymptotically efficient, following the classical arguments in Ferguson (1996); van der Vaart and Wellner (1996); Stuart. et al. (1999); Bickel and Doksum (2015a,b), i.e., $\widetilde{\boldsymbol{\theta}}^* \to \boldsymbol{\theta}$ almost surely as $n \to \infty$ and $\sqrt{n}(\widetilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}) \to N(\mathbf{0}_{q\times 1}, \mathbb{I}_1^{-1}(\boldsymbol{\theta}))$ in distribution as $n \to \infty$.

B.3.4 Proof of Theorem 3.2.1

Proof of Theorem 3.2.1. Recall an i.i.d. random sample $\boldsymbol{X}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}(\mathbf{A},\mathbf{B},\boldsymbol{p}))$ for $i = 1,\ldots,n$, where $\boldsymbol{\Sigma}(\mathbf{A},\mathbf{B},\boldsymbol{p})$ is supposed to be a UB matrix with diagonal matrix $\mathbf{A}$ and symmetric matrix $\mathbf{B}$, and $\boldsymbol{p} = (p_1,\ldots,p_K)^\top$ is a given partition-size vector with $p_k > 1$ for every k and $p = p_1 + \cdots + p_K$. Let $\boldsymbol{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}(\mathbf{A},\mathbf{B},\boldsymbol{p}))$ and $\bar{\boldsymbol{X}} = (\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)/n$. Let $\boldsymbol{X}_{i,p_k}, \bar{\boldsymbol{X}}_{p_k}, \boldsymbol{\mu}_{p_k} \in \mathbb{R}^{p_k}$ satisfying $\boldsymbol{X}_i = (\boldsymbol{X}_{i,p_1}^\top,\ldots,\boldsymbol{X}_{i,p_K}^\top)^\top$, $\bar{\boldsymbol{X}} = (\bar{\boldsymbol{X}}_{p_1}^\top,\ldots,\bar{\boldsymbol{X}}_{p_K}^\top)^\top$, and $\boldsymbol{\mu} = (\boldsymbol{\mu}_{p_1}^\top,\ldots,\boldsymbol{\mu}_{p_K}^\top)^\top$, respectively.

First, we prove that $\widetilde{\boldsymbol{\theta}}$ is unbiased. By $\boldsymbol{\Sigma}(\mathbf{A},\mathbf{B},\boldsymbol{p}) = (\boldsymbol{\Sigma}_{kk'})$, we can obtain that

$$\boldsymbol{\Sigma}_{kk'} = \mathrm{cov}(\boldsymbol{X}_{p_k}, \boldsymbol{X}_{p_{k'}}) = \mathrm{E}\{(\boldsymbol{X}_{p_k} - \boldsymbol{\mu}_{p_k})(\boldsymbol{X}_{p_{k'}} - \boldsymbol{\mu}_{p_{k'}})^\top\},$$

$$\boldsymbol{\Sigma}_{kk'}/n = \mathrm{cov}(\bar{\boldsymbol{X}}_{p_k}, \bar{\boldsymbol{X}}_{p_{k'}}) = \mathrm{E}\{(\bar{\boldsymbol{X}}_{p_k} - \boldsymbol{\mu}_{p_k})(\bar{\boldsymbol{X}}_{p_{k'}} - \boldsymbol{\mu}_{p_{k'}})^\top\}.$$

for every k and k'. Let $\mathbf{C}_{kk'} = \sum_{i=1}^{n}(\boldsymbol{X}_{i,p_k} - \bar{\boldsymbol{X}}_{i,p_k})(\boldsymbol{X}_{i,p_{k'}} - \bar{\boldsymbol{X}}_{i,p_{k'}})^\top$ for every k and k', which can be simplified to

$$\mathbf{C}_{kk'} = \sum_{i=1}^{n}(\boldsymbol{X}_{i,p_k} - \boldsymbol{\mu}_{p_k})(\boldsymbol{X}_{i,p_{k'}} - \boldsymbol{\mu}_{p_{k'}})^\top - n(\bar{\boldsymbol{X}}_{p_k} - \boldsymbol{\mu}_{p_k})(\bar{\boldsymbol{X}}_{p_{k'}} - \boldsymbol{\mu}_{p_{k'}})^\top.$$

Taking expectation on the both sides, we have $\mathrm{E}\left(\mathbf{C}_{kk'}\right) = (n-1)\boldsymbol{\Sigma}_{kk'}$ for every k and k'. Recall that, given k and k', $\boldsymbol{\Sigma}_{kk} = a_{kk}\mathbf{I}_{p_k} + b_{kk}\mathbf{J}_{p_k}$ for $k = k'$ and $\boldsymbol{\Sigma}_{kk'} = b_{kk'}\mathbf{1}_{p_k \times p_{k'}}$. Define

$$\alpha_{kk} = a_{kk} + b_{kk} = \mathrm{tr}\left(\boldsymbol{\Sigma}_{kk'}\right)/p_k, \quad \beta_{kk'} = \begin{cases} a_{kk}/p_k + b_{kk} = \mathrm{sum}\left(\boldsymbol{\Sigma}_{kk}\right)/p_k^2, & k = k' \\ b_{kk'} = \mathrm{sum}(\boldsymbol{\Sigma}_{kk'})/(p_k p_{k'}), & k \neq k' \end{cases}.$$

By $(n-1)\mathbf{S}_{kk'} = \mathbf{C}_{kk'}$, we have the following results:

$$\widetilde{\alpha}_{kk} = p_k^{-1}\,\mathrm{tr}(\mathbf{S}_{kk}), \quad \widetilde{\beta}_{kk} = p_k^{-2}\,\mathrm{sum}(\mathbf{S}_{kk}), \quad \widetilde{\beta}_{kk'} = (p_k p_{k'})^{-1}\,\mathrm{sum}(\mathbf{S}_{kk'})$$

are unbiased estimators for α_{kk}, β_{kk} for every k and for $\beta_{kk'}$ for every $k \neq k'$. Furthermore,

$$\widetilde{a}_{kk} = \widetilde{\alpha}_{kk} - \frac{p_k\widetilde{\beta}_{kk} - \widetilde{\alpha}_{kk}}{p_k - 1}, \quad \widetilde{b}_{kk'} = \begin{cases} \dfrac{p_k\widetilde{\beta}_{kk} - \widetilde{\alpha}_{kk}}{p_k - 1}, & k = k' \\ \widetilde{\beta}_{kk'}, & k \neq k' \end{cases},$$

are unbiased estimators for a_{kk} and $b_{kk'}$ every k and k'.

Next, we prove the optimal property of $\widetilde{\boldsymbol{\theta}}$. Let $\boldsymbol{Y} = (\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_n^\top)^\top \in \mathbb{R}^{pn}$ denote a vector consisting of all observations. By normality assumption, $\boldsymbol{Y} \sim N(\boldsymbol{\mu}_y, \mathbf{V})$, where $\boldsymbol{\mu}_y = (\mathbf{1}_n \otimes \mathbf{I}_p) \times \boldsymbol{\mu} \in \mathbb{R}^{pn}$ and $\mathbf{V} = \mathbf{I}_n \otimes \boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p}) \in \mathbb{R}^{pn \times pn}$.

First, we prove $\widetilde{\boldsymbol{\mu}} = \bar{\boldsymbol{X}}$ is the best linear unbiased estimator (BLUE). Equivalently, we need to check the conditions of Theorem 1 and Corollary 2 in Zmyślony (1976). Since $\boldsymbol{\Sigma}(\mathbf{I}_K, \mathbf{0}_{K\times K}, \boldsymbol{p}) = \mathbf{I}_p$, the identity matrix $\mathbf{I}_n \otimes \mathbf{I}_p \in \text{span}(\mathbf{V})$. Let M^- denote the generalized inverse for some matrix M, and let $\mathbf{P}_0 = (\mathbf{1}_n \otimes \mathbf{I}_p)\{(\mathbf{1}_n \otimes \mathbf{I}_p)^\top (\mathbf{1}_n \otimes \mathbf{I}_p)\}^-(\mathbf{1}_n \otimes \mathbf{I}_p)^\top$, which can be simplified to $n^{-1}\mathbf{J}_n \otimes \mathbf{I}_n$ by using the fact $\mathbf{1}_n^- = \mathbf{1}_n^\top/n$. Thus, $\mathbf{P}_0$ is an orthogonal projector because

$$\mathbf{P}_0^2 = (n^{-1}\mathbf{J}_n \otimes \mathbf{I}_p)(n^{-1}\mathbf{J}_n \otimes \mathbf{I}_p) = n^{-2}(\mathbf{J}_n \times \mathbf{J}_n) \otimes \mathbf{I}_p = n^{-1}\mathbf{J}_n \otimes \mathbf{I}_p = \mathbf{P}_0,$$

$$\mathbf{P}_0^\top = n^{-1}\mathbf{J}_n^\top \otimes \mathbf{I}_p^\top = \mathbf{P}_0.$$

Then, $\mathbf{V}$ and $\mathbf{P}_0$ are commutative, because

$$\mathbf{P}_0\mathbf{V} = (n^{-1}\mathbf{J}_n \otimes \mathbf{I}_p)\{\mathbf{I}_n \otimes \boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})\} = n^{-1}(\mathbf{J}_n\mathbf{I}_n) \otimes \{\mathbf{I}_p\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})\} = n^{-1}\mathbf{J}_n \otimes \boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p}),$$

$$\mathbf{V}\mathbf{P}_0 = \{\mathbf{I}_n \otimes \boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})\}(n^{-1}\mathbf{J}_n \otimes \mathbf{I}_p) = n^{-1}(\mathbf{I}_n\mathbf{J}_n) \otimes \{\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})\mathbf{I}_p\} = n^{-1}\mathbf{J}_n \otimes \boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p}).$$

Hence, all estimable functions have the best estimator, which can be expressed in terms of the solution to the following normal equation

$$(\mathbf{1}_n \otimes \mathbf{I}_p)^\top(\mathbf{1}_n \otimes \mathbf{I}_p)\boldsymbol{\mu} = (\mathbf{1}_n \otimes \mathbf{I}_p)^\top \boldsymbol{Y}.$$

So, $\widetilde{\boldsymbol{\mu}} = \bar{\boldsymbol{X}}$ is BLUE. This coincides with the arguments of the proofs in Roy et al. (2016) and Koziol et al. (2017).

Second, we can prove $\widetilde{\alpha}_{kk}$ and $\widetilde{\beta}_{kk'}$ are best quadratic unbiased estimator (BQUE) for α_{kk} and $\beta_{kk'}$ for every k and k'. We need to check the conditions of Theorem 2 in Zmyślony (1976), or the lines of analogous arguments in Roy et al. (2016) and Koziol et al. (2017), that is, given

$\mathbf{P}_0\mathbf{V} = \mathbf{V}\mathbf{P}_0$ and $\mathbf{R}_0 = \mathbf{I}_{pn} - \mathbf{P}_0$, there exist BQUE for the parameters of quadratic covariance if and only if $\text{span}\,(\mathbf{P}_0\mathbf{V}\mathbf{P}_0)$, i.e., the smallest linear space containing $\mathbf{P}_0\mathbf{V}\mathbf{P}_0$, is a quadratic space. Since $\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$ is a UB matrix, $\{\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})\}^2$ is a UB matrix, expressed by $(\boldsymbol{\Sigma}^2)\,(\mathbf{A}^2, \mathbf{AB} + \mathbf{BA} + \mathbf{BPB}, \boldsymbol{p})$, so $\text{span}\{\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})\}$ is a quadratic subspace and the identity matrix $\mathbf{I}(\boldsymbol{p})$ belongs to it. So, $\text{span}\,(\mathbf{V})$ is a quadratic subspace. In addition, it is clear that $\mathbf{P}_0$ is idempotent, because $\mathbf{P}_0^2 = (\mathbf{I}_{pn} - \mathbf{P}_0)^2 = \mathbf{I}_{pn}^2 - \mathbf{I}_{pn}\mathbf{P}_0 - \mathbf{P}_0\mathbf{I}_{pn} + \mathbf{P}_0^2 = \mathbf{I}_{pn} - 2\mathbf{P}_0 + \mathbf{P}_0 = \mathbf{I}_{pn} - \mathbf{P}_0 = \mathbf{P}_0$. By the result (2.e) from Seely (1971), $\text{span}(\mathbf{P}_0\mathbf{V}\mathbf{P}_0) = \{\mathbf{P}_0\mathbf{V}\mathbf{P}_0 : \mathbf{V} \in \text{span}(\mathbf{V})\}$ is a quadratic subspace. Since $\text{span}\{\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})\}$ and $\text{span}(\mathbf{V})$ are quadratic subspaces, we can find bases for them respectively. Recall the definition of the block Hadamard product. $\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{B}, \boldsymbol{p})$ has a base as follows:

$$\mathbf{E}_{kk} \circ \mathbf{I}(\boldsymbol{p}), \quad \mathbf{E}_{kk} \circ \mathbf{J}(\boldsymbol{p}), \quad \text{for every } k, \quad (\mathbf{E}_{kk'} + \mathbf{E}_{k'k}) \circ \mathbf{J}(\boldsymbol{p}), \quad \text{for every } k \neq k',$$

where $\mathbf{E}_{kk'}$ denotes a K by K matrix in which (k, k') entry is 1 and the other entries are 0's for every k and k'. Thus, the base for $\text{span}\,(\mathbf{V})$ is

$$\mathbf{I}_n \otimes \{\mathbf{E}_{kk} \circ \mathbf{I}(\boldsymbol{p})\}, \quad \mathbf{I}_n \otimes \{\mathbf{E}_{kk} \circ \mathbf{J}(\boldsymbol{p})\}, \quad \text{for every } k, \quad \mathbf{I}_n \otimes \{(\mathbf{E}_{kk'} + \mathbf{E}_{k'k}) \circ \mathbf{J}(\boldsymbol{p})\}, \quad \text{for every } k \neq k'.$$

By Result 2 from Roy et al. (2016); Koziol et al. (2017), $(\mathbf{1}_n^\top \otimes \mathbf{I}_p)\boldsymbol{Y}$ is the complete and minimal sufficient statistic for $\boldsymbol{\mu}$, and

$$\boldsymbol{Y}^\top \times \mathbf{P}_0 \times [\mathbf{I}_n \otimes \{\mathbf{E}_{kk} \circ \mathbf{I}(\boldsymbol{p})\}] \times \mathbf{P}_0 \times \boldsymbol{Y}, \quad \boldsymbol{Y}^\top \times \mathbf{P}_0 \times [\mathbf{I}_n \otimes \{\mathbf{E}_{kk} \circ \mathbf{J}(\boldsymbol{p})\}] \times \mathbf{P}_0 \times \boldsymbol{Y}, \quad \text{for every } k,$$

$$\boldsymbol{Y}^\top \times \mathbf{P}_0 \times [\mathbf{I}_n \otimes \{(\mathbf{E}_{kk'} + \mathbf{E}_{k'k}) \circ \mathbf{J}(\boldsymbol{p})\}] \times \mathbf{P}_0 \times \boldsymbol{Y}, \quad \text{for every } k \neq k',$$

are the complete and minimal sufficient statistics for $\boldsymbol{\Sigma}$. Given the above base for $\text{span}(\mathbf{V})$, we

follow the arguments about the coordinate-free approach (Roy et al., 2016; Koziol et al., 2017; Wichura, 2006), the BQUEs for α_{kk} and $\beta_{kk'}$ are the least square estimators satisfying the normal equations as below.

$$
\begin{aligned}
\operatorname{vec}\left(\mathbf{P}_0 \times \left[\mathbf{I}_n \otimes \left\{\mathbf{E}_{kk} \circ \mathbf{I}(\boldsymbol{p})\right\}\right]\right)^\top \times \operatorname{vec}\left(\mathbf{P}_0 \times \left[\mathbf{I}_n \otimes \left\{\mathbf{E}_{kk} \circ \mathbf{I}(\boldsymbol{p})\right\}\right]\right) \times \alpha_{kk} \\
= \operatorname{vec}\left(\mathbf{P}_0 \times \left[\mathbf{I}_n \otimes \left\{\mathbf{E}_{kk} \circ \mathbf{I}(\boldsymbol{p})\right\}\right]\right)^\top \\
\times \operatorname{vec}\left\{\left(\mathbf{P}_0 \times \boldsymbol{Y}\right) \times \left(\mathbf{P}_0 \times \boldsymbol{Y}\right)^\top\right\}, \\
\operatorname{vec}\left(\mathbf{P}_0 \times \left[\mathbf{I}_n \otimes \left\{\mathbf{E}_{kk} \circ \mathbf{J}(\boldsymbol{p})\right\}\right]\right)^\top \times \operatorname{vec}\left(\mathbf{P}_0 \times \left[\mathbf{I}_n \otimes \left\{\mathbf{E}_{kk} \circ \mathbf{J}(\boldsymbol{p})\right\}\right]\right) \times \beta_{kk} \\
= \operatorname{vec}\left(\mathbf{P}_0 \times \left[\mathbf{I}_n \otimes \left\{\mathbf{E}_{kk} \circ \mathbf{J}(\boldsymbol{p})\right\}\right]\right)^\top \\
\times \operatorname{vec}\left\{\left(\mathbf{P}_0 \times \boldsymbol{Y}\right) \times \left(\mathbf{P}_0 \times \boldsymbol{Y}\right)^\top\right\}, \\
\operatorname{vec}\left(\mathbf{P}_0 \times \left[\mathbf{I}_n \otimes \left\{\left(\mathbf{E}_{kk'} + \mathbf{E}_{k'k}\right) \circ \mathbf{J}(\boldsymbol{p})\right\}\right]\right)^\top \times \operatorname{vec}\left(\mathbf{P}_0 \times \left[\mathbf{I}_n \otimes \left\{\left(\mathbf{E}_{kk'} + \mathbf{E}_{k'k}\right) \circ \mathbf{J}(\boldsymbol{p})\right\}\right]\right) \times \beta_{kk'} \\
= \operatorname{vec}\left(\mathbf{P}_0 \times \left[\mathbf{I}_n \otimes \left\{\left(\mathbf{E}_{kk'} + \mathbf{E}_{k'k}\right) \circ \mathbf{J}(\boldsymbol{p})\right\}\right]\right)^\top \\
\times \operatorname{vec}\left\{\left(\mathbf{P}_0 \times \boldsymbol{Y}\right) \times \left(\mathbf{P}_0 \times \boldsymbol{Y}\right)^\top\right\},
\end{aligned}
$$

for every k, every k, and every $k \neq k'$, respectively, where $\operatorname{vec}(M)$ denotes a single vector by stacking the columns of M (Henderson and Searle, 1979). In addition to the fact that $\operatorname{vec}(A)^\top \operatorname{vec}(B) = \operatorname{tr}(A^\top B)$, the idempotent matrix $\mathbf{P}_0$ commutes with the base of $\operatorname{span}(\mathbf{V})$, the above equations can be simplified as below.

$$
\begin{aligned}
&\operatorname{tr}\left(\mathbf{P}_0 \left[\mathbf{I}_n \otimes \left\{\mathbf{E}_{kk} \circ \mathbf{I}(\boldsymbol{p})\right\}\right]^2\right) \alpha_{kk} = \left(\mathbf{P}_0 \boldsymbol{Y}\right)^\top \left[\mathbf{I}_n \otimes \left\{\mathbf{E}_{kk} \circ \mathbf{I}(\boldsymbol{p})\right\}\right] \left(\mathbf{P}_0 \boldsymbol{Y}\right), \quad \text{for every } k, \\
&\operatorname{tr}\left(\mathbf{P}_0 \left[\mathbf{I}_n \otimes \left\{\mathbf{E}_{kk} \circ \mathbf{J}(\boldsymbol{p})\right\}\right]^2\right) \beta_{kk} = \left(\mathbf{P}_0 \boldsymbol{Y}\right)^\top \left[\mathbf{I}_n \otimes \left\{\mathbf{E}_{kk} \circ \mathbf{J}(\boldsymbol{p})\right\}\right] \left(\mathbf{P}_0 \boldsymbol{Y}\right), \quad \text{for every } k, \\
&\operatorname{tr}\left(\mathbf{P}_0 \left[\mathbf{I}_n \otimes \left\{\left(\mathbf{E}_{kk'} + \mathbf{E}_{k'k}\right) \circ \mathbf{J}(\boldsymbol{p})\right\}\right]^2\right) \beta_{kk'} = \left(\mathbf{P}_0 \boldsymbol{Y}\right)^\top \left[\mathbf{I}_n \otimes \left\{\left(\mathbf{E}_{kk'} + \mathbf{E}_{k'k}\right) \circ \mathbf{J}(\boldsymbol{p})\right\}\right] \left(\mathbf{P}_0 \boldsymbol{Y}\right), \quad \text{for every } k \neq k'
\end{aligned}
$$

Define the residual vector $\boldsymbol{r}_0 = \mathbf{P}_0 \times \boldsymbol{Y} \in \mathbb{R}^{pn}$, the solutions to the simplified normal equations are

$$\widetilde{\alpha}_{kk} = \frac{\boldsymbol{r}_0^\top \left[\mathbf{I}_n \otimes \{\mathbf{E}_{kk} \circ \mathbf{I}(\boldsymbol{p})\}\right] \boldsymbol{r}_0}{\operatorname{tr}\left(\mathbf{P}_0 \left[\mathbf{I}_n \otimes \{\mathbf{E}_{kk} \circ \mathbf{I}(\boldsymbol{p})\}\right]^2\right)}, \quad \widetilde{\beta}_{kk} = \frac{\boldsymbol{r}_0^\top \left[\mathbf{I}_n \otimes \{\mathbf{E}_{kk} \circ \mathbf{J}(\boldsymbol{p})\}\right] \boldsymbol{r}_0}{\operatorname{tr}\left(\mathbf{P}_0 \left[\mathbf{I}_n \otimes \{\mathbf{E}_{kk} \circ \mathbf{J}(\boldsymbol{p})\}\right]^2\right)} \quad \text{for every } k,$$
$$\widetilde{\beta}_{kk'} = \frac{\boldsymbol{r}_0^\top \left[\mathbf{I}_n \otimes \{(\mathbf{E}_{kk'} + \mathbf{E}_{k'k}) \circ \mathbf{J}(\boldsymbol{p})\}\right] \boldsymbol{r}_0}{\operatorname{tr}\left(\mathbf{P}_0 \left[\mathbf{I}_n \otimes \{(\mathbf{E}_{kk'} + \mathbf{E}_{k'k}) \circ \mathbf{J}(\boldsymbol{p})\}\right]^2\right)} \quad \text{for every } k \neq k',$$

where the denominators can be further simplified respectively:

$$\operatorname{tr}\left(\mathbf{P}_0 \left[\mathbf{I}_n \otimes \{\mathbf{E}_{kk} \circ \mathbf{I}(\boldsymbol{p})\}\right]^2\right) = (n-1)\operatorname{tr}\{\mathbf{E}_{kk} \circ \mathbf{I}(\boldsymbol{p})\} = (n-1)p_k \quad \text{for every } k,$$

$$\operatorname{tr}\left(\mathbf{P}_0 \left[\mathbf{I}_n \otimes \{\mathbf{E}_{kk} \circ \mathbf{J}(\boldsymbol{p})\}\right]^2\right) = (n-1)\operatorname{tr}\{(\mathbf{E}_{kk} P \mathbf{E}_{kk}) \circ \mathbf{J}(\boldsymbol{p})\} = (n-1)p_k^2 \quad \text{for every } k,$$

$$\operatorname{tr}\left(\mathbf{P}_0 \left[\mathbf{I}_n \otimes \{(\mathbf{E}_{kk'} + \mathbf{E}_{k'k}) \circ \mathbf{J}(\boldsymbol{p})\}\right]^2\right) = (n-1)(2p_k p_{k'}) \quad \text{for every } k \neq k'.$$

Let $\boldsymbol{r}_i = (\boldsymbol{r}_{i,p_1}^\top, \boldsymbol{r}_{i,p_2}^\top, \ldots, \boldsymbol{r}_{i,p_K}^\top)^\top \in \mathbb{R}^p$ denote the i th subvector of $\boldsymbol{r}_0$ satisfying that $\boldsymbol{r}_0 = (\boldsymbol{r}_1^\top, \boldsymbol{r}_2^\top, \ldots, \boldsymbol{r}_n^\top)^\top$. For every k and k',

$$\widetilde{\alpha}_{kk} = \sum_{i=1}^{n} \boldsymbol{r}_{i,p_k}^\top \boldsymbol{r}_{i,p_k} / \{(n-1)p_k\},$$
$$\widetilde{\beta}_{kk} = \sum_{i=1}^{n} \operatorname{sum}(\boldsymbol{r}_{i,p_k}) \operatorname{sum}(\boldsymbol{r}_{i,p_k}) / \{(n-1)p_k^2\} = \sum_{i=1}^{n} \operatorname{sum}^2(\boldsymbol{r}_{i,p_k}) / \{(n-1)p_k^2\},$$
$$\widetilde{\beta}_{kk'} = \sum_{i=1}^{n} 2(\operatorname{sum}(\boldsymbol{r}_{i,p_k}) \operatorname{sum}(\boldsymbol{r}_{i,p_{k'}})) / \{2(n-1)p_k p_{k'}\} = \sum_{i=1}^{n} \{\operatorname{sum}(\boldsymbol{r}_{i,p_k}) \operatorname{sum}(\boldsymbol{r}_{i,p_{k'}})\} / \{(n-1)p_k p_{k'}\}.$$

By the facts that $\boldsymbol{r}_{i,p_k} = \boldsymbol{X}_{i,p_k} - \bar{\boldsymbol{X}}_{p_k}$ and $\mathbf{C}_{kk'} = \sum_{i=1}^{n} (\boldsymbol{X}_{i,p_k} - \bar{\boldsymbol{X}}_{p_k})(\boldsymbol{X}_{i,p_{k'}} - \bar{\boldsymbol{X}}_{p_{k'}})^\top =$

$\sum_{i=1}^{n} \boldsymbol{r}_{i,p_k} \boldsymbol{r}_{i,p_{k'}}^{\top}$ for every k and k', then,

$$
\begin{aligned}
k \neq k' : \operatorname{sum}(\mathbf{C}_{kk'}) &= \operatorname{sum}\left(\sum_{i=1}^{n} \boldsymbol{r}_{i,p_k} \boldsymbol{r}_{i,p_{k'}}^{\top}\right) = \sum_{i=1}^{n} \operatorname{sum}\left(\boldsymbol{r}_{i,p_k} \boldsymbol{r}_{i,p_{k'}}^{\top}\right) = \sum_{i=1}^{n} \operatorname{sum}\left(\boldsymbol{r}_{i,p_k}\right) \operatorname{sum}\left(\boldsymbol{r}_{i,p_{k'}}\right), \\
k = k' : \operatorname{sum}(\mathbf{C}_{kk'}) &= \operatorname{sum}\left(\sum_{i=1}^{n} \boldsymbol{r}_{i,p_k} \boldsymbol{r}_{i,p_k}^{\top}\right) = \sum_{i=1}^{n} \operatorname{sum}\left(\boldsymbol{r}_{i,p_k} \boldsymbol{r}_{i,p_k}^{\top}\right) = \sum_{i=1}^{n} \operatorname{sum}^2(\boldsymbol{r}_{i,p_k}), \\
k = k' : \operatorname{tr}(\mathbf{C}_{kk'}) &= \operatorname{tr}\left(\sum_{i=1}^{n} \boldsymbol{r}_{i,p_k} \boldsymbol{r}_{i,p_k}^{\top}\right) = \sum_{i=1}^{n} \operatorname{tr}\left(\boldsymbol{r}_{i,p_k} \boldsymbol{r}_{i,p_k}^{\top}\right) = \sum_{i=1}^{n} \boldsymbol{r}_{i,p_k}^{\top} \boldsymbol{r}_{i,p_k}.
\end{aligned}
$$

Finally, the estimators for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are respectively BLUE and BQUE, but also they are functions of complete statistics. Thus, we have the following best-unbiased estimators

$$
\widetilde{\boldsymbol{\mu}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i, \quad \widetilde{\alpha}_{kk} = \frac{1}{n-1} \frac{\operatorname{tr}(\mathbf{C}_{kk})}{p_k}, \quad \widetilde{\beta}_{kk'} = \begin{cases} \dfrac{1}{n-1} \dfrac{\operatorname{sum}(\mathbf{C}_{kk})}{p_k^2}, & k = k' \\ \dfrac{1}{n-1} \dfrac{\operatorname{sum}(\mathbf{C}_{kk'})}{p_k p_{k'}}, & k \neq k' \end{cases},
$$

for $\boldsymbol{\mu}$, α_{kk}, and $\beta_{kk'}$ for every k and k'. ■

B.3.5 Proof of Corollary 3.2.2

Proof of Corollary 3.2.2. The result can be obtained immediately from the properties of UB matrices in Chapter 2. ■

B.3.6 Proofs of Corollary 3.2.3 and Corollary B.2.1

of Corollary 3.2.3. Roy et al. (2016) and Koziol et al. (2017) provided their formula (4.17) for calculating the variance of a quadratic form $\boldsymbol{Y} A \boldsymbol{Y}^{\top}$ for some matrix A satisfying $A = \mathbf{P}_0 A \mathbf{P}_0$,

$$
\operatorname{var}\left(\boldsymbol{Y}^{\top} A \boldsymbol{Y}\right) = 2 \times \operatorname{tr}\left(\mathbf{P}_0 A \mathbf{V} A \mathbf{V}\right),
$$

where $\boldsymbol{Y} \sim N(\boldsymbol{\mu}_y, \mathbf{V})$ and $\mathbf{P}_0 = \mathbf{I}_{pn} - \mathbf{P}_0$ as defined in the proof of Theorem 3.2.1. Recall the proof of Theorem 3.2.1,

$$\widetilde{\alpha}_{kk} = \frac{\boldsymbol{r}_0^\top \left[\mathbf{I}_n \otimes \{\mathbf{E}_{kk} \circ \mathbf{I}(\boldsymbol{p})\}\right] \boldsymbol{r}_0}{(n-1)p_k}, \quad \widetilde{\beta}_{kk} = \frac{\boldsymbol{r}_0^\top \left[\mathbf{I}_n \otimes \{\mathbf{E}_{kk} \circ \mathbf{J}(\boldsymbol{p})\}\right] \boldsymbol{r}_0}{(n-1)p_k^2}, \quad \text{for every } k$$
$$\widetilde{\beta}_{kk'} = \frac{\boldsymbol{r}_0^\top \left[\mathbf{I}_n \otimes \{(\mathbf{E}_{kk'} + \mathbf{E}_{k'k}) \circ \mathbf{J}(\boldsymbol{p})\}\right] \boldsymbol{r}_0}{(n-1)(2p_k p_{k'})}, \quad \text{for every } k \neq k'.$$

Let $x = x(k,k)$, $y = y(k,k)$, and $z = z(k,k')$ denote the terms $\mathbf{E}_{kk} \circ \mathbf{I}(\boldsymbol{p})$, $\mathbf{E}_{kk} \circ \mathbf{J}(\boldsymbol{p})$, and $(\mathbf{E}_{kk'} + \mathbf{E}_{k'k}) \circ \mathbf{J}(\boldsymbol{p})$ respectively. Then, let ω denote any of x, y, and z and let $\mathbf{Q} = \mathbf{Q}(\omega) = \mathbf{P}_0 \times (\mathbf{I}_n \otimes \omega) \times \mathbf{P}_0$. So, $\mathbf{P}_0 \times \mathbf{Q} \times \mathbf{P}_0 = \mathbf{P}_0^2 \times (\mathbf{I}_n \otimes \omega) \times \mathbf{P}_0^2 = \mathbf{Q}$ due to the fact that $\mathbf{P}_0^2 = \mathbf{P}_0$. In addition, recall that $\mathbf{P}_0\mathbf{V} = \mathbf{V}\mathbf{P}_0$. Thus, $\mathrm{var}(\boldsymbol{Y}^\top \mathbf{Q} \boldsymbol{Y}) = 2 \times \mathrm{tr}(\mathbf{P}_0\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{V})$, which can be simplified as below:

$$\begin{aligned} 2 \times \mathrm{tr}\left(\mathbf{P}_0\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{V}\right) &= 2 \times \mathrm{tr}\left[\mathbf{P}_0\{\mathbf{P}_0(\mathbf{I}_n \otimes \omega)\mathbf{P}_0\}\mathbf{V}\{\mathbf{P}_0(\mathbf{I}_n \otimes \omega)\mathbf{P}_0\}\mathbf{V}\right] \\ &= 2 \times \mathrm{tr}\left\{\mathbf{P}_0(\mathbf{I}_n \otimes \omega)\mathbf{V}(\mathbf{I}_n \otimes \omega)\mathbf{V}\right\}, \end{aligned}$$

where we use the fact that $\mathbf{P}_0$ commutes with $\mathbf{I}_n \otimes \omega$ and $\mathbf{V}$, respectively. Therefore,

$$\begin{aligned} \mathrm{var}\left\{\boldsymbol{r}_0^\top(\mathbf{I}_n \otimes \omega)\boldsymbol{r}_0\right\} &= \mathrm{var}\left[\boldsymbol{Y}^\top\{\mathbf{P}_0(\mathbf{I}_n \otimes \omega)\mathbf{P}_0\}\boldsymbol{Y}\right] \\ &= 2 \times \mathrm{tr}\{\mathbf{P}_0(\mathbf{I}_n \otimes \omega)\mathbf{V}(\mathbf{I}_n \otimes \omega)\mathbf{V}\} \\ &= 2(n-1) \times \mathrm{tr}\{(\omega\boldsymbol{\Sigma})(\omega\boldsymbol{\Sigma})\} \\ &= 2(n-1)s_\omega, \end{aligned}$$

where s_ω denotes $\mathrm{tr}\{(\omega\boldsymbol{\Sigma})(\omega\boldsymbol{\Sigma})\}$. Hence, using these notations,

$$\mathrm{var}\,(\widetilde{\alpha}_{kk}) = \frac{1}{(n-1)^2 p_k^2}\,\mathrm{var}\left\{\boldsymbol{r}_0^\top(\mathbf{I}_n \otimes x)\boldsymbol{r}_0\right\} = \frac{2(n-1)s_x}{(n-1)^2 p_k^2} = \frac{2s_x}{(n-1)p_k^2} \quad \text{for every } k,$$

$$\mathrm{var}\left(\widetilde{\beta}_{kk}\right) = \frac{1}{(n-1)^2 p_k^4}\,\mathrm{var}\left\{\boldsymbol{r}_0^\top(\mathbf{I}_n \otimes y)\boldsymbol{r}_0\right\} = \frac{2(n-1)s_y}{(n-1)^2 p_k^4} = \frac{2s_y}{(n-1)p_k^4} \quad \text{for every } k,$$

$$\mathrm{var}\left(\widetilde{\beta}_{kk'}\right) = \frac{1}{4(n-1)^2 p_k^2 p_{k'}^2}\,\mathrm{var}\left\{\boldsymbol{r}_0^\top(\mathbf{I}_n \otimes z)\boldsymbol{r}_0\right\} = \frac{2(n-1)s_z}{4(n-1)^2 p_k^2 p_{k'}^2} = \frac{2s_z}{4(n-1)p_k^2 p_{k'}^2} \quad \text{for every } k \neq k',$$

where s_x, s_y and s_z are the traces of $(x\boldsymbol{\Sigma})^2$, $(y\boldsymbol{\Sigma})^2$ and $(z\boldsymbol{\Sigma})^2$ respectively. In fact, the expressions of the covariances among $\widetilde{\alpha}$'s and $\widetilde{\beta}$'s are associated with terms of the products of any two of $(x\boldsymbol{\Sigma})$, $(y\boldsymbol{\Sigma})$ and $(z\boldsymbol{\Sigma})$, where $x = x(k,k)$, $y = y(k,k)$, and $z = z(k,k')$ for every k and every $k \neq k'$. We focus on the calculations of $(x\boldsymbol{\Sigma})^2$, $(y\boldsymbol{\Sigma})^2$, and $(z\boldsymbol{\Sigma})^2$ only.

First, we calculate the $\mathrm{tr}(x\boldsymbol{\Sigma})^2$ as below:

$$\begin{aligned}
(x\boldsymbol{\Sigma})(x\boldsymbol{\Sigma}) &= \{(\mathbf{E}_{kk}\mathbf{A}) \circ \mathbf{I}(\boldsymbol{p}) + (\mathbf{E}_{kk}\mathbf{B}) \circ \mathbf{J}(\boldsymbol{p})\}\,\{(\mathbf{E}_{kk}\mathbf{A}) \circ \mathbf{I}(\boldsymbol{p}) + (\mathbf{E}_{kk}\mathbf{B}) \circ \mathbf{J}(\boldsymbol{p})\} \\
&= \{(\mathbf{E}_{kk}\mathbf{A})(\mathbf{E}_{kk}\mathbf{A})\} \circ \mathbf{I}(\boldsymbol{p}) + \{(\mathbf{E}_{kk}\mathbf{A})(\mathbf{E}_{kk}\mathbf{B})\} \circ \mathbf{J}(\boldsymbol{p}) + \{(\mathbf{E}_{kk}\mathbf{B})(\mathbf{E}_{kk}\mathbf{A})\} \circ \mathbf{J}(\boldsymbol{p}) \\
&\quad + \{(\mathbf{E}_{kk}\mathbf{B})\mathbf{P}(\mathbf{E}_{kk}\mathbf{B})\} \circ \mathbf{J}(\boldsymbol{p}) \\
&= \left\{(\mathbf{E}_{kk}\mathbf{A})^2\right\} \circ \mathbf{I}(\boldsymbol{p}) + \{(\mathbf{E}_{kk}\mathbf{A})(\mathbf{E}_{kk}\mathbf{B}) + (\mathbf{E}_{kk}\mathbf{B})(\mathbf{E}_{kk}\mathbf{A}) + (\mathbf{E}_{kk}\mathbf{B})\mathbf{P}(\mathbf{E}_{kk}\mathbf{B})\} \circ \mathbf{J}(\boldsymbol{p}),
\end{aligned}$$

therefore, for every k,

$$\mathrm{tr}\,\{(x\boldsymbol{\Sigma})(x\boldsymbol{\Sigma})\} = p_k(a_{kk}^2 + 2a_{kk}b_{kk} + p_k b_{kk}^2).$$

Then, we calculate the $\mathrm{tr}\,(y\boldsymbol{\Sigma})^2$ as below:

$$(y\boldsymbol{\Sigma})(y\boldsymbol{\Sigma}) = \{(\mathbf{E}_{kk}\mathbf{A} + \mathbf{E}_{kk}\mathbf{PB}) \circ \mathbf{J}(\boldsymbol{p})\}\{(\mathbf{E}_{kk}\mathbf{A} + \mathbf{E}_{kk}\mathbf{PB}) \circ \mathbf{J}(\boldsymbol{p})\}$$

$$= \{(\mathbf{E}_{kk}\mathbf{A} + \mathbf{E}_{kk}\mathbf{PB})\mathbf{P}(\mathbf{E}_{kk}\mathbf{A} + \mathbf{E}_{kk}\mathbf{PB})\} \circ \mathbf{J}(\boldsymbol{p}),$$

therefore, for every k,

$$\mathrm{tr}\{(y\boldsymbol{\Sigma})(y\boldsymbol{\Sigma})\} = p_k^2(a_{kk} + p_k b_{kk})^2.$$

Finally, we calculate the $\mathrm{tr}(z\boldsymbol{\Sigma})^2$ as below:

$$(z\boldsymbol{\Sigma})(z\boldsymbol{\Sigma}) = [\{(\mathbf{E}_{kk'} + \mathbf{E}_{k'k})\mathbf{A} + (\mathbf{E}_{kk'} + \mathbf{E}_{k'k})\mathbf{PB}\}\,\mathbf{P}\,\{(\mathbf{E}_{kk'} + \mathbf{E}_{k'k})\mathbf{A} + (\mathbf{E}_{kk'} + \mathbf{E}_{k'k})\mathbf{PB}\}] \circ \mathbf{J}(\boldsymbol{p}),$$

therefore, for every $k \neq k'$,

$$\mathrm{tr}\{(z_{kk'}\boldsymbol{\Sigma})(z_{kk'}\boldsymbol{\Sigma})\} = b_{k'k}^2 p_k^2 p_{k'}^2 + b_{kk'}^2 p_k^2 p_{k'}^2 + 2p_k p_{k'}(a_{kk} + p_k b_{kk})(a_{k'k'} + p_{k'k'} b_{k'k'}).$$

We complete the proof of variance estimators. ■

Proof of Corollary B.2.1. The expressions of the covariances among $\widetilde{\alpha}$'s and $\widetilde{\beta}$'s are associated with terms of the products of any two of $(x\boldsymbol{\Sigma})$, $(y\boldsymbol{\Sigma})$, and $(z\boldsymbol{\Sigma})$, where $x = x(k,k)$, $y = y(k,k)$, and $z = z(k,k')$ for every k and $k \neq k'$. One may obtain the covariance estimators by calculating $\mathrm{tr}\{(x\boldsymbol{\Sigma})(y\boldsymbol{\Sigma})\}$, $\mathrm{tr}\{(x\boldsymbol{\Sigma})(z\boldsymbol{\Sigma})\}$, and $\mathrm{tr}\{(y\boldsymbol{\Sigma})(z\boldsymbol{\Sigma})\}$, respectively in a similar manner. ■

B.3.7 Proof of Theorem 3.2.2

Lemma B.3.1. *(Gaussian Hanson-Wright inequality) Let $X \in \mathbb{R}^p$ be a Gaussian vector with mean zero and covariance matrix $\mathbf{\Sigma}$ and $A \in \mathbb{R}^{p \times p}$. Then, for every $t \geq 0$, we have*

$$\text{pr}\left(\left|X^\top AX - E\left(X^\top AX\right)\right| \geq t\right) \leq 2\exp\left(-C\min\left\{\frac{t^2}{\|\mathbf{\Sigma}^{1/2}A\mathbf{\Sigma}^{1/2}\|_F^2}, \frac{t}{\|\mathbf{\Sigma}^{1/2}A\mathbf{\Sigma}^{1/2}\|_{op}}\right\}\right),$$

where the positive constant C does not depend on p, A, and t.

Proof of Lemma B.3.1. It can be derived from the case with a sub-Gaussian variable. See the original proof in Hanson and Wright (1971) or Wright (1973) and a modern version in Rudelson and Vershynin (2013) or Vershynin (2018, Theorem 6.2.1). ■

Proof of Theorem 3.2.2. We aim to prove it within 4 steps.

Step 1, we would like to find out the upper bounds for the absolute values of the biases for the modified hard-threshold estimators under different cases.

Case 1: for a fixed λ and every fixed k and k', let $\sigma = b_{kk'}$ and $\widetilde{\sigma} = \widetilde{b}_{kk'}$, hence let $\widehat{\sigma} = \widetilde{\sigma}\mathbb{I}(|\widetilde{\sigma}| > \lambda)$. Assuming that $|\widetilde{\sigma} - \sigma| \leq \lambda/2$, or equivalently, $\sigma - \lambda/2 \leq \widetilde{\sigma} \leq \sigma + \lambda/2$, we obtain: if $\sigma \in [-\lambda/2, \lambda/2]$, then $\widetilde{\sigma} \in [-\lambda, \lambda]$, or equivalently, $|\widetilde{\sigma}| \leq \lambda$, and therefore, $\widehat{\sigma} = 0$ and $|\widehat{\sigma} - \sigma| = |\sigma| \in [0, \lambda/2]$; if $\sigma \in (3\lambda/2, \infty)$, then $\widetilde{\sigma} \in (\lambda, \infty)$, or equivalently, $|\widetilde{\sigma}| > \lambda$, and therefore, $\widehat{\sigma} = \widetilde{\sigma}$ and $|\widehat{\sigma} - \sigma| = |\widetilde{\sigma} - \sigma| \in [0, \lambda/2]$; if $\sigma \in (-\infty, -3\lambda/2)$, then $\widetilde{\sigma} \in (-\infty, -\lambda)$, or equivalently, $|\widetilde{\sigma}| > \lambda$, and therefore, $\widehat{\sigma} = \widetilde{\sigma}$ and $|\widehat{\sigma} - \sigma| = |\widetilde{\sigma} - \sigma| \in [0, \lambda/2]$; if $\sigma \in (\lambda/2, 3\lambda/2]$, then $\widetilde{\sigma} \in (0, 2\lambda]$ and it can be larger than λ or be smaller than λ, therefore

$\widehat{\sigma} \in \{0, \widetilde{\sigma}\}$ and $|\widehat{\sigma} - \sigma| = \begin{cases} |\widetilde{\sigma} - \sigma| \in [0, \lambda/2], & \widetilde{\sigma} \in (\lambda, 2\lambda] \\ |\sigma| \in (\lambda/2, 3\lambda/2], & \widetilde{\sigma} \in (0, \lambda] \end{cases}$, which is between 0 and $3\lambda/2$; if $\sigma \in [-3\lambda/2, -\lambda/2)$, then $\widetilde{\sigma} \in [-2\lambda, 0)$ and it can be larger than $-\lambda$ or be smaller than $-\lambda$, therefore $\widehat{\sigma} \in \{0, \widetilde{\sigma}\}$ and $|\widehat{\sigma} - \sigma| = \begin{cases} |\widetilde{\sigma} - \sigma| \in [0, \lambda/2], & \widetilde{\sigma} \in [-2\lambda, -\lambda) \\ |\sigma| \in [\lambda/2, 3\lambda/2], & \widetilde{\sigma} \in [-\lambda, 0) \end{cases}$, which is also between 0 and $3\lambda/2$. Put the above arguments together, for a fixed λ, under the assumption $|\widetilde{\sigma} - \sigma| \leq \lambda/2$,

$$|\widehat{\sigma} - \sigma| = \begin{cases} |\widetilde{\sigma} - \sigma| \in [0, \lambda/2], & \sigma \in (-\infty, -3\lambda/2) \\ \{|\widetilde{\sigma} - \sigma|, |\sigma|\} \leq |\sigma| \vee (\lambda/2), & \sigma \in [-3\lambda/2, -\lambda/2) \\ |\sigma| \in [0, \lambda/2], & \sigma \in [-\lambda/2, \lambda/2] \\ \{|\widetilde{\sigma} - \sigma|, |\sigma|\} \leq |\sigma| \vee (\lambda/2), & \sigma \in (\lambda/2, 3\lambda/2] \\ |\widetilde{\sigma} - \sigma| \in [0, \lambda/2], & \sigma \in (3\lambda/2, \infty) \end{cases},$$

therefore the maximum of $|\widehat{\sigma} - \sigma|$ might be either $|\sigma|$ if $\lambda/2 < |\sigma| \leq 3\lambda/2$ or $\lambda/2$ otherwise. In other words, we may write $|\widehat{\sigma} - \sigma| \leq |\sigma| \wedge (3\lambda/2)$.

Case 2: for a fixed λ and every fixed k, let $\sigma = a_{kk}$ and $\widetilde{\sigma} = \widetilde{a}_{kk}$, hence let $\widehat{\sigma} = \widetilde{\sigma}\mathbb{I}(|\widetilde{\sigma}| > \lambda)$. We assume that $|\widetilde{\alpha}_{kk} - \alpha_{kk}| \leq \lambda/2$ and $|\widetilde{b}_{kk} - b_{kk}| \leq \lambda/2$ holds simultaneously, then

$$|\widetilde{a}_{kk} - a_{kk}| = |\widetilde{\alpha}_{kk} - \widetilde{b}_{kk} - (\alpha_{kk} - b_{kk})| = |(\widetilde{\alpha}_{kk} - \alpha_{kk}) - (\widetilde{b}_{kk} - b_{kk})| \leq |\widetilde{\alpha}_{kk} - \alpha_{kk}| + |\widetilde{b}_{kk} - b_{kk}| \leq \lambda.$$

So, the above assumption is equivalent implies that $|\widetilde{\sigma} - \sigma| \leq \lambda$, or equivalently, $\sigma - \lambda \leq \widetilde{\sigma} \leq \sigma + \lambda$: if $\sigma \in (-\infty, -2\lambda)$, then $\widetilde{\sigma} \in (-\infty, -\lambda)$ and therefore $|\widetilde{\sigma}| > \lambda$, and $\widehat{\sigma} = \widetilde{\sigma}$, and thus

$|\widehat{\sigma}-\sigma| = |\widetilde{\sigma}-\sigma| \in [0,\lambda]$; if $\sigma = 0$, then $|\widetilde{\sigma}| \leq \lambda$ and therefore $\widehat{\sigma} = 0$, and $|\widehat{\sigma}-\sigma| = |\sigma| = 0$; if $\sigma \in (0, 2\lambda]$, then $\widetilde{\sigma} \in (-\lambda, 3\lambda]$ and therefore $|\widetilde{\sigma}|$ can be larger than λ or be smaller than λ, thus $\widehat{\sigma} = \{0, \widetilde{\sigma}\}$, and thus $|\widehat{\sigma}-\sigma| = \begin{cases} |\widetilde{\sigma}-\sigma| \in [0,\lambda], & \widetilde{\sigma} \in (\lambda, 3\lambda] \\ |\sigma| \in (0, 2\lambda], & \widetilde{\sigma} \in (-\lambda, \lambda] \end{cases}$, which is between 0 and 2λ; if $\sigma \in [-2\lambda, 0)$, then $\widetilde{\sigma} \in [-3\lambda, \lambda)$ and therefore $|\widetilde{\sigma}|$ can be larger than λ or be smaller than λ, thus $\widehat{\sigma} = \{\widetilde{\sigma}, 0\}$, and thus $|\widehat{\sigma}-\sigma| = \begin{cases} |\widetilde{\sigma}-\sigma| \in [0,\lambda], & \widetilde{\sigma} \in [-3\lambda, -\lambda) \\ |\sigma| \in (0, 2\lambda], & \widetilde{\sigma} \in [-\lambda, \lambda) \end{cases}$, which is between 0 and 2λ; if $\sigma \in (2\lambda, \infty)$, then $\widetilde{\sigma} \in (\lambda, \infty)$ and therefore $|\widetilde{\sigma}| > \lambda$, and $\widehat{\sigma} = \widetilde{\sigma}$, and thus $|\widehat{\sigma}-\sigma| = |\widetilde{\sigma}-\sigma| \in [0,\lambda]$. Put the above arguments together, for a fixed λ, under the assumption $|\widetilde{\sigma}-\sigma| \leq \lambda$,

$$|\widehat{\sigma}-\sigma| = \begin{cases} |\widetilde{\sigma}-\sigma| \in [0,\lambda], & \sigma \in (-\infty, -2\lambda) \\ \{|\widetilde{\sigma}-\sigma|, |\sigma|\} \leq |\sigma| \vee \lambda, & \sigma \in [-2\lambda, 0) \\ 0, & \sigma = 0 \\ \{|\widetilde{\sigma}-\sigma|, |\sigma|\} \leq |\sigma| \vee \lambda, & \sigma \in (0, 2\lambda] \\ |\widetilde{\sigma}-\sigma| \in [0,\lambda], & \sigma \in (2\lambda, \infty) \end{cases},$$

therefore the maximum of $|\widehat{\sigma}-\sigma|$ might be either $|\sigma|$ if $0 < |\sigma| \leq 2\lambda$ or λ otherwise. In other words, we may write $|\widehat{\sigma}-\sigma| \leq |\sigma| \wedge (2\lambda)$. Thus, at Step 1, we conclude that $|\widehat{a}_{kk} - a_{kk}| \leq |a_{kk}| \wedge (2\lambda)$ and $|\widehat{b}_{kk'} - b_{kk'}| \leq |b_{kk'}| \wedge (3\lambda/2)$ for a fixed λ and every k and k' under the assumptions: $\max_{1\leq k\leq K} |\widetilde{\alpha}_{kk} - \alpha_{kk}| \leq \lambda/2$ and $\max_{1\leq k,k'\leq K} |\widetilde{b}_{kk'} - b_{kk'}| \leq \lambda/2$. We denote the

following event sets

$$\mathbf{E}_N^{(0)} = \left\{ \max_{1\le k\le K} |\widetilde{a}_{kk} - a_{kk}| \le \lambda \right\}, \quad \mathbf{E}_N^{(1)} = \left\{ \max_{1\le k\le K} |\widetilde{\alpha}_{kk} - \alpha_{kk}| \le \lambda/2 \right\},$$
$$\mathbf{E}_N^{(2)} = \left\{ \max_{1\le k\le K} \left|\widetilde{b}_{kk} - b_{kk}\right| \le \lambda/2 \right\}, \quad \mathbf{E}_N^{(3)} = \left\{ \max_{1\le k\ne k'\le K} \left|\widetilde{b}_{kk'} - b_{kk'}\right| \le \lambda/2 \right\},$$

and the event sets $\mathbf{E}_N^{(1)}$ and $\mathbf{E}_N^{(2)}$ implies that $\mathbf{E}_N^{(0)}$. The probability of their complement sets can be bounded as below:

$$\mathrm{pr}\left(\mathbf{E}_N^{(1),\complement}\right) = \mathrm{pr}\left(\max_{1\le k\le K} |\widetilde{\alpha}_{kk} - \alpha_{kk}| > \lambda/2 \right) \overset{[1]}{\le} K \times \mathrm{pr}\left(|\widetilde{\alpha}_{kk} - \alpha_{kk}| > \lambda/2\right),$$
$$\mathrm{pr}\left(\mathbf{E}_N^{(2),\complement}\right) = \mathrm{pr}\left(\max_{1\le k\le K} \left|\widetilde{b}_{kk} - b_{kk}\right| > \lambda/2 \right) \le K \times \mathrm{pr}\left(\left|\widetilde{b}_{kk} - b_{kk}\right| > \lambda/2\right),$$
$$\mathrm{pr}\left(\mathbf{E}_N^{(3),\complement}\right) = \mathrm{pr}\left(\max_{1\le k\ne k'\le K} \left|\widetilde{b}_{kk'} - b_{kk'}\right| > \lambda/2 \right) \le K \times (K-1) \times \mathrm{pr}\left(\left|\widetilde{b}_{kk'} - b_{kk'}\right| > \lambda/2\right),$$

where $[1]$ holds due to the union bound.

Step 2, we would like to compute the upper bounds for the above three probabilities respectively.

Let $\mathbf{E}_{kk'} \in \mathbb{R}^{K\times K}$ denote the matrix whose (k,k') entry is 1 and the other entries are 0 for every k and k'. Then, let $\mathbf{A}_{kk} = (p_k^{-1}\mathbf{E}_{kk}) \circ \mathbf{I}(\boldsymbol{p})$, $\mathbf{B}_{kk} = \{-p_k^{-1}(p_k-1)^{-1}\mathbf{E}_{kk}\} \circ \mathbf{I}(\boldsymbol{p}) + \{p_k^{-1}(p_k-1)^{-1}\mathbf{E}_{kk}\} \circ \mathbf{J}(\boldsymbol{p})$ and $\mathbf{B}_{kk'} = \{(p_k p_{k'})^{-1}\mathbf{E}_{kk'}\} \circ \mathbf{J}(\boldsymbol{p})$ for every k and $k \ne k'$. Let $\boldsymbol{W} = (\boldsymbol{X}_1^\top, \ldots, \boldsymbol{X}_n^\top)^\top \in \mathbb{R}^{pn}$ and decompose $\boldsymbol{X} = (\boldsymbol{X}_{p_1}^\top, \ldots, \boldsymbol{X}_{p_K}^\top)^\top$, where $\boldsymbol{X}_{p_k} = (\boldsymbol{X}_{\bar{p}_{k-1}+1}, \ldots, \boldsymbol{X}_{\bar{p}_k})^\top \in \mathbb{R}^{p_k}$ for every k, $\bar{p}_0 = 0$ and $\bar{p}_k = \sum_{k'=1}^k p_{k'}$ denote the sum of the first

k elements of $\boldsymbol{p}$ for $k = 1, \ldots, K$. Note that $\boldsymbol{\mu} = \mathbf{0}_{p\times 1}$ is known and $\mathbf{S} = \mathbf{X}^\top\mathbf{X}/n$, thus,

$$\boldsymbol{W}^\top (\mathbf{I}_n \otimes \mathbf{A}_{kk})\, \boldsymbol{W} = \sum_{i=1}^{n} \boldsymbol{X}_i^\top \mathbf{A}_{kk} \boldsymbol{X}_i = \frac{1}{p_k} \sum_{i=1}^{n} \operatorname{tr}\left(\boldsymbol{X}_{i,p_k} \boldsymbol{X}_{i,p_k}^\top\right) = \frac{n}{p_k} \operatorname{tr}(\mathbf{S}_{kk}) = n\widetilde{\alpha}_{kk},$$
$$\boldsymbol{W}^\top (\mathbf{I}_n \otimes \mathbf{B}_{kk})\, \boldsymbol{W} = \sum_{i=1}^{n} \boldsymbol{X}_i^\top \mathbf{B}_{kk} \boldsymbol{X}_i = \frac{n}{p_k(p_k-1)} \left\{\operatorname{sum}(\mathbf{S}_{kk}) - \operatorname{tr}(\mathbf{S}_{kk})\right\} = n\widetilde{b}_{kk},$$
$$\boldsymbol{W}^\top (\mathbf{I}_n \otimes \mathbf{B}_{kk'})\, \boldsymbol{W} = \sum_{i=1}^{n} \boldsymbol{X}_i^\top \mathbf{B}_{kk'} \boldsymbol{X}_i = \frac{1}{p_k p_{k'}} \sum_{i=1}^{n} \operatorname{sum}(\boldsymbol{X}_{i,p_k} \boldsymbol{X}_{i,p_{k'}}^\top) = \frac{n}{p_k p_{k'}} \operatorname{sum}(\mathbf{S}_{kk'}) = n\widetilde{b}_{kk'},$$

for every k and for every $k \neq k'$ respectively.

On the one hand, $\widetilde{\alpha}_{kk}$, $\widetilde{b}_{kk'}$ are unbiased estimators of α_{kk} and $b_{kk'}$ for every k and k', let

$$Q_{kk}^{(\alpha)} = \boldsymbol{W}^\top (\mathbf{I}_n \otimes \mathbf{A}_{kk})\, \boldsymbol{W} - \mathrm{E}\left\{\boldsymbol{W}^\top (\mathbf{I}_n \otimes \mathbf{A}_{kk})\, \boldsymbol{W}\right\} = n\left(\widetilde{\alpha}_{kk} - \alpha_{kk}\right),$$
$$Q_{kk'}^{(b)} = \boldsymbol{W}^\top (\mathbf{I}_n \otimes \mathbf{B}_{kk'})\, \boldsymbol{W} - \mathrm{E}\left\{\boldsymbol{W}^\top (\mathbf{I}_n \otimes \mathbf{B}_{kk'})\, \boldsymbol{W}\right\} = n\left(\widetilde{b}_{kk'} - b_{kk'}\right),$$

for every k and k'. On the other hand, applying the result of Lemma B.3.1, given a $\delta > 0$,

$$
\begin{aligned}
&\mathrm{pr}\left(|\widetilde{\alpha}_{kk} - \alpha_{kk}| > \delta\right) \\
&= \mathrm{pr}\left(\left|Q_{kk}^{(\alpha)}\right| > n\delta\right) \\
&\leq 2\exp\left(-C\min\left\{\frac{(n\delta)^2}{\left\|\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\left(\mathbf{I}_n \otimes \mathbf{A}_{kk}\right)\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\right\|_{\mathrm{F}}^2}, \frac{n\delta}{\left\|\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\left(\mathbf{I}_n \otimes \mathbf{A}_{kk}\right)\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\right\|_{\mathrm{op}}}\right\}\right), \\
&\mathrm{pr}\left(\left|\widetilde{b}_{kk} - b_{kk}\right| > \delta\right) \\
&\leq 2\exp\left(-C\min\left\{\frac{(n\delta)^2}{\left\|\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\left(\mathbf{I}_n \otimes \mathbf{B}_{kk}\right)\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\right\|_{\mathrm{F}}^2}, \frac{n\delta}{\left\|\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\left(\mathbf{I}_n \otimes \mathbf{B}_{kk}\right)\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\right\|_{\mathrm{op}}}\right\}\right), \\
&\mathrm{pr}\left(\left|\widetilde{b}_{kk'} - b_{kk'}\right| > \delta\right) \\
&\leq 2\exp\left(-C\min\left\{\frac{(n\delta)^2}{\left\|\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\left(\mathbf{I}_n \otimes \mathbf{B}_{kk'}\right)\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\right\|_{\mathrm{F}}^2}, \frac{n\delta}{\left\|\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\left(\mathbf{I}_n \otimes \mathbf{B}_{kk'}\right)\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\right\|_{\mathrm{op}}}\right\}\right),
\end{aligned}
$$

for every k, for every k, and for every $k \neq k'$ respectively, where C is a positive constant. By the properties of Kronecker product and Frobenius norm,

$$
\begin{aligned}
\left\|\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\left(\mathbf{I}_n \otimes \mathbf{A}_{kk}\right)\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\right\|_{\mathrm{F}}^2 &= \left\|\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}^{1/2}\right)\left(\mathbf{I}_n \otimes \mathbf{A}_{kk}\right)\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}^{1/2}\right)\right\|_{\mathrm{F}}^2 \\
&= n\left\|\boldsymbol{\Sigma}^{1/2}\mathbf{A}_{kk}\boldsymbol{\Sigma}^{1/2}\right\|_{\mathrm{F}}^2 \\
&= n \times \mathrm{tr}\left\{\left(\boldsymbol{\Sigma}^{1/2}\mathbf{A}_{kk}\boldsymbol{\Sigma}^{1/2}\right)\left(\boldsymbol{\Sigma}^{1/2}\mathbf{A}_{kk}\boldsymbol{\Sigma}^{1/2}\right)\right\} \\
&= n \times \mathrm{tr}\left(\boldsymbol{\Sigma}^{1/2}\mathbf{A}_{kk}\boldsymbol{\Sigma}\mathbf{A}_{kk}\boldsymbol{\Sigma}^{1/2}\right) \\
&= n \times \mathrm{tr}\left(\mathbf{A}_{kk}\boldsymbol{\Sigma}\mathbf{A}_{kk}\boldsymbol{\Sigma}\right),
\end{aligned}
$$

and

$$\left\|\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\left(\mathbf{I}_n \otimes \mathbf{B}_{kk}\right)\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\right\|_{\mathrm{F}}^2 = n \times \operatorname{tr}\left(\mathbf{B}_{kk}\boldsymbol{\Sigma}\mathbf{B}_{kk}\boldsymbol{\Sigma}\right),$$
$$\left\|\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\left(\mathbf{I}_n \otimes \mathbf{B}_{kk'}\right)\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\right\|_{\mathrm{F}}^2 = n \times \operatorname{tr}\left(\mathbf{B}_{kk'}\boldsymbol{\Sigma}\mathbf{B}_{kk'}\boldsymbol{\Sigma}\right).$$

Then, calculate the traces,

$$n \times \operatorname{tr}\left(\mathbf{A}_{kk}\boldsymbol{\Sigma}\mathbf{A}_{kk}\boldsymbol{\Sigma}\right) = \frac{n}{p_k}\left(a_{kk}^2 + 2a_{kk}b_{kk} + p_k b_{kk}^2\right),$$
$$n \times \operatorname{tr}\left(\mathbf{B}_{kk}\boldsymbol{\Sigma}\mathbf{B}_{kk}\boldsymbol{\Sigma}\right) = \frac{n}{p_k(p_k-1)^2}\left\{a_{kk}^2 - 2a_{kk}(a_{kk} - b_{kk} + p_k b_{kk}) + p_k\left(a_{kk} + p_k b_{kk} - b_{kk}\right)^2\right\},$$
$$n \times \operatorname{tr}\left(\mathbf{B}_{kk'}\boldsymbol{\Sigma}\mathbf{B}_{kk'}\boldsymbol{\Sigma}\right) = n b_{k'k}^2.$$

By the properties of the operator norm, let $\lambda_{\max}(M)$ denote the largest eigenvalue of M,

$$\begin{aligned}
\left\|\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\left(\mathbf{I}_n \otimes \mathbf{A}_{kk}\right)\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\right\|_{\mathrm{op}}^2 &= \left\|\mathbf{I}_n \otimes \left(\boldsymbol{\Sigma}^{1/2}\mathbf{A}_{kk}\boldsymbol{\Sigma}^{1/2}\right)\right\|_{\mathrm{op}}^2 \\
&= \lambda_{\max}\left[\left\{\mathbf{I}_n \otimes \left(\boldsymbol{\Sigma}^{1/2}\mathbf{A}_{kk}\boldsymbol{\Sigma}^{1/2}\right)\right\}\left\{\mathbf{I}_n \otimes \left(\boldsymbol{\Sigma}^{1/2}\mathbf{A}_{kk}\boldsymbol{\Sigma}^{1/2}\right)\right\}\right] \\
&= \lambda_{\max}\left[\mathbf{I}_n \otimes \left\{\left(\boldsymbol{\Sigma}^{1/2}\mathbf{A}_{kk}\boldsymbol{\Sigma}^{1/2}\right)\left(\boldsymbol{\Sigma}^{1/2}\mathbf{A}_{kk}\boldsymbol{\Sigma}^{1/2}\right)\right\}\right] \\
&= \lambda_{\max}\left\{\left(\boldsymbol{\Sigma}^{1/2}\mathbf{A}_{kk}\boldsymbol{\Sigma}^{1/2}\right)\left(\boldsymbol{\Sigma}^{1/2}\mathbf{A}_{kk}\boldsymbol{\Sigma}^{1/2}\right)\right\} \\
&= \lambda_{\max}\left(\boldsymbol{\Sigma}^{1/2}\mathbf{A}_{kk}\boldsymbol{\Sigma}\mathbf{A}_{kk}\boldsymbol{\Sigma}^{1/2}\right),
\end{aligned}$$

and

$$\left\|\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\left(\mathbf{I}_n \otimes b_{kk}\right)\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\right\|_{\mathrm{op}}^2 = \lambda_{\max}\left(\boldsymbol{\Sigma}^{1/2}\mathbf{B}_{kk}\boldsymbol{\Sigma}\mathbf{B}_{kk}\boldsymbol{\Sigma}^{1/2}\right),$$
$$\left\|\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\left(\mathbf{I}_n \otimes \mathbf{B}_{kk'}\right)\left(\mathbf{I}_n \otimes \boldsymbol{\Sigma}\right)^{1/2}\right\|_{\mathrm{op}}^2 = \lambda_{\max}\left(\boldsymbol{\Sigma}^{1/2}\mathbf{B}_{kk'}\boldsymbol{\Sigma}\mathbf{B}_{kk'}\boldsymbol{\Sigma}^{1/2}\right).$$

Therefore, under the assumption about $\boldsymbol{\Sigma}$, there exist positive constants $C_{1,*}$ and $C_{2,*}$ satisfying

that

$$C_{1,*}^2 = C_{\mathbf{A},\mathbf{B},\boldsymbol{p}}^2 = \max_K \max_{1\le k,k'\le K}$$

$$\left\{\lambda_{\max}\left(\boldsymbol{\Sigma}^{1/2}\mathbf{A}_{kk}\boldsymbol{\Sigma}\mathbf{A}_{kk}\boldsymbol{\Sigma}^{1/2}\right), \lambda_{\max}\left(\boldsymbol{\Sigma}^{1/2}\mathbf{B}_{kk}\boldsymbol{\Sigma}\mathbf{B}_{kk}\boldsymbol{\Sigma}^{1/2}\right), \lambda_{\max}\left(\boldsymbol{\Sigma}^{1/2}\mathbf{B}_{kk'}\boldsymbol{\Sigma}\mathbf{B}_{kk'}\boldsymbol{\Sigma}^{1/2}\right)\right\} < \infty,$$

and

$$C_{2,*}^2 = C_{\mathbf{A},\mathbf{B},\boldsymbol{p}}^2 = \max_K \max_{1\le k,k'\le K} \left\{\mathrm{tr}\left(\mathbf{A}_{kk}\boldsymbol{\Sigma}\mathbf{A}_{kk}\boldsymbol{\Sigma}\right), \mathrm{tr}\left(\mathbf{B}_{kk}\boldsymbol{\Sigma}\mathbf{B}_{kk}\boldsymbol{\Sigma}\right), \mathrm{tr}\left(\mathbf{B}_{kk'}\boldsymbol{\Sigma}\mathbf{B}_{kk'}\boldsymbol{\Sigma}\right)\right\} < \infty.$$

Therefore,

$$\begin{aligned}
K \times \mathrm{pr}\left(|\widetilde{\alpha}_{kk} - \alpha_{kk}| > \delta\right) &\le 2K \exp\left(-C \min\left\{\frac{n\delta^2}{C_{2,*}^2}, \frac{n\delta}{C_{1,*}}\right\}\right), \\
K \times \mathrm{pr}\left(\left|\widetilde{b}_{kk} - b_{kk}\right| > \delta\right) &\le 2K \exp\left(-C \min\left\{\frac{n\delta^2}{C_{2,*}^2}, \frac{n\delta}{C_{1,*}}\right\}\right), \\
K \times (K-1) \times \mathrm{pr}\left(\left|\widetilde{b}_{kk'} - b_{kk'}\right| > \delta\right) &\le 2K(K-1) \exp\left(-C \min\left\{\frac{n\delta^2}{C_{2,*}^2}, \frac{n\delta}{C_{1,*}}\right\}\right).
\end{aligned}$$

Step 3, let $U = Cn \min\left\{\lambda^2(2C_{2,*}^2)^{-1}, \lambda(2C_{1,*})^{-1}\right\} = Cn \min\{\lambda^2, \lambda\}$, where constants $C_{1,*}$ and $C_{2,*}$ are absorbed into constant C. Set $n^{-1}\log(K) \to 0$ as $K > n \to \infty$. Let $\lambda = \eta\sqrt{\log(K)/n} \to 0^+$, then $n\lambda^2 = \eta^2 \log(K) \to \infty$, while $\lambda^2 < \lambda$ since $\lambda \to 0^+$. So, we have

$$\lim_{n\to\infty} \frac{U}{\log(K)} = \lim_{n\to\infty} \frac{Cn\lambda^2}{\log(K)} = \lim_{n\to\infty} \frac{C\eta^2 \log(K)}{\log(K)} = C\eta^2.$$

Therefore, the Forbenius norm is bounded as below,

$$\begin{aligned}\left\|\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\right\|_{\mathrm{F}}^{2} &= \sum_{k=1}^{K} p_k\left(\widehat{\alpha}_{kk}-\alpha_{kk}\right)^2+\sum_{k=1}^{K} p_k(p_k-1)\left(\widehat{b}_{kk}-b_{kk}\right)^2+\sum_{k\neq k'} p_k p_{k'}\left(\widehat{b}_{kk'}-b_{kk'}\right)^2 \\ &\leq \sum_{k=1}^{K} p_k\left\{\alpha_{kk}^2\wedge(2\lambda)^2\right\}+\sum_{k=1}^{K} p_k(p_k-1)\left\{b_{kk}^2\wedge(3\lambda/2)^2\right\}+\sum_{k\neq k'} p_k p_{k'}\left\{b_{kk'}^2\wedge(3\lambda/2)^2\right\},\end{aligned}$$

and the spectral norm is bounded below

$$\begin{aligned}\left\|\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\right\|_{\mathrm{S}} &\leq \left\|\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\right\|_{1} \\ &= \sup_{1\leq j'\leq p}\left\{\left|\alpha_{k(j')k(j')}\right|\wedge(2\lambda)+\left(p_{k(j')}-1\right)\left|b_{k(j')k(j')}\right|\wedge(3\lambda/2)+\sum_{k\neq k(j')} p_k\left|b_{kk(j')}\right|\wedge(3\lambda/2)\right\},\end{aligned}$$

where $k(j')=k$ if $j'\in\{\bar{p}_{k-1}+1,\bar{p}_{k-1}+2,\ldots,\bar{p}_k\}$ for every k.

Finally, step 4, set $0<p_0<2$ and therefore x^{2-p_0} is monotonically increasing on $(0,\infty)$. Let $\omega_{kk}^{(\alpha),2}=\alpha_{kk}^2\wedge(2\lambda)^2$, $\omega_{kk'}^{(b),2}=b_{kk'}^2\wedge(3\lambda/2)^2$ for every k and k'. After doing some algebra, we have

$$\begin{aligned}\sum_{k=1}^{K} p_k\left\{\alpha_{kk}^2\wedge(2\lambda)^2\right\} &= \sum_{k=1}^{K} p_k\left|\omega_{kk}^{(\alpha)}\right|^{p_0}\left|\omega_{kk}^{(\alpha)}\right|^{2-p_0} \\ &\leq \sum_{k=1}^{K} p_k\left|\omega_{kk}^{(\alpha)}\right|^{p_0}\left(\max_{1\leq k\leq K}\left|\omega_{kk}^{(\alpha)}\right|\right)^{2-p_0} \\ &= \left(\max_{1\leq k\leq K}\left|\omega_{kk}^{(\alpha)}\right|\right)^{2-p_0}\left(\sum_{k=1}^{K} p_k\left|\omega_{kk}^{(\alpha)}\right|^{p_0}\right) \\ &\leq (2\lambda)^{2-p_0}\left(\sum_{k=1}^{K} p_k\left|\alpha_{kk}\right|^{p_0}\right),\end{aligned}$$

and

$$\begin{aligned}
\sum_{k=1}^{K} p_k(p_k-1)\left\{b_{kk}^2 \wedge (3\lambda/2)^2\right\} &\le \left(\max_{1\le k\le K}\left|\omega_{kk}^{(b)}\right|\right)^{2-p_0}\left\{\sum_{k=1}^{K} p_k(p_k-1)\left|\omega_{kk}^{(b)}\right|^{p_0}\right\} \\
&\le (3\lambda/2)^{2-p_0}\left\{\sum_{k=1}^{K} p_k(p_k-1)\left|b_{kk}\right|^{p_0}\right\}, \\
\sum_{k\neq k'} p_k p_{k'}\left\{b_{kk'}^2 \wedge (3\lambda/2)^2\right\} &\le \left(\max_{1\le k\neq k'\le K}\left|\omega_{kk'}^{(b)}\right|\right)^{2-p_0}\left(\sum_{k=1}^{K} p_k p_{k'}\left|\omega_{kk'}^{(b)}\right|^{p_0}\right) \\
&\le (3\lambda/2)^{2-p_0}\left(\sum_{k\neq k'} p_k p_{k'}\left|b_{kk'}\right|^{p_0}\right).
\end{aligned}$$

Put them together,

$$\begin{aligned}
&\left\|\widehat{\mathbf{\Sigma}}-\mathbf{\Sigma}\right\|_{\mathrm{F}}^2 \\
&\le \sum_{k=1}^{K} p_k\left\{\alpha_{kk}^2 \wedge (2\lambda)^2\right\} + \sum_{k=1}^{K} p_k(p_k-1)\left\{b_{kk}^2 \wedge (3\lambda/2)^2\right\} + \sum_{k\neq k'} p_k p_{k'}\left\{b_{kk'}^2 \wedge (3\lambda/2)^2\right\} \\
&\le (2\lambda)^{2-p_0}\left(\sum_{k=1}^{K} p_k\left|\alpha_{kk}\right|^{p_0}\right) + (3\lambda/2)^{2-p_0}\left(\sum_{k=1}^{K} p_k(p_k-1)\left|b_{kk}\right|^{p_0}\right) + (3\lambda/2)^{2-p_0}\left(\sum_{k\neq k'} p_k p_{k'}\left|b_{kk'}\right|^{p_0}\right) \\
&\le (2\lambda)^{2-p_0}\left\|\mathbf{\Sigma}\right\|_{p_0}^{p_0,(\text{vector})},
\end{aligned}$$

where $\|\cdot\|_p^{(\text{vector})}$ denotes the entrywise norm. For the spectral norm, consider $0 < q_0 < 1$,

$$
\begin{aligned}
&\left\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\right\|_{\text{S}} \\
&= \sup_{1\le j'\le p}\left\{\left|\alpha_{k(j')k(j')}\right| \wedge (2\lambda) + \left(p_{k(j')} - 1\right)\left|b_{k(j')k(j')}\right| \wedge (3\lambda/2) + \sum_{k\neq k(j')} p_k \left|b_{kk(j')}\right| \wedge (3\lambda/2)\right\} \\
&\le \sup_{1\le j'\le p}\left\{(2\lambda)^{1-q_0}\left|\alpha_{k(j')k(j')}\right|^{q_0} + (3\lambda/2)^{1-q_0}\left(p_{k(j')} - 1\right)\left|b_{k(j')k(j')}\right|^{q_0} + (3\lambda/2)^{1-q_0}\sum_{k\neq k(j')} p_k \left|b_{kk(j')}\right|^{q_0}\right\} \\
&\le (2\lambda)^{1-q_0} \sup_{1\le j'\le p}\left\{\left|\alpha_{k(j')k(j')}\right|^{q_0} + \left(p_{k(j')} - 1\right)\left|b_{k(j')k(j')}\right|^{q_0} + \sum_{k\neq k(j')} p_k \left|b_{kk(j')}\right|^{q_0}\right\} \\
&\le (2\lambda)^{1-q_0}\left\|\boldsymbol{\Sigma}^{q_0}\right\|_1,
\end{aligned}
$$

where $\|\cdot^{q_0}\|_1$ denotes the maximum absolute column sum (with q_0-th power entrywisely).

In summary, given $\log(K)/n \to 0$ as $K = K_n > n \to \infty$, we select $C\eta^2 = 5$ and $\lambda = \eta\sqrt{\log(K)/n} \to 0$, then $2K(K-1)\exp(-U) \to 0$. Furthermore, for $0 < p_0 < 2$ and $0 < q_0 < 1$, $\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\text{F}}^2 \le (2\lambda)^{2-p_0}\|\boldsymbol{\Sigma}\|_{p_0}^{p_0,(\text{vector})}$ and $\|\widehat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\text{S}} \le (2\lambda)^{1-q_0}\|\boldsymbol{\Sigma}^{q_0}\|_1$ with probability 1. ■

B.4 Extra Simulation Studies

B.4.1 Extra simulation studies in Scenario 1

Examination of the accuracy of the covariance estimator. Based on Corollary B.2.1, we would like to verify the accuracy of covariance estimators in Scenario 1, for each pair of $\alpha_{0,11}, \ldots, \alpha_{0,KK}$ and $\beta_{0,11}, \ldots, \beta_{0,KK}$ for the sample sizes $n = 50$, 100, or 150, with 1000 Monte Carlo replicates. For each replicate, we compute the covariance estimate of every pair of α_0's and β_0's by substituting the estimates $\widetilde{\alpha}_{kk}$ and $\widetilde{\beta}_{kk'}$ in Corollary B.2.1. Based on 1000 replicates, the average

of estimated covariance is obtained (denoted by AC) for every pair of α_0's and β_0's. We also substitute the real values in the covariance formulas (denoted by RC) for every pair of α_0's and β_0's. As a standard, the Monte Carlo covariance is also computed (denoted by MCC) for every pair of α_0's and β_0's. The results of pairs involving α_0's, β_0's, and α_0's-β_0's are tabulated in an Excel file, which is available at `https://github.com/yiorfun/UBCovEst`.

The results in the tables show that the covariances evaluated at the true values are almost identical to the empirical ones computed by the Monte Carlo method. In addition, given a relatively small sample size, say $n = 50$, the proposed method may slightly overestimate the covariances for some pairs of α_0's and β_0's. It may be because the number of unknown parameters (i.e., $q = 20$) is much close to $n = 50$. As the sample size is larger than 50, the proposed method is applicable to provide reasonable covariance estimation.

Evaluation of the estimated covariance matrix estimator on multiple testing. We also conduct an extra simulation study to investigate the advantage of a covariance matrix estimate with less bias. In other words, a covariance matrix estimate with a larger bias might produce an inappropriate type 1 error or cause a loss in statistical power. Specifically, we set $p_{\text{ind}} = 100$ (therefore $p = 500$) and generate a sample with $n = 50$ from $N\left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_{0,1}\left(\mathbf{A}_0, \mathbf{B}_0, \boldsymbol{p}_1\right)\right)$, where $\boldsymbol{\mu}_0 = (\mu_{01}, \ldots, \mu_{0p})^\top$, $\mu_{0j} = 0.3$ for $j = 1, \ldots, 10$ and $\mu_{0j} = 0$ otherwise, and the other settings are the same as those in Scenario 1. To simultaneously perform hypothesis tests

$$H_{0j} : \mu_j = 0, \quad \text{against} \quad H_{1j} : \mu_j \neq 0, \quad j = 1, \ldots, p,$$

we apply the principal factor approximation (PFA) method to estimate the false discovery proportion (FDP) due to the presence of the dependence among test statistics (Fan et al., 2012;

Fan and Han, 2017). Among 100 Monte Carlo replicates, given a sequence of threshold values t, we calculate the medians and standard errors (s.d.) of the total number of rejections $R(t)$, the number of correct rejections $S(t)$, and the false discovery proportion $\mathrm{FDP}(t)$ using the true covariance matrix $\boldsymbol{\Sigma}_{0,1}\left(\mathbf{A}_0, \mathbf{B}_0, \boldsymbol{p}_1\right)$, its POET estimator $\widetilde{\boldsymbol{\Sigma}}_{\text{POET}}$, and the proposed estimator $\widetilde{\boldsymbol{\Sigma}}_{\text{prop.}} = \widetilde{\boldsymbol{\Sigma}}_1\left(\widetilde{\mathbf{A}}_1, \widetilde{\mathbf{B}}_1, \boldsymbol{p}_1\right)$, respectively.

The results presented in Table B.1 demonstrate that the proposed covariance matrix estimator behaves more similarly with the true covariance matrix than the POET estimator due to the less bias. For example, given a threshold value $t = 0.0527$, the approximate $\mathrm{FDP}(t)$ are 0.00, 0.79, and 0.13 for the truth, the POET estimator, and the proposed estimator, respectively. It implies the bias of the POET estimator might lead to an inappropriate PDP. Also, $S(t)$ is estimated by 10, 4, and 10 for the truth, the POET estimator, and the proposed estimator, respectively, which implies that there is a loss in statistical power for the POET estimator.

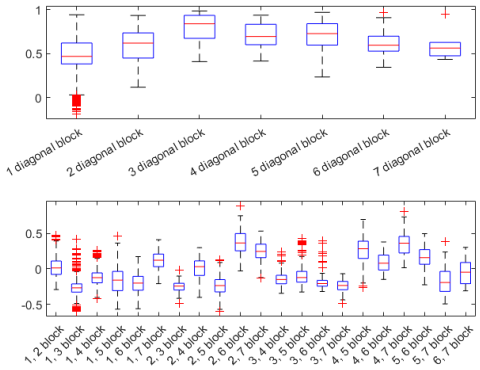
multicolumn1l	multicolumn6ctrue	multicolumn6cPOET	multicolumn6cproposed method															
threshold	median	s.d.	median	s.d.	median	s.d.	median	s.d.	median	s.d.	median	s.d.	median	s.d.	median	s.d.	median	s.d.
t	$R(t)$	$R(t)$	$S(t)$	$S(t)$	$\mathrm{FDP}(t)$	$\mathrm{FDP}(t)$	$R(t)$	$R(t)$	$S(t)$	$S(t)$	$\mathrm{FDP}(t)$	$\mathrm{FDP}(t)$	$R(t)$	$R(t)$	$S(t)$	$S(t)$	$\mathrm{FDP}(t)$	$\mathrm{FDP}(t)$
0.0001	0.00	0.00	10.00	0.00	0.00	0.00	0.00	0.00	0.00	1.90	0.00	0.00	0.00	0.00	10.00	4.23	0.00	0.00
0.0054	0.00	3.47	10.00	0.00	0.00	0.22	0.00	11.40	0.00	3.21	0.00	0.31	0.00	10.97	10.00	3.18	0.00	0.23
0.0106	0.00	11.96	10.00	0.00	0.00	0.31	0.00	15.95	0.00	3.42	0.00	0.38	0.00	14.68	10.00	2.81	0.00	0.34
0.0159	0.00	20.02	10.00	0.00	0.00	0.37	0.00	24.75	0.00	3.70	0.00	0.41	0.00	22.95	10.00	2.76	0.00	0.39
0.0211	0.00	25.19	10.00	0.00	0.00	0.40	0.00	30.70	0.00	3.78	0.00	0.45	0.00	30.01	10.00	2.70	0.00	0.40
0.0264	0.00	30.84	10.00	0.00	0.00	0.43	0.00	35.79	0.00	3.96	0.00	0.47	0.00	34.22	10.00	2.46	0.00	0.42
0.0316	0.00	36.12	10.00	0.00	0.00	0.44	0.00	39.71	0.00	3.99	0.00	0.47	0.00	38.52	10.00	2.41	0.00	0.43
0.0369	0.00	40.75	10.00	0.00	0.00	0.45	0.00	44.15	0.00	4.19	0.00	0.47	0.00	42.58	10.00	2.36	0.00	0.44
0.0422	0.00	45.65	10.00	0.00	0.00	0.45	0.00	49.61	1.50	4.29	0.00	0.48	0.00	47.61	10.00	2.18	0.00	0.45
0.0474	0.00	49.77	10.00	0.00	0.00	0.46	0.50	55.04	2.00	4.43	0.03	0.48	0.00	53.13	10.00	2.09	0.00	0.45
0.0527	1.00	52.49	10.00	0.00	0.00	0.46	1.50	59.48	4.00	4.52	0.79	0.48	1.00	57.79	10.00	2.04	0.13	0.46
0.0579	1.00	55.14	10.00	0.00	0.00	0.46	2.00	63.10	5.00	4.61	0.85	0.48	2.00	61.26	10.00	1.97	0.42	0.46
0.0632	4.00	57.63	10.00	0.00	0.59	0.46	4.00	65.92	6.50	4.65	0.87	0.47	3.50	64.55	10.00	1.96	0.44	0.46
0.0685	8.00	59.61	10.00	0.00	0.64	0.46	8.00	67.90	7.00	4.63	0.97	0.46	5.50	66.84	10.00	1.96	0.60	0.45
0.0737	10.00	62.86	10.00	0.00	0.74	0.46	10.00	69.55	7.50	4.61	1.00	0.45	8.00	68.44	10.00	1.96	0.73	0.46
0.0790	10.00	65.62	10.00	0.00	0.74	0.45	12.00	71.26	8.00	4.45	1.00	0.44	10.00	70.06	10.00	1.96	0.84	0.45
0.0842	10.50	68.02	10.00	0.00	0.80	0.45	13.00	72.40	9.00	4.37	1.00	0.43	12.00	71.64	10.00	1.94	0.88	0.44
0.0895	10.50	70.18	10.00	0.00	0.87	0.45	16.00	73.83	10.00	4.26	1.00	0.43	13.00	72.85	10.00	1.94	0.82	0.44
0.0947	14.00	71.77	10.00	0.00	0.88	0.44	21.50	75.19	10.00	4.16	1.00	0.42	18.00	74.40	10.00	1.94	0.79	0.43
0.1000	20.00	72.97	10.00	0.00	0.88	0.44	27.00	77.10	10.00	3.95	1.00	0.40	22.00	75.90	10.00	1.93	0.83	0.43

Table B.1: Estimation results of the total number of rejections $R(t)$, the number of correct rejections $S(t)$, and the false discovery proportion (FDP) using the true covariance matrix, the POET estimator, and the proposed estimator

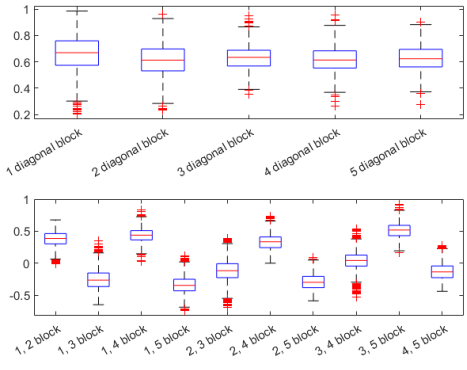
B.4.2 Extra simulation study in Scenario 3

Justification of the proposed parameterization strategy. To justify the proposed parameterization for the structure of uniform blocks, we compare the box plots of the off-diagonal entries in the diagonal blocks and those of all entries in the off-diagonal blocks based on the real data examples, to those based on the simulated datasets in Scenario 3 (Section 3.3.4). The box plots are presented below, which demonstrate that although the true covariance matrix $\boldsymbol{\Upsilon}_{0,\sigma}$ is not a UB matrix, the proposed estimation procedure can yield a more competitive estimate for the covariance matrix and a not-worse estimate for the precision matrix than the conventional methods do, if the variations in blocks are small. Therefore, these results show that it is reasonable to consider the structure of uniform blocks for the real datasets because the variations in blocks are close to those in the simulated data with $\sigma = 0.8$ for the proteomics study and the brain imaging study, respectively.

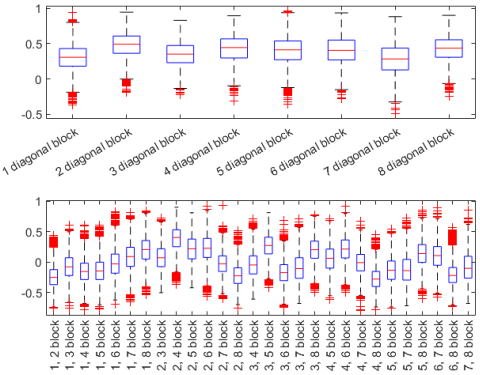
To conduct simulations in Section 3.3 and in Section B.4, we load the R packages "pfa" (Fan et al., 2012) (with some modifications), "CovTools" (Lee et al., 2021), "POET" (Fan et al., 2016), "CVTuningCov" (Wang, 2015) (with some modifications), and implement the estimation procedures by using the software R (R Core Team, 2021). R code for the numerical studies in Section 3.3 and the data examples in Section B.4 are available at `https://github.com/yiorfun/UBCovEst`.



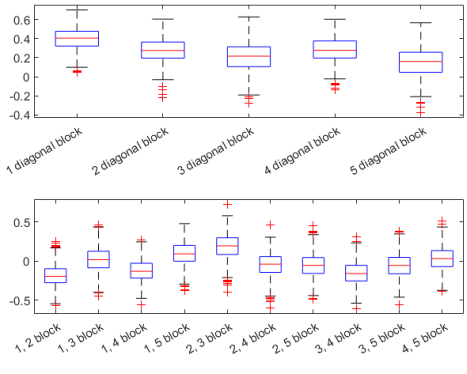
(a) Box plots for the proteomics dataset



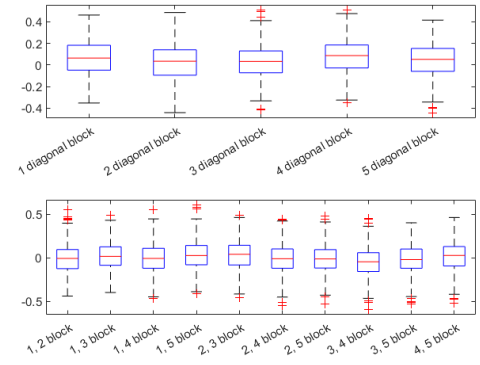
(b) Box plots for the brain imaging dataset



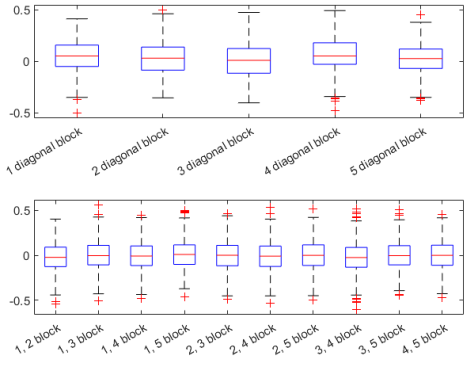
(c) Box plots for the Spellman dataset



(d) Box plots for the simulated dataset with $\sigma = 0.1$ in Scenario 3



(e) Boxplots for the simulated dataset with $\sigma = 0.5$ in Scenario 3



(f) Boxplots for the simulated dataset with $\sigma = 0.8$ in Scenario 3

Figure B.2: Box plots of the off-diagonal entries in the diagonal blocks and of all entries in the diagonal blocks.

Appendix C: Supplementary Materials for Chapter 4

C.1 Definition

Definition C.1.1 (consistent estimator of a covariance matrix, also see Definition 8.2.1 in Fomby et al. (1984) or page 68 on Schmidt (2020))**.** *If a covariance matrix* $\boldsymbol{\Sigma}$ *depends on a finite number of parameters* $\theta_1, \ldots, \theta_p$*, and if* $\widehat{\boldsymbol{\Sigma}}$ *depends on consistent estimators* $\widehat{\theta}_1, \ldots, \widehat{\theta}_p$*, then* $\widehat{\boldsymbol{\Sigma}}$ *is said to be a consistent estimator of* $\boldsymbol{\Sigma}$*.*

C.2 Properties of UB-Matrices

Corollary C.2.1 (square-inverse transformations)**.** *Let* $\boldsymbol{\ell} = (L_1, \ldots, L_G)^\top$ *be a partition-size vector satisfying that* $R = L_1 + \cdots + L_G$ *and* $L_g > 1$ *for all* g*, and* $\boldsymbol{N}_1, \boldsymbol{N}_2 \in \mathbb{R}^{R \times R}$ *be two matrices satisfying the square-inverse relationship, i.e.,* $\boldsymbol{N}_2 = (\boldsymbol{I}_R - \boldsymbol{N}_1)^{-1} (\boldsymbol{I}_R - \boldsymbol{N}_1)^{-1}$*.*

(1) If $\boldsymbol{N}_1 = \boldsymbol{N}_1 (\boldsymbol{A}_1, \boldsymbol{B}_1, \boldsymbol{\ell})$ *is a UB matrix with diagonal matrix* $\boldsymbol{A}_1$ *and symmetric matrix* $\boldsymbol{B}_1$*, then* $\boldsymbol{N}_2$ *is a UB matrix, expressed by* $\boldsymbol{N}_2 (\boldsymbol{A}_2, \boldsymbol{B}_2, \boldsymbol{\ell})$*. Specifically, let* $\boldsymbol{L} = \operatorname{diag}(L_1, \ldots, L_G)$*,*

$\boldsymbol{A}_2$ and $\boldsymbol{B}_2$ can be calculated using $\boldsymbol{A}_1$, $\boldsymbol{B}_1$, and $\boldsymbol{L}$ as below:

$$\boldsymbol{I}_R - \boldsymbol{N}_1 = \boldsymbol{N}^* \left(\boldsymbol{A}^*, \boldsymbol{B}^*, \boldsymbol{\ell}\right), \quad \boldsymbol{A}^* = \boldsymbol{I}_G - \boldsymbol{A}_1, \quad \boldsymbol{B}^* = -\boldsymbol{B}_1, \quad \boldsymbol{\Delta}^* = \boldsymbol{I}_G - \boldsymbol{A}_1 - \boldsymbol{B}_1\boldsymbol{L};$$

$$\left(\boldsymbol{I}_R - \boldsymbol{N}_1\right)^{-1} = \boldsymbol{N}^\star \left(\boldsymbol{A}^\star, \boldsymbol{B}^\star, \boldsymbol{\ell}\right), \quad \boldsymbol{A}^\star = \boldsymbol{A}^{*,-1}, \quad \boldsymbol{B}^\star = -\boldsymbol{\Delta}^{*,-1}\boldsymbol{B}^*\boldsymbol{A}^{*,-1};$$

$$\left(\boldsymbol{I}_R - \boldsymbol{N}_1\right)^{-1}\left(\boldsymbol{I}_R - \boldsymbol{N}_1\right)^{-1} = \boldsymbol{N}_2 \left(\boldsymbol{A}_2, \boldsymbol{B}_2, \boldsymbol{\ell}\right), \quad \boldsymbol{A}_2 = \boldsymbol{A}^{\star,2}, \quad \boldsymbol{B}_2 = \boldsymbol{A}^\star\boldsymbol{B}^\star + \boldsymbol{B}^\star\boldsymbol{A}^\star + \boldsymbol{B}^\star\boldsymbol{L}\boldsymbol{B}^\star.$$

(2) If $\boldsymbol{N}_2 = \boldsymbol{N}_2 \left(\boldsymbol{A}_2, \boldsymbol{B}_2, \boldsymbol{\ell}\right)$ is a UB matrix with diagonal matrix $\boldsymbol{A}_2$ and symmetric matrix $\boldsymbol{B}_2$, then $\boldsymbol{N}_1$ is a UB matrix, expressed by $\boldsymbol{N}_1 \left(\boldsymbol{A}_1, \boldsymbol{B}_1, \boldsymbol{\ell}\right)$. Specifically, let $\boldsymbol{L} = \mathrm{diag}\left(L_1, \ldots, L_G\right)$, $\boldsymbol{A}_1$ and $\boldsymbol{B}_1$ can be calculated using $\boldsymbol{A}_2$, $\boldsymbol{B}_2$, $\boldsymbol{L}$, and $\boldsymbol{\Delta}_2 = \boldsymbol{A}_2 + \boldsymbol{B}_2\boldsymbol{L}$ as below:

$$\boldsymbol{N}_2^{-1} = \boldsymbol{N}^\dagger \left(\boldsymbol{A}^\dagger, \boldsymbol{B}^\dagger, \boldsymbol{\ell}\right), \quad \boldsymbol{A}^\dagger = \boldsymbol{A}_2^{-1}, \quad \boldsymbol{B}^\dagger = -\boldsymbol{\Delta}_2^{-1}\boldsymbol{B}_2\boldsymbol{A}_2^{-1};$$

$\boldsymbol{I}_R - \boldsymbol{N}_1 = \boldsymbol{N}^* \left(\boldsymbol{A}^*, \boldsymbol{B}^*, \boldsymbol{\ell}\right)$, $\boldsymbol{A}^* = \boldsymbol{A}^{\dagger,\frac{1}{2}}$, $\boldsymbol{B}^*$ *is the solution to the algebraic Riccati equation:*

$$\boldsymbol{A}^* \times \boldsymbol{B}^* + \boldsymbol{B}^* \times \boldsymbol{A}^* + \boldsymbol{B}^* \times \boldsymbol{L} \times \boldsymbol{B}^* - \boldsymbol{B}^\dagger = \boldsymbol{0}_{G\times G};$$

$$\boldsymbol{N}_1 = \boldsymbol{N}_1 \left(\boldsymbol{A}_1, \boldsymbol{B}_1, \boldsymbol{\ell}\right), \quad \boldsymbol{A}_1 = \boldsymbol{I}_G - \boldsymbol{A}^*, \quad \boldsymbol{B}_1 = -\boldsymbol{B}^*.$$

C.3 Plug-In Estimators

The plug-in matrix estimators are given by

$$\widehat{\boldsymbol{\Upsilon}}\left(\widehat{\mathbf{A}}_{\boldsymbol{\Upsilon}},\widehat{\mathbf{B}}_{\boldsymbol{\Upsilon}},\boldsymbol{\ell}\right)=\widehat{\mathbf{A}}_{\boldsymbol{\Upsilon}}\circ\mathbf{I}(\boldsymbol{\ell})+\widehat{\mathbf{B}}_{\boldsymbol{\Upsilon}}\circ\mathbf{J}(\boldsymbol{\ell}),\ \text{with}\ \begin{cases}\widehat{\mathbf{A}}_{\boldsymbol{\Upsilon}} &= \operatorname{diag}\left(-\widehat{\gamma}_{11},\ldots,-\widehat{\gamma}_{GG}\right)\\ \widehat{\mathbf{B}}_{\boldsymbol{\Upsilon}} &= \left(\widehat{\gamma}_{gg'}\right)\end{cases}, \tag{C.3.1}$$

$$\widehat{\boldsymbol{\Omega}}\left(\widehat{\mathbf{A}}_{\boldsymbol{\Omega}},\widehat{\mathbf{B}}_{\boldsymbol{\Omega}},\boldsymbol{\ell}\right)=\widehat{\mathbf{A}}_{\boldsymbol{\Omega}}\circ\mathbf{I}(\boldsymbol{\ell})+\widehat{\mathbf{B}}_{\boldsymbol{\Omega}}\circ\mathbf{J}(\boldsymbol{\ell}),\ \text{with}\ \begin{cases}\widehat{\mathbf{A}}_{\boldsymbol{\Omega}} &= \left(\mathbf{I}_G-\widehat{\mathbf{A}}_{\boldsymbol{\Upsilon}}\right)^2\\ \widehat{\mathbf{B}}_{\boldsymbol{\Omega}} &= -2\widehat{\mathbf{B}}_{\boldsymbol{\Upsilon}}+\widehat{\mathbf{A}}_{\boldsymbol{\Upsilon}}\widehat{\mathbf{B}}_{\boldsymbol{\Upsilon}}+\widehat{\mathbf{B}}_{\boldsymbol{\Upsilon}}\widehat{\mathbf{A}}_{\boldsymbol{\Upsilon}}+\widehat{\mathbf{B}}_{\boldsymbol{\Upsilon}}\mathbf{L}\widehat{\mathbf{B}}_{\boldsymbol{\Upsilon}}\end{cases}, \tag{C.3.2}$$

$$\widehat{\boldsymbol{\Sigma}}\left(\widehat{\mathbf{A}}_{\boldsymbol{\Sigma}},\widehat{\mathbf{B}}_{\boldsymbol{\Sigma}},\boldsymbol{\ell}\right)=\widehat{\mathbf{A}}_{\boldsymbol{\Sigma}}\circ\mathbf{I}(\boldsymbol{\ell})+\widehat{\mathbf{B}}_{\boldsymbol{\Sigma}}\circ\mathbf{J}(\boldsymbol{\ell}),\ \text{with}\ \begin{cases}\widehat{\mathbf{A}}_{\boldsymbol{\Sigma}} &= \widehat{\mathbf{A}}_{\boldsymbol{\Omega}}^{-1}\\ \widehat{\mathbf{B}}_{\boldsymbol{\Sigma}} &= -\widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Omega}}^{-1}\widehat{\mathbf{B}}_{\boldsymbol{\Omega}}\widehat{\mathbf{A}}_{\boldsymbol{\Omega}}^{-1}\end{cases}, \tag{C.3.3}$$

where we assume $\widehat{\mathbf{A}}_{\boldsymbol{\Omega}}\succ 0$, $\widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Omega}}=\widehat{\mathbf{A}}_{\boldsymbol{\Omega}}+\widehat{\mathbf{B}}_{\boldsymbol{\Omega}}\mathbf{L}$ has positive eigenvalues only.

By Theorem 4.2.2, we note that the matrix estimators $\widehat{\boldsymbol{\Upsilon}}\left(\widehat{\mathbf{A}}_{\boldsymbol{\Upsilon}},\widehat{\mathbf{B}}_{\boldsymbol{\Upsilon}},\boldsymbol{\ell}\right)$ in (C.3.1), $\widehat{\boldsymbol{\Omega}}\left(\widehat{\mathbf{A}}_{\boldsymbol{\Omega}},\widehat{\mathbf{B}}_{\boldsymbol{\Omega}},\boldsymbol{\ell}\right)$ in (C.3.2), and $\widehat{\boldsymbol{\Sigma}}\left(\widehat{\mathbf{A}}_{\boldsymbol{\Sigma}},\widehat{\mathbf{B}}_{\boldsymbol{\Sigma}},\boldsymbol{\ell}\right)$ in (C.3.3) are consistent estimators in the sense of Definition C.1.1.

C.4 Technical Conditions and Proofs

Condition 1. *Covariance matrix* $\boldsymbol{\Sigma}\left(\boldsymbol{A}_{\boldsymbol{\Sigma}},\boldsymbol{B}_{\boldsymbol{\Sigma}},\boldsymbol{\ell}\right)$ *and the proposed estimator* $\widehat{\boldsymbol{\Sigma}}\left(\widehat{\boldsymbol{A}}_{\boldsymbol{\Sigma}},\widehat{\boldsymbol{B}}_{\boldsymbol{\Sigma}},\boldsymbol{\ell}\right)$ *are positive definite, or equivalently,* $\boldsymbol{A}_{\boldsymbol{\Sigma}},\widehat{\boldsymbol{A}}_{\boldsymbol{\Sigma}}\succ 0$ *and both* $\boldsymbol{\Delta}_{\boldsymbol{\Sigma}}=\boldsymbol{A}_{\boldsymbol{\Sigma}}+\boldsymbol{B}_{\boldsymbol{\Sigma}}\times\boldsymbol{L}$ *and* $\widehat{\boldsymbol{\Delta}}_{\boldsymbol{\Sigma}}=\widehat{\boldsymbol{A}}_{\boldsymbol{\Sigma}}+\widehat{\boldsymbol{B}}_{\boldsymbol{\Sigma}}\times\boldsymbol{L}$ *have positive eigenvalues only.*

Condition 2. *Covariate vectors* $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\in\mathbb{R}^p$ *are linear independent.*

Condition 3. *Sample size* $n > \max\{p, G(G+1)/2\}$.

Condition 4. *Unknown regression coefficient vector* $\boldsymbol{\beta}$ *does not depend on unknown scaled dependence parameter vector* $\boldsymbol{\gamma}$.

Condition 5. *Each element of* $n^{-1}\boldsymbol{\Phi}_{\boldsymbol{\beta}}$ *converges to a finite function with respect to* $\boldsymbol{\gamma}$*, as* n *goes to infinity, uniformly for* $\boldsymbol{\gamma}$ *in the compact set* Θ.

Condition 6. *Each diagonal element of*

$$n^{-2}\left\{\left[\frac{\partial \boldsymbol{\Omega}\left(\boldsymbol{A}_{\boldsymbol{\Omega}}, \boldsymbol{B}_{\boldsymbol{\Omega}}, \boldsymbol{\ell}\right)}{\partial \boldsymbol{\gamma}_j}\right] \boldsymbol{\Sigma}\left(\boldsymbol{A}_{\boldsymbol{\Sigma}}, \boldsymbol{B}_{\boldsymbol{\Sigma}}, \boldsymbol{\ell}\right)\left[\frac{\partial \boldsymbol{\Omega}\left(\boldsymbol{A}_{\boldsymbol{\Omega}}, \boldsymbol{B}_{\boldsymbol{\Omega}}, \boldsymbol{\ell}\right)}{\partial \boldsymbol{\gamma}_j}\right]\right\} \otimes\left(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i^{\top}\right)$$

converges to 0*, as* n *goes to infinity, uniformly for* $\boldsymbol{\gamma}$ *in the compact set* Θ*, for all* j.

Proof of Corollary C.2.1. The requirement for a unique solution to the algebraic Riccati equation can be found in Ran and Rodman (1984) and Abou-Kandil et al. (2003). ■

Proof of Corollay 4.2.1. Using the result in Corollary C.2.1,the proof is straightforward. ■

Proofs of Theorem 4.2.1, Theorem 4.2.2, and Theorem 4.2.3. We check that Conditions 1, 2, 3, 4 satisfy Assumptions 1-4 in Magnus (1978). In particular, Conditions 2 and 3 imply the design matrix $\mathbf{x} \in \mathbb{R}^{(nR)\times(Rp)}$ has a full rank; and the partial derivatives (C.4.1) and (C.4.2) imply that each element of the matrix $\boldsymbol{\Sigma}\left(\mathbf{A}_{\boldsymbol{\Sigma}}, \mathbf{B}_{\boldsymbol{\Sigma}}, \boldsymbol{\ell}\right)$ are twice differentiable function with respect to $\boldsymbol{\gamma}_j$, where $\boldsymbol{\gamma}_j$ belonging to $\{\gamma_{11}, \ldots, \gamma_{1G}, \ldots, \gamma_{2G}, \ldots, \gamma_{GG}\}$ denotes the j-th component of $\boldsymbol{\gamma}$ for $j = 1, \ldots, G(G+1)/2$. Therefore, the properties hold by Theorem 1 in Magnus (1978).

Then, we start with the following equality

$$\mathbf{x}_{(Rp)\times(nR)}^{\top}\left(\mathbf{I}_n \otimes \mathbf{N}_{R\times R}\right)_{(nR)\times(nR)} \mathbf{x}_{(nR)\times(Rp)} = \mathbf{N} \otimes \left(\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\top}\right)$$

for $\mathbf{N} \in \mathbb{R}^{R\times R}$. Taking $\mathbf{N} = \mathbf{I}_R$, we obtain that an alternative form of $\mathbf{x}^\top\mathbf{x}$, i.e., $\mathbf{x}^\top\mathbf{x} = \mathbf{I}_R \otimes \left(\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top\right)$. By Condition 2, $\left(\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top\right)$ is invertible, so is $\mathbf{x}^\top\mathbf{x}$. Suppose $\mathbf{N}$ is invertible, then

$$\begin{aligned}\left(\mathbf{x}^\top\mathbf{x}\right)^{-1}\mathbf{x}^\top\left(\mathbf{I}_n\otimes\mathbf{N}\right)\mathbf{x}\left(\mathbf{x}^\top\mathbf{x}\right)^{-1} &= \left[\mathbf{I}_R\otimes\left(\sum_{i=1}^n\mathbf{x}_i\mathbf{x}_i^\top\right)^{-1}\right]\left[\mathbf{N}\otimes\left(\sum_{i=1}^n\mathbf{x}_i\mathbf{x}_i^\top\right)\right]\left[\mathbf{I}_R\otimes\left(\sum_{i=1}^n\mathbf{x}_i\mathbf{x}_i^\top\right)^{-1}\right]\\ &= \mathbf{N}\otimes\left(\sum_{i=1}^n\mathbf{x}_i\mathbf{x}_i^\top\right)^{-1} = \left[\mathbf{N}^{-1}\otimes\left(\sum_{i=1}^n\mathbf{x}_i\mathbf{x}_i^\top\right)\right]^{-1} = \left[\mathbf{x}^\top\left(\mathbf{I}_n\otimes\mathbf{N}\right)^{-1}\mathbf{x}\right]^{-1}.\end{aligned}$$

Following Condition 1 and Theorem 1(A) in Lu and Schmidt (2012), we replace $\mathbf{N} = \boldsymbol{\Sigma}\left(\mathbf{A}_{\boldsymbol{\Sigma}}, \mathbf{B}_{\boldsymbol{\Sigma}}, \boldsymbol{\ell}\right)$ and $\mathbf{N} = \widehat{\boldsymbol{\Sigma}}\left(\widehat{\mathbf{A}}_{\boldsymbol{\Sigma}}, \widehat{\mathbf{B}}_{\boldsymbol{\Sigma}}, \boldsymbol{\ell}\right)$, respectively, and complete the proof of the equality of the OLS estimator and the FGLS estimator (see more discussions in Puntanen and Styan (1989) and a case of $G = 1$ in He and Wang (2022)). Under the normality assumption for the MAUD, we obtain the normal distribution for $\widehat{\boldsymbol{\beta}}$.

We follow the lines of arguments in the proof of Theorem 3 in Magnus (1978). Specifically, let $\boldsymbol{\gamma}_j \in \{\gamma_{11}, \ldots, \gamma_{1G}, \ldots, \gamma_{2G}, \ldots, \gamma_{GG}\}$ denote the j-th component of $\boldsymbol{\gamma}$ for $j = 1, \ldots, G(G+1)/2$. Then, $\boldsymbol{\Phi}_{\boldsymbol{\gamma}} = \left(\psi_{jj'}^{(\boldsymbol{\gamma})}\right)$ and $\psi_{jj'}^{(\boldsymbol{\gamma})}$ is given by

$$\begin{aligned}\psi_{jj'}^{(\boldsymbol{\gamma})} &= \frac{1}{2}\operatorname{tr}\left\{\left[\frac{\partial\left(\mathbf{I}_n\otimes\boldsymbol{\Sigma}\left(\mathbf{A}_{\boldsymbol{\Sigma}},\mathbf{B}_{\boldsymbol{\Sigma}},\boldsymbol{\ell}\right)\right)^{-1}}{\partial\boldsymbol{\gamma}_j}\right]\right.\\ &\quad\left.\times\left(\mathbf{I}_n\otimes\boldsymbol{\Sigma}\left(\mathbf{A}_{\boldsymbol{\Sigma}},\mathbf{B}_{\boldsymbol{\Sigma}},\boldsymbol{\ell}\right)\right)\left[\frac{\partial\left(\mathbf{I}_n\otimes\boldsymbol{\Sigma}\left(\mathbf{A}_{\boldsymbol{\Sigma}},\mathbf{B}_{\boldsymbol{\Sigma}},\boldsymbol{\ell}\right)\right)^{-1}}{\partial\boldsymbol{\gamma}_{j'}}\right]\left(\mathbf{I}_n\otimes\boldsymbol{\Sigma}\left(\mathbf{A}_{\boldsymbol{\Sigma}},\mathbf{B}_{\boldsymbol{\Sigma}},\boldsymbol{\ell}\right)\right)\right\}\\ &= \frac{n}{2}\operatorname{tr}\left\{\left[\frac{\partial\boldsymbol{\Omega}\left(\mathbf{A}_{\boldsymbol{\Omega}},\mathbf{B}_{\boldsymbol{\Omega}},\boldsymbol{\ell}\right)}{\partial\boldsymbol{\gamma}_j}\right]\boldsymbol{\Sigma}\left(\mathbf{A}_{\boldsymbol{\Sigma}},\mathbf{B}_{\boldsymbol{\Sigma}},\boldsymbol{\ell}\right)\left[\frac{\partial\boldsymbol{\Omega}\left(\mathbf{A}_{\boldsymbol{\Omega}},\mathbf{B}_{\boldsymbol{\Omega}},\boldsymbol{\ell}\right)}{\partial\boldsymbol{\gamma}_{j'}}\right]\boldsymbol{\Sigma}\left(\mathbf{A}_{\boldsymbol{\Sigma}},\mathbf{B}_{\boldsymbol{\Sigma}},\boldsymbol{\ell}\right)\right\}\end{aligned}$$

for $j, j' = 1, \ldots, G(G+1)/2$, where $\dfrac{\partial\boldsymbol{\Omega}\left(\mathbf{A}_{\boldsymbol{\Omega}},\mathbf{B}_{\boldsymbol{\Omega}},\boldsymbol{\ell}\right)}{\partial\boldsymbol{\gamma}_j} = \left(\dfrac{\partial\mathbf{A}_{\boldsymbol{\Omega}}}{\partial\boldsymbol{\gamma}_j}\right)\circ\mathbf{I}(\boldsymbol{\ell}) + \left(\dfrac{\partial\mathbf{B}_{\boldsymbol{\Omega}}}{\partial\boldsymbol{\gamma}_j}\right)\circ\mathbf{J}(\boldsymbol{\ell})$. Recall

the following "coordinate matrices",

$$\begin{aligned}
&\mathbf{A}_{\boldsymbol{\Upsilon}} = \operatorname{diag}\left(-\gamma_{11}, \ldots, -\gamma_{GG}\right), \quad && \mathbf{B}_{\boldsymbol{\Upsilon}} = \left(\gamma_{gg'}\right); \\
&\mathbf{A}_{\boldsymbol{\Omega}} = \left(\mathbf{I}_G - \mathbf{A}_{\boldsymbol{\Upsilon}}\right)^2, && \mathbf{B}_{\boldsymbol{\Omega}} = -2\mathbf{B}_{\boldsymbol{\Upsilon}} + \mathbf{A}_{\boldsymbol{\Upsilon}}\mathbf{B}_{\boldsymbol{\Upsilon}} + \mathbf{B}_{\boldsymbol{\Upsilon}}\mathbf{A}_{\boldsymbol{\Upsilon}} + \mathbf{B}_{\boldsymbol{\Upsilon}}\mathbf{L}\mathbf{B}_{\boldsymbol{\Upsilon}}; \\
&\mathbf{A}_{\boldsymbol{\Sigma}} = \mathbf{A}_{\boldsymbol{\Omega}}^{-1}, && \mathbf{B}_{\boldsymbol{\Sigma}} = -\boldsymbol{\Delta}_{\boldsymbol{\Omega}}^{-1}\mathbf{B}_{\boldsymbol{\Omega}}\mathbf{A}_{\boldsymbol{\Omega}}^{-1};
\end{aligned}$$

and let $\mathbf{E}_{gg'} \in \mathbb{R}^{G\times G}$ denote a matrix whose (g, g')-th element is 1 and the others are 0.

If $\boldsymbol{\gamma}_j = \gamma_{gg}$ for some g, then $\dfrac{\partial \boldsymbol{\Omega}\left(\mathbf{A}_{\boldsymbol{\Omega}}, \mathbf{B}_{\boldsymbol{\Omega}}, \boldsymbol{\ell}\right)}{\partial \boldsymbol{\gamma}_j} = \left(\dfrac{\partial \mathbf{A}_{\boldsymbol{\Omega}}}{\partial \boldsymbol{\gamma}_j}\right) \circ \mathbf{I}(\boldsymbol{\ell}) + \left(\dfrac{\partial \mathbf{B}_{\boldsymbol{\Omega}}}{\partial \boldsymbol{\gamma}_j}\right) \circ \mathbf{J}(\boldsymbol{\ell})$, and

$$\begin{aligned}
\frac{\partial \mathbf{A}_{\boldsymbol{\Omega}}}{\partial \boldsymbol{\gamma}_j} &= 2(1+\gamma_{gg})\mathbf{E}_{gg}, \\
\frac{\partial \mathbf{B}_{\boldsymbol{\Omega}}}{\partial \boldsymbol{\gamma}_j} &= -2(1+\gamma_{gg})\mathbf{E}_{gg} - \left(\mathbf{E}_{gg}\mathbf{B}_{\boldsymbol{\Upsilon}} + \mathbf{B}_{\boldsymbol{\Upsilon}}\mathbf{E}_{gg}\right) + \mathbf{E}_{gg}\mathbf{L}\mathbf{B}_{\boldsymbol{\Upsilon}} + \mathbf{B}_{\boldsymbol{\Upsilon}}\mathbf{L}\mathbf{E}_{gg}.
\end{aligned} \tag{C.4.1}$$

If $\boldsymbol{\gamma}_j = \gamma_{gg'}$ for some $g \neq g'$, then $\dfrac{\partial \boldsymbol{\Omega}\left(\mathbf{A}_{\boldsymbol{\Omega}}, \mathbf{B}_{\boldsymbol{\Omega}}, \boldsymbol{\ell}\right)}{\partial \boldsymbol{\gamma}_j} = \mathbf{0}_{G\times G} \circ \mathbf{I}(\boldsymbol{\ell}) + \left(\dfrac{\partial \mathbf{B}_{\boldsymbol{\Omega}}}{\partial \boldsymbol{\gamma}_j}\right) \circ \mathbf{J}(\boldsymbol{\ell})$, and

$$\begin{aligned}
\frac{\partial \mathbf{A}_{\boldsymbol{\Omega}}}{\partial \boldsymbol{\gamma}_j} &= \mathbf{0}_{G\times G}, \\
\frac{\partial \mathbf{B}_{\boldsymbol{\Omega}}}{\partial \boldsymbol{\gamma}_j} &= -2\left(\mathbf{E}_{gg'} + \mathbf{E}_{g'g}\right) + \left(\mathbf{E}_{gg'} + \mathbf{E}_{g'g}\right)\left(\mathbf{A}_{\boldsymbol{\Upsilon}} + \mathbf{L}\mathbf{B}_{\boldsymbol{\Upsilon}}\right) + \left(\mathbf{A}_{\boldsymbol{\Upsilon}} + \mathbf{B}_{\boldsymbol{\Upsilon}}\mathbf{L}\right)\left(\mathbf{E}_{gg'} + \mathbf{E}_{g'g}\right).
\end{aligned} \tag{C.4.2}$$

Due to $\mathbf{E}_{gg'}$ for $g, g' = 1, \ldots, G$, $\operatorname{vec}\left[\partial \boldsymbol{\Omega}\left(\mathbf{A}_{\boldsymbol{\Omega}}, \mathbf{B}_{\boldsymbol{\Omega}}, \boldsymbol{\ell}\right)/\partial \boldsymbol{\gamma}_j\right]$ for $j = 1, \ldots, G(G+1)/2$ are linearly independent, yielding Assumption 5 in Magnus (1978) is satisfied. Thus, $\boldsymbol{\Psi} \succ 0$ follows the result of Lemma 1 in Magnus (1978).

Now, we follow the lines of the proof of Theorem 5 in Magnus (1978). Given the first-order partial derivatives in closed form, we can obtain the second-order partial derivatives in closed form, which is omitted here. So, Assumption 10 and Assumption 11 in Magnus (1978)

are satisfied because both the following facts: $2n^{-2}\psi_{jj'}^{(\gamma)}$ converges to a finite function with respect to $\boldsymbol{\gamma}$, as n goes to infinity, uniformly for $\boldsymbol{\gamma}$ in the compact set Θ, for all j and j'; and $n^{-1}\operatorname{tr}\left\{\left[\partial^2\boldsymbol{\Omega}\left(\mathbf{A}_{\boldsymbol{\Omega}},\mathbf{B}_{\boldsymbol{\Omega}},\boldsymbol{\ell}\right)/\left(\partial\boldsymbol{\gamma}_j\partial\boldsymbol{\gamma}_{j'}\right)\right]\boldsymbol{\Sigma}\left(\mathbf{A}_{\boldsymbol{\Sigma}},\mathbf{B}_{\boldsymbol{\Sigma}},\boldsymbol{\ell}\right)\right\}^2$ converges to 0, as n goes to infinity, uniformly for $\boldsymbol{\gamma}$ in the compact set Θ, for all j and j'. Given the consistency, $\widehat{\boldsymbol{\gamma}}$ is the unique ML estimator due to the result of Lemma 2 in Magnus (1978). ■

C.5 The Case of $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} \neq \mathbf{I}_R$

We set a new parametric covariance matrix

$$\boldsymbol{\Sigma} = \left(\mathbf{I}_R - \boldsymbol{\Upsilon}\right)^{-1}\boldsymbol{\Sigma}_{\epsilon}\left(\mathbf{I}_R - \boldsymbol{\Upsilon}\right)^{-1},$$

where $\boldsymbol{\Upsilon} = \boldsymbol{\Upsilon}\left(\mathbf{A}_{\boldsymbol{\Upsilon}},\mathbf{B}_{\boldsymbol{\Upsilon}},\boldsymbol{\ell}\right)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}} = \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}\left(\mathbf{A}_{\boldsymbol{\epsilon}},\mathbf{B}_{\boldsymbol{\epsilon}},\boldsymbol{\ell}\right)$ with $\mathbf{A}_{\boldsymbol{\Upsilon}} = \operatorname{diag}\left(-\gamma_{11},\ldots,-\gamma_{GG}\right)$, $\mathbf{B}_{\boldsymbol{\Upsilon}} = \left(\gamma_{gg'}\right)$, $\gamma_{g'g} = \gamma_{gg'}$, $\mathbf{A}_{\boldsymbol{\epsilon}} = \operatorname{diag}\left(\omega_{11},\ldots,\omega_{GG}\right)$, and $\mathbf{B}_{\boldsymbol{\epsilon}} = \mathbf{0}_{G\times G}$.

Therefore, we can prove $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ are also uniform-block matrices.

$$\begin{aligned}
\boldsymbol{\Omega}\left(\mathbf{A}_{\boldsymbol{\Omega}},\mathbf{B}_{\boldsymbol{\Omega}},\boldsymbol{\ell}\right) &= \left(\mathbf{I}_R - \boldsymbol{\Upsilon}\right)\boldsymbol{\Sigma}_{\epsilon}^{-1}\left(\mathbf{I}_R - \boldsymbol{\Upsilon}\right);\\
\mathbf{A}_{\boldsymbol{\Omega}} &= \mathbf{A}_{\boldsymbol{\epsilon}}^{-1} - 2\mathbf{A}_{\boldsymbol{\Upsilon}}\mathbf{A}_{\boldsymbol{\epsilon}}^{-1} + \mathbf{A}_{\boldsymbol{\Upsilon}}^{2}\mathbf{A}_{\boldsymbol{\epsilon}}^{-1},\\
\mathbf{B}_{\boldsymbol{\Omega}} &= -\mathbf{A}_{\boldsymbol{\epsilon}}^{-1}\mathbf{B}_{\boldsymbol{\Upsilon}} + \mathbf{A}_{\boldsymbol{\Upsilon}}\mathbf{A}_{\boldsymbol{\epsilon}}^{-1}\mathbf{B}_{\boldsymbol{\Upsilon}} - \mathbf{B}_{\boldsymbol{\Upsilon}}\mathbf{A}_{\boldsymbol{\epsilon}}^{-1} + \mathbf{B}_{\boldsymbol{\Upsilon}}\mathbf{A}_{\boldsymbol{\epsilon}}^{-1}\mathbf{A}_{\boldsymbol{\Upsilon}} + \mathbf{B}_{\boldsymbol{\Upsilon}}\mathbf{A}_{\boldsymbol{\epsilon}}^{-1}\mathbf{L}\mathbf{B}_{\boldsymbol{\Upsilon}}.
\end{aligned}$$

We can observe that both $\mathbf{A}_{\boldsymbol{\Omega}}$ and $\mathbf{B}_{\boldsymbol{\Omega}}$ are symmetric.

Finally, extra simulation results and R code for the numerical studies are available at `https://github.com/yiorfun/MAUD`.

Appendix D: Supplementary Materials for Chapter 5

D.1 Technical Proof

D.1.1 Proof of Corollary 5.2.1

Proof of Corollary 5.2.1. Based on

$$\boldsymbol{\Sigma}\left(\mathbf{A}, \mathbf{B}, \boldsymbol{p}\right) = \mathrm{Bdiag}\left(\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_K\right) \times \boldsymbol{\Sigma}_{\boldsymbol{f}} \times \mathrm{Bdiag}\left(\boldsymbol{\ell}_1^\top, \ldots, \boldsymbol{\ell}_K^\top\right) + \boldsymbol{\Sigma}_{\boldsymbol{u}},$$

we rewrite it:

$$\begin{aligned}\boldsymbol{\Sigma}\left(\mathbf{A}, \mathbf{B}, \boldsymbol{p}\right) &= \mathbf{A} \circ \mathbf{I}(\boldsymbol{p}) + \mathbf{B} \circ \mathbf{J}(\boldsymbol{p}) = \mathrm{Bdiag}\left(a_{1,1}\mathbf{I}_{p_1}, \ldots, a_{K,K}\mathbf{I}_{p_K}\right) + \left(b_{k,k'}\mathbf{1}_{p_k \times p_{k'}}\right) \\ &= \mathrm{Bdiag}\left(\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_K\right) \times \boldsymbol{\Sigma}_{\boldsymbol{f}} \times \mathrm{Bdiag}\left(\boldsymbol{\ell}_1^\top, \ldots, \boldsymbol{\ell}_K^\top\right) + \boldsymbol{\Sigma}_{\boldsymbol{u}},\end{aligned}$$

where matrices $\mathbf{A} = \mathrm{diag}\left(a_{1,1}, \ldots, a_{K,K}\right)$ and $\mathbf{B} = \left(b_{k,k'}\right)$ with $b_{k',k} = b_{k,k'}$ for every $k \neq k'$ are known, $\boldsymbol{\Sigma}_{\boldsymbol{f}} = \left(\sigma_{\boldsymbol{f},kk'}\right) \in \mathbb{R}^{K \times K}$ is an unknown symmetric matrix with $\sigma_{\boldsymbol{f},k'k} = \sigma_{\boldsymbol{f},k'k}$ for every $k \neq k'$, and $\boldsymbol{\Sigma}_{\boldsymbol{u}} = \mathrm{diag}\left(\sigma_{\boldsymbol{u},11}, \ldots, \sigma_{\boldsymbol{u},pp}\right) \in \mathbb{R}^{p \times p}$ is an unknown diagonal matrix.

Do some algebraic, we have that

$$
\begin{aligned}
&\mathrm{Bdiag}\left(\boldsymbol{\ell}_1,\ldots,\boldsymbol{\ell}_K\right)\times\boldsymbol{\Sigma}_{\boldsymbol{f}}\times\mathrm{Bdiag}\left(\boldsymbol{\ell}_1^\top,\ldots,\boldsymbol{\ell}_K^\top\right)+\boldsymbol{\Sigma}_{\boldsymbol{u}}\\
&=\begin{pmatrix}\boldsymbol{\ell}_1 & \ldots & \mathbf{0}_{p_1\times 1}\\ \mathbf{0}_{p_2\times 1} & \ldots & \mathbf{0}_{p_2\times 1}\\ & \ddots & \\ \mathbf{0}_{p_K\times 1} & \ldots & \boldsymbol{\ell}_K\end{pmatrix}_{p\times K}\begin{pmatrix}\sigma_{\boldsymbol{f},11} & \ldots & \sigma_{\boldsymbol{f},1K}\\ \sigma_{\boldsymbol{f},21} & \ldots & \sigma_{\boldsymbol{f},2K}\\ \vdots & \ddots & \vdots\\ \sigma_{\boldsymbol{f},K1} & \ldots & \sigma_{\boldsymbol{f},KK}\end{pmatrix}_{K\times K}\begin{pmatrix}\boldsymbol{\ell}_1^\top & \ldots & \mathbf{0}_{1\times p_K}\\ \mathbf{0}_{1\times p_1} & \ldots & \mathbf{0}_{1\times p_K}\\ & \ddots & \\ \mathbf{0}_{1\times p_1} & \ldots & \boldsymbol{\ell}_K^\top\end{pmatrix}_{K\times p}+\boldsymbol{\Sigma}_{\boldsymbol{u}}\\
&=\left(\left(\sigma_{\boldsymbol{f},kk'}\boldsymbol{\ell}_k\boldsymbol{\ell}_{k'}^\top\right)_{p_k\times p_{k'}}\right)+\mathrm{diag}\left(\sigma_{\boldsymbol{u},11},\ldots,\sigma_{\boldsymbol{u},pp}\right).
\end{aligned}
$$

In other words, $\mathrm{Bdiag}\left(\boldsymbol{\ell}_1,\ldots,\boldsymbol{\ell}_K\right)\times\boldsymbol{\Sigma}_{\boldsymbol{f}}\times\mathrm{Bdiag}\left(\boldsymbol{\ell}_1^\top,\ldots,\boldsymbol{\ell}_K^\top\right)+\boldsymbol{\Sigma}_{\boldsymbol{u}}$ has the diagonal blocks $\left(\sigma_{\boldsymbol{f},kk}\boldsymbol{\ell}_k\boldsymbol{\ell}_k^\top\right)_{p_k\times p_k}+\mathrm{diag}\left(\sigma_{\boldsymbol{u},\bar{p}_{k-1}+1,\bar{p}_{k-1}+1},\ldots,\sigma_{\boldsymbol{u},\bar{p}_k,\bar{p}_k}\right)$ for every k and the off-diagonal blocks $\left(\sigma_{\boldsymbol{f},kk'}\boldsymbol{\ell}_k\boldsymbol{\ell}_{k'}^\top\right)_{p_k\times p_{k'}}$ for every $k\neq k'$.

Since $\boldsymbol{\Sigma}\left(\mathbf{A},\mathbf{B},\boldsymbol{p}\right)$ has the diagonal blocks $a_{k,k}\mathbf{I}_{p_k}+b_{k,k}\mathbf{J}_{p_k}$ for every k and the off-diagonal blocks $b_{k,k'}\mathbf{1}_{p_k\times p_{k'}}$ for every $k\neq k'$, by the definition of the equality of two matrices, we obtain that

$$
\begin{aligned}
a_{k,k}\mathbf{I}_{p_k}+b_{k,k}\mathbf{J}_{p_k}&=\sigma_{\boldsymbol{f},kk}\boldsymbol{\ell}_k\boldsymbol{\ell}_k^\top+\mathrm{diag}\left(\sigma_{\boldsymbol{u},\bar{p}_{k-1}+1,\bar{p}_{k-1}+1},\ldots,\sigma_{\boldsymbol{u},\bar{p}_k,\bar{p}_k}\right),\quad k=k'=1,\ldots,K\\
b_{k,k'}\mathbf{1}_{p_k\times p_{k'}}&=\sigma_{\boldsymbol{f},kk'}\boldsymbol{\ell}_k\boldsymbol{\ell}_{k'}^\top,\quad k\neq k',\quad k,k'=1,\ldots,K.
\end{aligned}
$$

First, we recall that the first element of $\boldsymbol{\ell}_k$, i.e., $\ell_{\bar{p}_{k-1}+1,k}=\iota_k\neq 0$, for every k.

Second, for fixed k, the equalities of diagonal elements yield $a_{k,k}+b_{k,k}=\sigma_{\boldsymbol{u},mm}+\sigma_{\boldsymbol{f},kk}\ell_{m,k}^2$ for all $m=\bar{p}_{k-1}+1,\ldots,\bar{p}_k$, while the equalities of off-diagonal elements yield $b_{k,k}=\sigma_{\boldsymbol{f},kk}\ell_{m,k}\ell_{m',k}$ for all pairs $m\neq m'=\bar{p}_{k-1}+1,\ldots,\bar{p}_k$.

(case 1) We focus on the off-diagonal equalities. If $\sigma_{\boldsymbol{f},kk}=0$, then $b_{k,k}=0$ but this

is impossible because $\mathbf{B} \succ 0$, by Hadamard's inequality, $b_{k,k} \neq 0$. If $\sigma_{\boldsymbol{f},kk} \neq 0$, we define $b^*_{k,k} = b_{k,k}/\sigma_{\boldsymbol{f},kk}$ thus $b^*_{k,k} = \ell_{m,k}\ell_{m',k}$ for all pairs $m \neq m' = \bar{p}_{k-1}+1,\ldots,\bar{p}_k$. Taking $m = \bar{p}_{k-1}+1$ and $m' \neq m$, we obtain that $b^*_{k,k} = \iota_k \ell_{m',k}$. Fixing m', and due to $p_k > 2$, taking another $m'' \in \{\bar{p}_{k-1}+1,\ldots,\bar{p}_k\}$ while $m'' \neq m'$ and $m'' \neq m$, we obtain that $b^*_{k,k} = \iota_k \ell_{m'',k}$. By $b^*_{k,k} = \ell_{m',k}\ell_{m'',k} = (b^*_{k,k}/\iota_k)^2$, we derive $b^*_{k,k} = \iota_k^2$, or equivalently, $\ell_{m,k} = b^*_{k,k}/\iota_k = \iota_k$ for all $m = \bar{p}_{k-1}+1,\ldots,\bar{p}_k$ and $\sigma_{\boldsymbol{f},kk} = b_{k,k}/\iota_k^2$.

(case 2) We focus on the diagonal equalities. Since $a_{k,k} + b_{k,k} = \sigma_{\boldsymbol{u},mm} + \sigma_{\boldsymbol{f},kk}\ell^2_{m,k} = \sigma_{\boldsymbol{u},mm} + \sigma_{\boldsymbol{f},kk}\iota_k^2 = \sigma_{\boldsymbol{u},mm} + b_{k,k}$, we have $\sigma_{\boldsymbol{u},mm} = a_{k,k}$ for all $m = \bar{p}_{k-1}+1,\ldots,\bar{p}_k$.

Third, for fixed pair $k \neq k'$, if $\sigma_{\boldsymbol{f},kk'} = 0$, then $b_{k,k'}\mathbf{1}_{p_k \times p_{k'}} = \mathbf{0}_{p_k \times p_{k'}}$, which implies that $b_{k,k'} = 0 = \sigma_{\boldsymbol{f},kk'}$. If $\sigma_{\boldsymbol{f},kk'} \neq 0$, then we define $b^*_{k,k'} = b_{k,k'}/\sigma_{\boldsymbol{f},kk'}$, therefore, $b^*_{k,k'}\mathbf{1}_{p_k \times p_{k'}} = \boldsymbol{\ell}_k \boldsymbol{\ell}_{k'}^\top = (\ell_{m,k}\ell_{m',k'})$, or equivalently, $b^*_{k,k'} = \ell_{m,k}\ell_{m',k'}$ for all pairs $m = \bar{p}_{k-1}+1,\ldots,\bar{p}_k$ and $m' = \bar{p}_{k'-1}+1,\ldots,\bar{p}_{k'}$. By the result in (case 1), $b^*_{k,k'} = \ell_{m,k}\ell_{m',k'} = \iota_k\iota_{k'}$ for all pairs m and m'. So, we have $b^*_{k,k'} = \iota_k\iota_{k'}$, or equivalently, $\sigma_{\boldsymbol{f},kk'} = b_{k,k'}/(\iota_k\iota_{k'})$.

In summary, we conclude $\boldsymbol{\ell}_k = \iota_k \mathbf{1}_{p_k \times 1}$ for every k, $\boldsymbol{\Sigma}_{\boldsymbol{f}} = \operatorname{diag}\left(\iota_1^{-1},\ldots,\iota_K^{-1}\right) \times \mathbf{B} \times \operatorname{diag}\left(\iota_1^{-1},\ldots,\iota_K^{-1}\right)$, and $\boldsymbol{\Sigma}_{\boldsymbol{u}} = \mathbf{A} \circ \mathbf{I}(\boldsymbol{p})$. ■

Finally, extra simulation results and R code for the numerical studies are available at `https://github.com/yiorfun/SCFA`.

Bibliography

Abou-Kandil, H., Freiling, G., Ionescu, V., and Jank, G. (2003). *Matrix Riccati equations in control and systems theory*. Birkhäuser.

Afifi, A. A. and Elashoff, R. M. (1969). Multivariate Two Sample Tests with Dichotomous and Continuous Variables. I. The Location Model. *The Annals of Mathematical Statistics* **40,** 290 – 298.

Aitkin, M. A., Nelson, W. C., and Reinfurt, K. H. (1968). Tests for correlation matrices. *Biometrika* **55,** 327–334.

An, B., Guo, J., and Liu, Y. (2014). Hypothesis testing for band size detection of high-dimensional banded precision matrices. *Biometrika* **101,** 477–483.

Anderson, T. (1992). Introduction to hotelling (1931) the generalization of student's ratio. In *Breakthroughs in Statistics*, pages 45–53. Springer.

Anderson, T. (2003). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons.

Anderson, T. W. (1969). Statistical inference for covariance matrices with linear structure. *Multivariate Analysis II* pages 55–66.

Anderson, T. W. (1973). Asymptotically Efficient Estimation of Covariance Matrices with Linear Structure. *The Annals of Statistics* **1,** 135 – 141.

Anderson, T. W. et al. (1970). Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. *Essays in probability and statistics* pages 1–24.

Antoniadis, A. and Fan, J. (2001). Regularization of wavelet approximations. *Journal of the American Statistical Association* **96,** 939–967.

Archakov, I. and Hansen, P. R. (2022). A Canonical Representation of Block Matrices with Applications to Covariance and Correlation Matrices. *The Review of Economics and Statistics* pages 1–39.

Arnold, S. F. (1973). Application of the theory of products of problems to certain patterned covariance matrices. *Annals of Statistics* **1,** 682–699.

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

Banerjee, S. and Ghosal, S. (2014). Posterior convergence rates for estimating large precision matrices using graphical models. *Electronic Journal of Statistics* **8,** 2111 – 2137.

Basilevsky, A. T. (2009). *Statistical factor analysis and related methods: theory and applications*. John Wiley & Sons.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57,** 289–300.

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics* **29,** 1165–1188.

Bhoj, D. S. (1987). Testing hypotheses on the mean vector under an intraclass correlation structure. *Biometrical Journal* **29,** 783–789.

Bickel, P. J. and Doksum, K. A. (2015a). *Mathematical statistics: basic ideas and selected topics*, volume 1. CRC Press, 2 edition.

Bickel, P. J. and Doksum, K. A. (2015b). *Mathematical statistics: basic ideas and selected topics*, volume 2. CRC Press, 2 edition.

Bickel, P. J. and Gel, Y. R. (2011). Banded regularization of autocovariance matrices in application to parameter estimation and forecasting of time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73,** 711–728.

Bickel, P. J. and Levina, E. (2008a). Regularized estimation of large covariance matrices. *Annals of Statistics* **36,** 199–227.

Bickel, P. J. and Levina, E. (2008b). Covariance regularization by thresholding. *Annals of Statistics* **36,** 2577–2604.

Bien, J. (2019). Graph-guided banding of the covariance matrix. *Journal of the American Statistical Association* **114,** 782–792.

Bove, D. S., Dedic, J., Kelkhoff, D., Kunzmann, K., Lang, B. M., Li, L., Wang, Y., and Gower-Page, C. (2022). *Mixed Models for Repeated Measures*. R package version 0.2.2.

Bowman, F. D. (2005). Spatio-temporal modeling of localized brain activity. *Biostatistics* **6,** 558–575.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.

Buaphim, N., Onsaard, K., So-ngoen, P., and Rungratgasame, T. (2018). Some reviews on ranks of upper triangular block matrices over a skew field. In *International Mathematical Forum*, volume 13, pages 323–335.

Cadima, J., Calheiros, F. L., and Preto, I. P. (2010). The eigenstructure of block-structured correlation matrices and its implications for principal component analysis. *Journal of Applied Statistics* **37,** 577–589.

Cai, T. and Liu, W. (2011). Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association* **106,** 672–684.

Cai, T. T., Ren, Z., and Zhou, H. H. (2013). Optimal rates of convergence for estimating toeplitz covariance matrices. *Probability Theory and Related Fields* **156,** 101–143.

Cai, T. T., Ren, Z., and Zhou, H. H. (2016). Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electronic Journal of Statistics* **10,** 1–59.

Cai, T. T. and Yuan, M. (2012). Adaptive covariance matrix estimation through block thresholding. *The Annals of Statistics* **40,** 2014–2042.

Chen, S., Bowman, F. D., and Mayberg, H. S. (2016). A bayesian hierarchical framework for modeling brain connectivity for neuroimaging data. *Biometrics* **72,** 596–605.

Chen, S., Kang, J., Xing, Y., Zhao, Y., and Milton, D. K. (2018). Estimating large covariance matrix with network topology for high-dimensional biomedical data. *Computational Statistics and Data Analysis* **127,** 82 – 95.

Chen, S., Li, M., Hong, D., Billheimer, D., Li, H., Xu, B. J., and Shyr, Y. (2009). A novel comprehensive wave-form MS data processing method. *Bioinformatics* **25,** 808–814.

Chen, S., Zhang, Y., Wu, Q., Bi, C., Kochunov, P., and Hong, L. E. (2023). Identifying covariate-related subnetworks for whole-brain connectome analysis. *Biostatistics* kxad007.

Chiappelli, J., Rowland, L. M., Wijtenburg, S. A., Chen, H., Maudsley, A. A., Sheriff, S., Chen, S., Savransky, A., Marshall, W., Ryan, M. C., Bruce, H. A., Shuldiner, A. R., Mitchell, B. D., Kochunov, P., and Hong, L. E. (2019). Cardiovascular risks impact human brain n-acetylaspartate in regionally specific patterns. *Proceedings of the National Academy of Sciences* **116,** 25243–25249.

Clement, B., Chakraborty, S., Sinha, B. K., and Giri, N. C. (1981). Tests for the mean vector under intraclass covariance structure. *Journal of Statistical Computation and Simulation* **12,** 237–245.

Colizza, V., Flammini, A., Serrano, M. A., and Vespignani, A. (2006). Detecting rich-club ordering in complex networks. *Nature physics* **2,** 110–115.

Consul, P. C. (1968). On the distribution of votaw's likelihood ratio criterion l for testing the bipolarity of a covariance matrix. *Mathematische Nachrichten* **36,** 1–13.

Craig, A. T. (1943). Note on the independence of certain quadratic forms. *The Annals of Mathematical Statistics* **14,** 195–197.

Derado, G., Bowman, F. D., and Kilts, C. D. (2010). Modeling the spatial and temporal dependence in fmri data. *Biometrics* **66,** 949–957.

Devijver, E. and Gallopin, M. (2018). Block-diagonal covariance selection for high-dimensional gaussian graphical models. *Journal of the American Statistical Association* **113,** 306–314.

Driscoll, M. F. and Gundberg, W. R. (1986). A history of the development of craig's theorem. *The American Statistician* **40,** 65–70.

Dyer, D. (1982). The convolution of generalized f distributions. *Journal of the American Statistical Association* **77,** 184–189.

Dykstra, R. L. (1970). Establishing the positive definiteness of the sample covariance matrix. *Annals of Mathematical Statistics* **41,** 2153–2154.

Efron, B. (2004). Large-scale simultaneous hypothesis testing. *Journal of the American Statistical Association* **99,** 96–104.

Efron, B. (2007). Correlation and large-scale simultaneous significance testing. *Journal of the American Statistical Association* **102,** 93–103.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *Annals of Statistics* **32,** 407–499.

Fan, J. (2005). Rejoinder: A Selective Overview of Nonparametric Methods in Financial Econometrics. *Statistical Science* **20,** 351 – 357.

Fan, J., Fan, Y., and Lv, J. (2008). High dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* **147,** 186–197.

Fan, J. and Han, X. (2017). Estimation of the false discovery proportion with unknown dependence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79,** 1143–1164.

Fan, J., Han, X., and Gu, W. (2012). Estimating false discovery proportion under arbitrary covariance dependence. *Journal of the American Statistical Association* **107,** 1019–1035. PMID: 24729644.

Fan, J., Ke, Y., Sun, Q., and Zhou, W.-X. (2019). Farmtest: Factor-adjusted robust multiple testing with approximate false discovery control. *Journal of the American Statistical Association* **114,** 1880–1893. PMID: 33033420.

Fan, J., Li, R., Zhang, C.-H., and Zou, H. (2020). *Statistical foundations of data science*. Chapman and Hall/CRC.

Fan, J., Liao, Y., and Liu, H. (2016). An overview of the estimation of large covariance and precision matrices. *The Econometrics Journal* **19,** C1–C32.

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

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

Fan, J., Liao, Y., and Mincheva, M. (2016). *POET: Principal Orthogonal ComplEment Thresholding (POET) Method*. R package version 2.0.

Fan, J., Liu, H., and Wang, W. (2018). Large covariance estimation through elliptical factor models. *The Annals of Statistics* **46,** 1383 – 1414.

Fan, J. and Lv, J. (2008). Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70,** 849–911.

Ferguson, T. (1996). *A Course in Large Sample Theory*. Chapman & Hall Texts in Statistical Science Series. Springer US.

Fleiss, J. L. (1966). Assessing the accuracy of multivariate observations. *Journal of the American Statistical Association* **61,** 403–412.

Fomby, T. B., Hill, R. C., and Johnson, S. R. (1984). *Advanced econometric methods*. Springer, New York, NY.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports* **486,** 75–174.

Fox, J. (2006). Teacher's corner: Structural equation modeling with the sem package in r. *Structural Equation Modeling: A Multidisciplinary Journal* **13,** 465–486.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **9,** 432–441.

Friguet, C., Kloareg, M., and Causeur, D. (2009). A factor model approach to multiple testing under dependence. *Journal of the American Statistical Association* **104,** 1406–1415.

Friston, K. J., Holmes, A. P., Poline, J., Grasby, P., Williams, S., Frackowiak, R. S., Turner, R., et al. (1995). Analysis of fmri time-series revisited. *Neuroimage* **2,** 45–53.

Gana, K. and Broc, G. (2019). *Structural equation modeling with lavaan*. John Wiley & Sons.

Geisser, S. (1963). Multivariate analysis of variance for a special covariance case. *Journal of the American Statistical Association* **58,** 660–669.

Geisser, S. (1964). Estimation in the uniform covariance case. *Journal of the Royal Statistical Society. Series B (Methodological)* **26,** 477–483.

Girvan, M. and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* **99,** 7821–7826.

Günther, M. and Klotz, L. (2012). Schur's theorem for a block hadamard product. *Linear Algebra and its Applications* **437,** 948–956.

Hanson, D. L. and Wright, F. T. (1971). A Bound on Tail Probabilities for Quadratic Forms in Independent Random Variables. *The Annals of Mathematical Statistics* **42,** 1079 – 1083.

Haq, M. S. (1974). A multivariate model with intra-class covariance structure. *Annals of the Institute of Statistical Mathematics* **26,** 413–420.

Hastie, T., Tibshirani, R., and Wainwright, M. (2015). *Statistical learning with sparsity: the lasso and generalizations*. CRC press.

Hayashi, F. (2011). *Econometrics*. Princeton University Press.

He, K., Kang, J., Hong, H. G., Zhu, J., Li, Y., Lin, H., Xu, H., and Li, Y. (2019). Covariance-insured screening. *Computational Statistics and Data Analysis* **132,** 100–114. Special Issue on Biostatistics.

He, K., Li, Y., Zhu, J., Liu, H., Lee, J. E., Amos, C. I., Hyslop, T., Jin, J., Lin, H., Wei, Q., and Li, Y. (2015). Component-wise gradient boosting and false discovery control in survival analysis with high-dimensional covariates. *Bioinformatics* **32,** 50–57.

He, T. and Wang, L. (2022). On seemingly unrelated regressions with uniform correlation error. *Communications in Statistics - Theory and Methods* **51,** 5714–5727.

Henderson, H. V. and Searle, S. R. (1979). Vec and vech operators for matrices, with some uses in jacobians and multivariate statistics. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* **7,** 65–81.

Homans, G. C. (2013). *The human group*. Routledge.

Horn, R. A., Mathias, R., and Nakamura, Y. (1991). Inequalities for unitarily invariant norms and bilinear matrix products. *Linear and Multilinear Algebra* **30,** 303–314.

Hotelling, H. (1947). Multivariate quality control, illustrated by the air testing of sample bombsights.

Hotelling, H. (1951). A generalized t test and measure of multivariate dispersion. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1951*. University of California Press.

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

Huang, P.-H. (2020). lslx: Semi-confirmatory structural equation modeling via penalized likelihood. *Journal of Statistical Software* **93,** 1–37.

Huttlin, E. L., Bruckner, R. J., Paulo, J. A., Cannon, J. R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M. P., Parzen, H., et al. (2017). Architecture of the human interactome defines protein communities and disease networks. *Nature* **545,** 505–509.

ISGlobal (2021). Barcelona institute for global health. `https://www.isglobal.org/en`. Accessed: 2022-07-30.

Ito, K. (1956). Asymptotic Formulae for the Distribution of Hotelling's Generalized T_0^2 Statistic. *The Annals of Mathematical Statistics* **27,** 1091 – 1105.

Ito, K. (1960). Asymptotic Formulae for the Distribution of Hotelling's Generalized T_0^2 Statistic. II. *The Annals of Mathematical Statistics* **31,** 1148 – 1153.

Jackson, D. L., Gillaspy Jr, J. A., and Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychological methods* **14,** 6.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics* **29,** 295–327.

Johnstone, I. M. and Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association* **104,** 682–693. PMID: 20617121.

Johnstone, I. M. and Paul, D. (2018). Pca in high dimensions: An orientation. *Proceedings of the IEEE* **106,** 1277–1292.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* **34,** 183–202.

Karoui, N. E. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *The Annals of Statistics* **36,** 2717–2756.

Ke, H., Ren, Z., Qi, J., Chen, S., Tseng, G. C., Ye, Z., and Ma, T. (2022). High-dimension to high-dimension screening for detecting genome-wide epigenetic and noncoding rna regulators of gene expression. *Bioinformatics* **38,** 4078–4087.

Korin, B. P. (1968). On the distribution of a statistic used for testing a covariance matrix. *Biometrika* **55,** 171–178.

Koziol, A., Roy, A., Zmyślony, R., Leiva, R., and Fonseca, M. (2017). Best unbiased estimates for parameters of three-level multivariate data with doubly exchangeable covariance structure. *Linear Algebra and its Applications* **535,** 87–104.

Lam, C. and Fan, J. (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Annals of Statistics* **37,** 4254–4278.

Lancaster, H. O. (1965). The helmert matrices. *The American Mathematical Monthly* **72,** 4–12.

Lawley, D. (1958). Estimation in factor analysis under various initial assumptions. *British journal of statistical Psychology* **11,** 1–12.

Lawley, D. N. (1938). A generalization of fisher's z test. *Biometrika* **30,** 180–187.

Ledoit, O. and Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* **88,** 365–411.

Lee, C. and Wilkinson, D. J. (2019). A review of stochastic block models and extensions for graph clustering. *Applied Network Science* **4,** 1–50.

Lee, H., Chen, C., Kochunov, P., Hong, E. L., and Chen, S. (2023). Fast autoregressive model for multivariate dependent outcomes with application to whole-brain region-wise neuroimaging data analysis. submitted.

Lee, J. C. and Hu, L. (1996). On the distribution of linear functions of independent f and u variates. *Statistics & probability letters* **26,** 339–346.

Lee, K. and Lee, J. (2021). Estimating large precision matrices via modified cholesky decomposition. *Statistica Sinica* **31,** 173–196.

Lee, K., Lin, L., and You, K. (2021). *CovTools: Statistical Tools for Covariance Analysis*. R package version 0.5.4.

Leek, J. T. and Storey, J. D. (2008). A general framework for multiple testing dependence. *Proceedings of the National Academy of Sciences* **105,** 18718–18723.

Leiva, R. (2007). Linear discrimination with equicorrelated training vectors. *Journal of Multivariate Analysis* **98,** 384–409.

Levine, J. H., Simonds, E. F., Bendall, S. C., Davis, K. L., El-ad, D. A., Tadmor, M. D., Litvin, O., Fienberg, H. G., Jager, A., Zunder, E. R., et al. (2015). Data-driven phenotypic dissection of aml reveals progenitor-like cells that correlate with prognosis. *Cell* **162,** 184–197.

Lu, C. and Schmidt, P. (2012). Conditions for the numerical equality of the ols, gls and amemiya–cragg estimators. *Economics Letters* **116,** 538–540.

Magnus, J. R. (1978). Maximum likelihood estimation of the gls model with unknown parameters in the disturbance covariance matrix. *Journal of Econometrics* **7,** 281 – 312.

Magwene, P. (2021). Biology 723: Statistical computing for biologists. `https://bio723-class.github.io/Bio723-book/index.html`. Accessed: 2022-07-30.

Mathai, A. and Katiyar, R. (1979). The distribution and the exact percentage points for wilks' l mvc criterion. *Annals of the Institute of Statistical Mathematics* **31,** 215–224.

Mathai, A. M. and Provost, S. B. (1992). *Quadratic forms in random variables: theory and applications*. Dekker.

Mathai, A. M. and Rathie, P. (1970). The exact distribution of votaw's criteria. *Annals of the Institute of Statistical Mathematics* **22,** 89–116.

Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *Annals of Mathematical Statistics* **11,** 204–209.

McKeon, J. J. (1974). F approximations to the distribution of hotelling's t2 0. *Biometrika* **61,** 381–383.

Morrison, D. F. (1971). The distribution of linear functions of independent f variates. *Journal of the American Statistical Association* **66,** 383–385.

Morrison, D. F. (1972). The analysis of a single sample of repeated measurements. *Biometrics* **28,** 55–71.

Muirhead, R. J. (2005). *Aspects of multivariate statistical theory*. John Wiley & Sons.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* **103,** 8577–8582.

Newman, M. E. J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E* **69,** 026113.

Oberski, D. (2014). lavaan.survey: An r package for complex survey analysis of structural equation models. *Journal of Statistical Software* **57,** 1–27.

Ogawa, J. and Olkin, I. (2008). A tale of two countries: The craig–sakamoto–matusita theorem. *Journal of Statistical Planning and Inference* **138,** 3419–3428. Special Issue in Honor of Junjiro Ogawa (1915 - 2000): Design of Experiments, Multivariate Analysis and Statistical Inference.

Olkin, I. (1972). Testing and estimation for structures which are circularly symmetric in blocks. *ETS Research Bulletin Series* **1972,** i–20.

Olkin, I. and Press, S. J. (1969). Testing and estimation for a circular stationary model. *Annals of Mathematical Statistics* **40,** 1358–1373.

Pal, S., Malekmohammadi, S., Regol, F., Zhang, Y., Xu, Y., and Coates, M. (2020). Non parametric graph learning for bayesian graph neural networks. In Peters, J. and Sontag, D., editors, *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124 of *Proceedings of Machine Learning Research*, pages 1318–1327. PMLR.

Perrot-Dockès, M., Lévy-Leduc, C., and Rajjou, L. (2022). Estimation of large block structured covariance matrices: Application to 'multi-omic' approaches to study seed quality. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **71,** 119–147.

Pillai, K. and Young, D. (1971). On the exact distribution of hotelling's generalized t02. *Journal of Multivariate Analysis* **1,** 90–107.

Pourahmadi, M. (2013). *High-dimensional covariance estimation: with high-dimensional data*, volume 882. John Wiley & Sons.

Puntanen, S. and Styan, G. P. H. (1989). The equality of the ordinary least squares estimator and the best linear unbiased estimator. *The American Statistician* **43,** 153–161.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ran, A. and Rodman, L. (1984). The algebraic matrix riccati equation. In *Topics in Operator Theory Systems and Networks*, pages 351–381. Springer.

Ravikumar, P., Wainwright, M. J., Raskutti, G., and Yu, B. (2011). High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence. *Electronic Journal of Statistics* **5,** 935 – 980.

Ravishanker, N. and Dey, D. K. (2002). *A first course in linear model theory*. Chapman and Hall/CRC.

Revelle, W. (2023). *Procedures for Psychological, Psychometric, and Personality Research*. R package version 2.3.6.

Risk, B. B., Matteson, D. S., Spreng, R. N., and Ruppert, D. (2016). Spatiotemporal mixed modeling of multi-subject task fmri via method of moments. *NeuroImage* **142,** 280–292.

Ritchie, S. C., Surendran, P., Karthikeyan, S., Lambert, S. A., Bolton, T., Pennells, L., Danesh, J., Di Angelantonio, E., Butterworth, A. S., and Inouye, M. (2023). Quality control and removal of technical variation of nmr metabolic biomarker data in˜ 120,000 uk biobank participants. *Scientific Data* **10,** 64.

Rogers, G. and Young, D. (1974). Testing and estimation when a normal covariance matrix has intraclass structure of arbitrary order. *Communications in Statistics* **3,** 343–359.

Rogers, G. S. and Young, D. L. (1975). Some likelihood ratio tests when a normal covariance matrix has certain reducible linear structures. *Communications in Statistics* **4,** 537–554.

Rosseel, Y. (2012). lavaan: An r package for structural equation modeling. *Journal of Statistical Software* **48,** 1–36.

Rosseel, Y., Jorgensen, T. D., Rockwood, N., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., Scharf, F., and Du, H. (2023). *Latent Variable Analysis*. R package version 0.6-15.

Rothman, A. J., Bickel, P. J., Levina, E., and Zhu, J. (2008). Sparse permutation invariant covariance estimation. *Electronic Journal of Statistics* **2,** 494–515.

Rothman, A. J., Levina, E., and Zhu, J. (2009). Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association* **104,** 177–186.

Roustant, O. and Deville, Y. (2017). On the validity of parametric block correlation matrices with constant within and between group correlations. *arXiv preprint arXiv:1705.09793* .

Roustant, O., Padonou, E., Deville, Y., Clément, A., Perrin, G., Giorla, J., and Wynn, H. (2020). Group kernels for gaussian process metamodels with categorical inputs. *SIAM/ASA Journal on Uncertainty Quantification* **8,** 775–806.

Roy, A. and Leiva, R. (2008). Likelihood ratio tests for triply multivariate data with structured correlation on spatial repeated measurements. *Statistics & Probability Letters* **78,** 1971–1980.

Roy, A. and Leiva, R. (2011). Estimating and testing a structured covariance matrix for three-level multivariate data. *Communications in Statistics - Theory and Methods* **40,** 1945–1963.

Roy, A., Leiva, R., Žežula, I., and Klein, D. (2015). Testing the equality of mean vectors for paired doubly multivariate observations in blocked compound symmetric covariance matrix setup. *Journal of Multivariate Analysis* **137,** 50–60.

Roy, A., Zmyślony, R., Fonseca, M., and Leiva, R. (2016). Optimal estimation for doubly multivariate data in blocked compound symmetric covariance structure. *Journal of Multivariate Analysis* **144,** 81–90.

Roy, J. and Murthy, V. (1960). Percentage points of wilks' l mvc and l vc criteria. *Psychometrika* **25,** 243–250.

Rudelson, M. and Vershynin, R. (2013). Hanson-Wright inequality and sub-gaussian concentration. *Electronic Communications in Probability* **18,** 1 – 9.

Schmidt, P. (2020). *Econometrics*. CRC Press.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., and King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research* **99,** 323–338.

Seely, J. (1971). Quadratic Subspaces and Completeness. *The Annals of Mathematical Statistics* **42,** 710 – 721.

Seely, J. (1977). Minimal sufficient statistics and completeness for multivariate normal families. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* **39,** 170–185.

Silva, E. R. D. O. E., Foster, D., Monnie McGee Harper, M., Seidman, C. E., Smith, J. D., Breslow, J. L., and Brinton, E. A. (2000). Alcohol consumption raises hdl cholesterol levels by increasing the transport rate of apolipoproteins a-i and a-ii. *Circulation* **102,** 2347–2352.

Simpson, S. L., Bowman, F. D., and Laurienti, P. J. (2013). Analyzing complex functional brain networks: Fusing statistics and network science to understand the brain. *Statistics Surveys* **7,** 1 – 36.

Sinha, B. K. and Wieand, H. S. (1979). Union-intersection test for the mean vector when the covariance matrix is totally reducible. *Journal of the American Statistical Association* **74,** 340–343.

Siotani, M. (1971). An Asymptotic Expansion of the Non-Null Distribution of Hotelling's Generalized T_0^2-Statistic. *The Annals of Mathematical Statistics* **42,** 560 – 571.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D., and Futcher, B. (1998). Comprehensive identification of cell cycle–regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization. *Molecular Biology of the Cell* **9,** 3273–3297. PMID: 9843569.

Spjøtvoll, E. (1972). Joint confidence intervals for all linear functions of means in the one-way layout with unknown group variances. *Biometrika* **59,** 683–685.

Srivastava, J. N. (1966). On testing hypotheses regarding a class of covariance structures. *Psychometrika* **31,** 147–164.

Storey, J. D., Taylor, J. E., and Siegmund, D. (2004). Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66,** 187–205.

Stuart., A., Ord, J., and Arnold, S. (1999). *Advanced Theory of Statistics, Volume 2A: Classical Inference and the Linear Model*. London: Oxford University Press.

Szatrowski, T. H. (1976). Estimation and testing for block compound symmetry and other patterned covariance matrices with linear and nonlinear structure. Technical Report 107, Stanford University.

Szatrowski, T. H. (1980). Necessary and Sufficient Conditions for Explicit Solutions in the Multivariate Normal Estimation Problem for Patterned Means and Covariances. *The Annals of Statistics* **8,** 802 – 810.

Szatrowski, T. H. (1982). Testing and estimation in the block compound symmetry problem. *Journal of Educational Statistics* **7,** 3–18.

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

van der Vaart, A. (2000). *Asymptotic Statistics*. Asymptotic Statistics. Cambridge University Press.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer.

Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Votaw, D. F. (1948). Testing compound symmetry in a normal multivariate distribution. *Annals of Mathematical Statistics* **19,** 447–473.

Votaw, D. F., Kimball, A. W., and Rafferty, J. A. (1950). Compound symmetry tests in the multivariate analysis of medical experiments. *Biometrics* **6,** 259–281.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press.

Wang, B. (2015). *Regularized Estimators of Covariance Matrices with CV Tuning*. R package version 1.0.

Wichura, M. J. (2006). *The coordinate-free approach to linear models*, volume 19. Cambridge University Press.

Wilks, S. S. (1946). Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution. *Annals of Mathematical Statistics* **17,** 257–281.

Worsley, K. J. and Friston, K. J. (1995). Analysis of fmri time-series revisited—again. *Neuroimage* **2,** 173–181.

Wright, F. T. (1973). A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *The Annals of Probability* **1,** 1068–1070.

Wu, Q., Ma, T., Liu, Q., Milton, D. K., Zhang, Y., and Chen, S. (2021). ICN: extracting interconnected communities in gene co-expression networks. *Bioinformatics* **37,** 1997–2003.

Wu, W. B. and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90,** 831–844.

Yildiz, P. B., Shyr, Y., Rahman, J. S., Wardwell, N. R., Zimmerman, L. J., Shakhtour, B., Gray, W. H., Chen, S., Li, M., Roder, H., Liebler, D. C., Bigbee, W. L., Siegfried, J. M., Weissfeld, J. L., Gonzalez, A. L., Ninan, M., Johnson, D. H., Carbone, D. P., Caprioli, R. M., and Massion, P. P. (2007). Diagnostic accuracy of maldi mass spectrometric analysis of unfractionated serum in lung cancer. *Journal of Thoracic Oncology* **2,** 893 – 901.

Young, D. L. (1976). Inference concerning the mean vector when the covariance matrix is totally reducible. *Journal of the American Statistical Association* **71,** 696–699.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in the gaussian graphical model. *Biometrika* **94,** 19–35.

Žežula, I., Klein, D., and Roy, A. (2018). Testing of multivariate repeated measures data with block exchangeable covariance structure. *TEST* **27,** 360–378.

Zhang, J. (2018). On the distribution of a quadratic form in normal variates. *REVSTAT-Statistical Journal* **16,** 315–322.

Zhao, Y. (2017). A survey on theoretical advances of community detection in networks. *WIREs Computational Statistics* **9,** e1403.

Zitnik, M., Sosič, R., and Leskovec, J. (2018). Prioritizing network communities. *Nature communications* **9,** 2544.

Zmyślony, R. (1976). On estimation of parameters in linear models. *Applicationes Mathematicae* **3,** 271–276.

Zmyślony, R. (1980). Completeness for a family of normal distributions. *Banach Center Publications* **6,** 355–357.