Methods for Correlated Data: Large-scale Linear Mixed Models and Brain Connectivity Networks

Kun Yue

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Ali Shojaie, Chair

Eardi Lila, Chair

Adam A. Szpiro

Program Authorized to Offer Degree:

Biostatistics

University of Washington

Abstract

Methods for Correlated Data: Large-scale Linear Mixed Models and Brain Connectivity Networks

Kun Yue

Co-Chairs of the Supervisory Committee:

Ali Shojaie

Department of Biostatistics

Eardi Lila

Department of Biostatistics

This dissertation addresses the challenges associated with correlated data in diverse fields, emphasizing both statistical methodology and applications in the realms of genetics and neuroimaging. The overarching theme revolves around the development and refinement of statistical tools, particularly focusing on large-scale linear mixed models and brain connectivity networks.

In the second chapter, we confront the computational inefficiencies and instability issues of standard methods for estimating variance components in linear mixed models, commonly used in genetic studies. Utilizing regularized estimation strategies, we propose the restricted Haseman-Elston (REHE) regression and its resampling variant (reREHE) estimators, along with an inference framework for REHE, as fast and robust alternatives that provide non-negative estimates with comparable accuracy to REML. The merits of REHE are illustrated using real data and benchmark simulation studies.

The third chapter is motivated by the problem of inferring the graph structure of functional

connectivity networks from multi-level functional magnetic resonance imaging (fMRI) data. We develop a valid inference framework for high-dimensional graphical models that accounts for group-level heterogeneity. We introduce a neighborhood-based method to learn the graph structure and reframe the problem as that of inferring fixed effect parameters in a doubly high-dimensional linear mixed model. Specifically, we propose a LASSO-based estimator and a de-biased LASSO-based inference framework for the fixed effect parameters of the linear mixed model. Moreover, we introduce consistent estimators for the variance components in order to identify subject-specific edges in the inferred graph. We also adapt our method to account for serial correlation by learning heterogeneous graphs in the setting of a vector autoregressive model. We demonstrate the performance of the proposed framework using real data and benchmark simulation studies.

The fourth chapter delves into the temporally dynamic brain connectivity of the default mode network as a potential biomarker for Alzheimer's Disease (AD). Existing amyloid beta ($A\beta$) biomarkers, though effective, confront practical limitations. Brain functional connectivity alterations linked to AD pathology propose a non-invasive avenue for $A\beta$ detection. However, current FC measurements lack standalone sensitivity. We investigate temporally dynamic FC through resting-state functional MRI and introduce the Generalized Autoregressive Conditional Heteroscedastic Dynamic Conditional Correlation (DCC-GARCH) model. To fulfill the model assumptions, we employ whitening procedures to remove the serial correlations. Recognizing the limitations of traditional methods, we introduce an iterative data-adaptive autoregressive model (IDAR) capable of modeling complex serial correlation structures for both long- and short-TR datasets. We comprehensively illustrate IDAR's performance by assessing residual serial correlations post-whitening and type-I error rates in task testing. After applying the IDAR approach to pre-process fMRI signals, we estimate dynamic functional connectivity profiles with DCC-GARCH models. Our results demonstrate superior sensitivity to CSF $A\beta$ status and provide crucial insights into dynamic functional connectivity analysis in AD.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am deeply grateful to my advisors, Dr. Ali Shojaie and Dr. Eardi Lila, for their unwavering guidance, support, and expertise throughout my journey in pursuing the degree. Their patience and the invaluable advice they generously offered have been pivotal in refining my work and fostering my academic development. Beyond shaping my scholarly endeavors, their mentorship has made substantial contributions to my personal and professional growth. Their dedication to my success has cultivated an environment where learning not only thrives but flourishes.

My heartfelt appreciation extends to my committee members, Dr. Adam Szpiro and Dr. Aakanksha Singhvi, for their insightful feedback, constructive critique, and invaluable contributions that have intricately shaped the refinement of my research. The diversity in their perspectives has undeniably enriched both the depth and breadth of my work. Additionally, I would like to express my appreciation to Dr. Hesam Jahanian, Dr. Thomas J. Grabowski, Dr. Jing Ma, and Dr. Jason Webster for their generous assistance and insightful input on my projects.

I express my deepest gratitude to my husband, Yikun Li, whose constant emotional support, tender care, and boundless love have served as a steadfast anchor during the challenging phases of this academic pursuit. His encouragement and profound understanding have played an instrumental role in navigating the intricate complexities of doctoral research.

I owe a profound debt of gratitude to my parents for their steadfast support and unyielding belief in my decisions. Their encouragement has been an unwavering source of motivation, empowering me to relentlessly pursue and realize my academic aspirations.

A special note of appreciation goes to my dear friends Sijia Li, Jingyi Mao, Tianyu Zhang,

Xiudi Li, Xiaoshun Zeng, Xu Wang, Yezi Yang, Edward Zhao, Yunhua Xiang. Their companionship, support, and shared moments during both study sessions and leisure times have added immeasurable richness to my academic journey.

Heartfelt thanks to my ever-joyful companion, JoJo, whose presence provided solace and moments of joy amidst the long hours of research. His wagging tail and playful spirit served as a constant reminder of the importance of maintaining balance and finding joy amid the rigors of academia.

To everyone, whether playing a big or small role in this academic odyssey, I extend my deepest gratitude. Your collective support has been instrumental in shaping my trajectory and achieving this significant milestone.

DEDICATION

to my family

Chapter 1

INTRODUCTION

Correlated data is ubiquitous across various scientific domains. In genetic studies, correlations can arise from the genetic relatedness [118] and social dependence [41] among subjects. In neuroscience experiments, correlations in functional magnetic resonance imaging (fMRI) signals stem from the prolonged hemodynamic responses induced by neural activities [30]. In addition, correlations commonly arise with clustered observation structures, such as in multi-subject experiments [154]. Correlated data demand dedicated statistical methodologies to extract meaningful insights. More importantly, the validity of analyses hinges on proper accounting for these correlation structures. As technological advancements enable the collection of large datasets with unique features, novel challenges emerge, necessitating innovative statistical tools. This dissertation addresses these challenges, with a particular focus on methods for large-scale linear mixed models and brain connectivity network analysis.

Chapter 2 delves into the realm of genetic studies, where linear mixed models serve as a powerful tool with broad applications, including heritability estimation [206], genome-wide association studies [5], and network-based pathway enrichment analysis [147, 148]. The gold-standard approach for estimating variance components in linear mixed models is the restricted maximum likelihood (REML) method [169]. However, its computational demands, especially with non-sparse covariance structures in large datasets, limit its scalability. In addition, REML exhibits numerical instability and unreliability in certain settings. Recognizing these shortcomings, alternative moment estimators, such as analysis of variance and the Haseman-Elston regression estimator [175, 91], have been used because of their computational efficiency. However, these estimators are not guaranteed to provide non-negative variance estimates, posing challenges in interpretation and downstream analyses. To address these challenges, the chapter proposes the restricted Haseman-Elston (REHE) regression as a

robust and efficient estimator of variance components for large-scale linear mixed models. REHE ensures non-negative estimates, maintaining accuracy comparable to REML when it performs well and proving robust in settings where REML falters. We also introduce REHE with resampling (reREHE), a variant of REHE that offers positive variance component estimates with high probability. Additionally, we propose bootstrap confidence intervals for REHE estimates to facilitate inference. We demonstrate the advantages of the proposed approach through simulation studies and applications to data from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) [210, 41] and The Cancer Genome Atlas (TCGA) breast cancer dataset [216].

In Chapter 3, our focus remains on methodologies for linear mixed models, with an emphasis on those with high-dimensional covariates. However, the application emphasis shifts toward neuroscience, specifically the inference of functional brain connectivity networks from multi-level fMRI data. We bring in Gaussian graphical models to characterize the brain connectivity network, and adopt the neighbourhood selection approach for network inference. The multi-subject experiment design introduces subject-level heterogeneity in connectivity networks [53, 154, 157]. We demonstrate through toy examples and simulation studies that methods neglecting this heterogeneity, such as the standard LASSO framework, are inherently flawed. Additionally, widely used techniques such as two-stage strategies exhibit limitations in this context. To address these issues, we integrate mixed effects into the neighborhood-selection approach and reframe the problem as inferring fixed effect parameters within a doubly high-dimensional linear mixed model. Given the absence of an established theoretical framework for doubly high-dimensional linear mixed models, Chapter 3 proposes a LASSO-based estimator and a de-biased LASSO-based inference framework for the fixed effect parameters. We leverage random matrix theory to address challenges arising from overlapping columns in the fixed and random effect design matrices, which hold particular relevance to the application of brain connectivity network inference. Furthermore, consistent estimators for the variance components are introduced to identify subject-specific edges in the inferred graph. To highlight the versatility of our proposed approach, we adapt our method to accommodate serial correlation by learning heterogeneous graphs in the context of a mixed-effect vector

autoregressive model. The performance of the proposed framework is demonstrated through benchmark simulation studies and the analysis of data from the Human Connectome Project [221].

Chapter 4 is dedicated exclusively to the analysis of functional brain connectivity networks, with a particular focus on characterizing the dynamic correlation profile within the fMRI dataset, and addressing challenges associated with serial correlations in fMRI observations. The driving force behind this investigation is the quest for sensitive biomarkers crucial for early-stage Alzheimer's Disease (AD) diagnosis. Practical limitations of existing Cerebrospinal fluid (CSF) amyloid beta ($A\beta$) biomarkers have prompted an examination of brain functional connectivity alterations associated with AD pathology. Given the constraints of traditional methods that predominantly rely on stationary measurements of functional connectivities [217, 102, 34, 213, 168], this chapter introduces the application of the Generalized Autoregressive Conditional Heteroscedastic Dynamic Conditional Correlation (DCC-GARCH) model to describe the temporally dynamic correlation profile in resting-state fMRI signals. To meet the assumptions of the DCC-GARCH model, a whitening procedure based on the Iterative Data-Adaptive Autoregressive Model (IDAR) is proposed to address serial correlations in the observations. The IDAR method is efficient in handling complex serial correlation structures for both long- and short-TR fMRI datasets, which is demonstrated through comprehensive evaluations. Subsequently, we apply the DCC-GARCH model coupled with the IDAR whitening procedure to estimate dynamic functional connectivity profiles. This application showcases the superior sensitivity of the dynamic correlation profile characteristics to CSF $A\beta$ status when compared to stationary functional brain connectivity measurements, providing crucial insights into the development of non-invasive, MR-based AD biomarkers for the detection of early-stage AD.

The above chapters collectively contribute to the overarching theme of addressing the challenges of analyzing correlated data in modern applications. In the subsequent chapters, we delve deeper into the specifics of each methodology, providing a comprehensive exploration of their development, performance and application. The dissertation concludes with a discussion of potential future work in Chapter 5.

Chapter 2

REHE: FAST VARIANCE COMPONENTS ESTIMATION FOR LINEAR MIXED MODELS

2.1 Introduction

Linear mixed model are a convenient and powerful tool for analyzing correlated data, with a wide range of applications in scientific research. It is especially useful for genetic studies of complex traits, including heritability estimation [206], genome-wide association studies (GWAS) [5], and network-based pathway enrichment analysis (NetGSA) [198]. Variance components estimation is an essential step when applying linear mixed models and the restricted maximum likelihood (REML) approach is considered to be the gold-standard for this task [169]. REML works by iteratively maximizing the residual likelihood with respect to the variance component parameters. During each iteration, REML computes the inverse of two $n \times n$ matrices, where n is the sample size of the data set. As a result, the computation for REML quickly becomes prohibitive for large sample sizes, especially when the correlations among observations are non-sparse – such as between-subject correlations due to genetic relatedness [118]. Despite efforts to improve the computational efficiency of REML, such as average information REML [72], Monte Carlo REML [149], and REML based on grid search [114], the scalability of REML to very large data sets is often limited in many applications. On the other hand, consistency and asymptotic normality of REML estimates are large sample properties. In this paper we illustrate that REML can be numerically unstable and can also provide unreliable estimates and/or confidence intervals of variance components in some settings (Section 2.4, Appendix A.3.3 and Appendix A.3.4). Given these shortcomings, new approaches for fast and reliable estimators of variance components for linear mixed model are needed.

When computational efficiency is a primary concern, moment estimators of variance compo-

nents have frequently been used as alternatives to REML. These include analysis of variance (ANOVA), minimum norm quadratic unbiased estimation (MINQUE), and the Haseman-Elston (HE) regression estimator [176, 175, 91, 206]. These methods bypass the most time-consuming step in REML — the inversion of $n \times n$ matrices. An essential component of of these alternative methods to REML is to set up estimating equations by equating the mean squared errors to its expectation, the error variance. ANOVA, originated from ideas by R. A. Fisher in 1920s, has been well established for estimating variance components. The resulting estimators are minimum variance quadratic unbiased [82], and minimum variance unbiased under normality assumptions on the random effects and the errors [81, 83]. MINQUE [175], which can be viewed as an extension of the ANOVA method, is equivalent to the first iteration of REML [190]. It relaxes the assumption of normality using estimating equations that rely on initial values for the variance components. The HE estimator, first introduced in [91], has been recently used for linear mixed model variance component estimation in genetic studies [206, 253]. Its simple implementation and fast computation make it favorable when working with large and densely correlated data sets. A key limitation of these moment estimators, however, is that they do not guarantee non-negative estimates for the variance components. This leads to difficulties in both interpretation and downstream analyses.

To address the shortcomings of existing approaches in variance component estimation, we propose a new estimation method based on restricted Haseman-Elston (REHE) regression. REHE is computationally efficient, and ensures non-negative estimates of variance components. We demonstrate that the REHE estimates are comparably accurate to the REML estimates when REML performs well, and are robust in the settings where REML does not provide good estimates. To accommodate the need for a strictly positive variance estimates in some applications, we also propose REHE with resampling (reREHE), which provides positive variance components estimates with high probability. Furthermore, to facilitate inference, we propose bootstrap confidence intervals for REHE estimates. We demonstrate that the REHE boostrap confidence intervals are more robust than their REML counterparts. Finally, we also show that REHE in combination with correlation matrix sparsification [75], when applicable, can result in a substantial increase in computational speed for variance

component inference. REHE is both computationally efficient and flexible, and can be used across a broad range of study designs. Here we demonstrate the utility of REHE in three different contexts: heritability estimation, GWAS and NetGSA. We benchmark the proposed methods' performance with simulation studies, and illustrate their advantages using data from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL) [210, 41], where there is extensive non-zero genetic correlations among the 12,803 study subjects with GWAS data available, and The Cancer Genome Atlas (TCGA) breast cancer data set [216].

The rest of the chapter is organized as follows. In Section 2.2, we introduce the REHE estimator, discuss its properties and propose a bootstrap inference framework. We also introduce the reREHE estimator as an alternative to the REHE estimator. We demonstrate the performance of the REHE and reREHE estimators with real data applications in Section 2.3. In Section 2.4, we benchmark their performance with extensive simulation studies. Section 2.5 concludes the paper with discussions on the results and potential improvements. Additional results on REHE, reREHE and matrix sparsification are provided in the Appendix.

2.2 Methods

Consider a generic linear mixed model for an outcome vector Y of length n:

$$Y = X\boldsymbol{\beta} + \sum_{k=1}^{K} \sigma_k \boldsymbol{\gamma_k} + \sigma_0 \boldsymbol{\epsilon}. \tag{2.1}$$

Here, X is an $n \times p$ design matrix for p covariates, and $\boldsymbol{\beta}$ is a p-dimensional fixed effect coefficient vector. For $k = 1,\ 2, \ldots, K$, $\boldsymbol{\gamma_k}$ is a length n vector of random effects following $N_n(0, D_k)$, where each $n \times n$ matrix D_k defines one source of correlation among the observations, and is assumed to be known. The noise $\boldsymbol{\epsilon}$ is a length n vector following $N_n(0, \boldsymbol{I}_n)$. The parameters σ_k's ($k = 0,\ 1, \ldots, K$) are the variance components. For $k = 1, \ldots, K$, $h_k = \sigma_k^2 / \sum_{l=0}^{K} \sigma_l^2$ estimates the proportion of variation explained by D_k. In the context of genetic studies, a genetic relatedness matrix (GRM) is often used for D_1 and h_1 is viewed as a measure of heritability, i.e., the proportion of the total trait variation that is due to genetic variation.

Our main objective is to estimate the variance components σ_k^2 $(k = 0, 1, \ldots, K)$. For expositional clarity, we assume the model has no fixed effect, and the outcome vector Y is centered. We also assume $K = 1$ such that the correlations among the observations can be modeled by a single random effect $\boldsymbol{\gamma_1}$. However, our methods can be easily extended to models with fixed effects (Section 2.2.2), or with more than one random effects. Denoting $D_0 = \boldsymbol{I}_n$ and $\boldsymbol{\gamma_0} = \boldsymbol{\epsilon}$, the simplified form of model (2.1) becomes

$$Y = \sigma_0 \boldsymbol{\gamma_0} + \sigma_1 \boldsymbol{\gamma_1}. \tag{2.2}$$

2.2.1 The Haseman-Elston Regression

The Haseman-Elston (HE) regression approach [91, 206, 253] estimates the variance components via the method of moments. Specifically, since model (2.2) implies that $\mathrm{V}ar(Y) = \sigma_0^2 D_0 + \sigma_1^2 D_1$, n^2 estimating equations are constructed:

$$\mathrm{E}[Y_i Y_j] = \sigma_0^2 D_0^{ij} + \sigma_1^2 D_1^{ij}, \quad i, j = 1,\ 2, \ldots\ n,$$

where D_k^{ij} denotes the (i, j) entry of matrix D_k $(k = 0, 1)$. Estimation of the variance components can thus be recast as a linear regression problem. Let $\tilde{Y} = \mathrm{vec}\left(YY^\top\right)$ denote the vectorization of the $n \times n$ matrix $YY^\top$ by stacking its columns, $\tilde{X} = (\mathrm{v}ec(D_0), \mathrm{v}ec(D_1))$ and $\boldsymbol{\sigma^2} = (\sigma_0^2, \sigma_1^2)^\top$. We then have $\mathrm{E}\big(\tilde{Y}\big) = \tilde{X}\boldsymbol{\sigma^2}$. HE solves for variance components $\boldsymbol{\sigma^2}$ by linear regression, specifically, by minimizing the residual sum of squares

$$\left(\tilde{Y} - \tilde{X}\boldsymbol{\sigma^2}\right)^\top \left(\tilde{Y} - \tilde{X}\boldsymbol{\sigma^2}\right). \tag{2.3}$$

The resulting estimator has a closed form expression $\hat{\boldsymbol{\sigma}}^2 = (\tilde{X}^\top \tilde{X})^{-1} \tilde{X}^\top \tilde{Y}$. A nice property of the HE estimator is unbiasedness, even if we only use a subset of the n^2 estimating equations to compute the estimates; see the Appendix A.1.1 for details.

The computational complexity of HE is $O(Kn^2)$ compared to $O(n^3)$ for REML. When the sample size n is large and the number of variance components K is small, as is typically the case in practice, HE offers substantial improvement in computation over REML. However,

the ordinary least squares solution for variance components by HE is not guaranteed to be non-negative, leading to difficulties in downstream analyses and interpretation. In practice, negative estimates from HE are often truncated at zero; yet such naive truncation does not minimize the residual sum of squares in (2.3) within the parameter space ($\sigma_k^2 \geq 0,\ k = 0, 1$). In addition, as will be illustrated in simulation studies in Section 2.4 and the Appendix A.3.3, naive truncation-based HE estimates generally have larger mean square error than estimates by both REML and our proposed REHE method, which we introduce in the next subsection.

2.2.2 The Restricted Haseman-Elston Regression

To prevent negative estimation of variance components by HE, while still preserving its computational efficiency, we propose a new variance components estimation method, termed the restricted Haseman-Elston (REHE) regression. Similar to HE, REHE is a moment estimator which regresses the empirical covariance of the observations on pre-specified correlation matrices that encode sample relatedness. However, instead of the ordinary least squares estimate by HE, REHE finds the non-negative minimizer of the residual sum of squares, ensuring sensible estimation of variance components (Supplementary Figure S10). Following (2.3), the REHE estimates of the variance components are expressed as:

$$(\tilde{\sigma}_0^2, \tilde{\sigma}_1^2) = \arg \min_{(\sigma_0^2 \geq 0, \sigma_1^2 \geq 0)} \sum_{l=1}^{n^2} \left(\tilde{Y}_l - \tilde{X}_{l1}\sigma_0^2 - \tilde{X}_{l2}\sigma_1^2 \right)^2. \tag{2.4}$$

There is a closed form solution to (2.4) with only two variance components (Appendix A.2.1). With more than two variance components, iterative algorithms for non-negative least squares (NNLS) can easily solve (2.4) [125, 76, 120, 64]. The convexity of (2.4) guarantees the numerical solutions of different solvers converge to the same global minimizer. Using the R package `quadprog` (*v1.5-7*, 14), REHE estimation has approximately the same computational cost as HE, and is thus substantially faster than REML. In addition, the REHE estimator is consistent under mild conditions, and asymptotically normal when the correlation matrix for the random effect is sparse and block-diagonal, such as a sparse empirical kinship matrix; see the Appendix A.1.2 for details.

Algorithm 1 reREHE Approach Estimation.

for $b = 1$ to B **do**

(a) Sample with replacement from Y to obtain a length $[nr_s]$ vector $Y^{(b)}$ with sampling rate r_s, where $[x]$ is a function to round x to the nearest integer; subset the correlation matrices accordingly as $D_k^{(b)}$, for $k = 0,\ 1$.

(b) Compute the variance component estimates $\left(\tilde{\sigma}^{2(b)}_{0,re}, \tilde{\sigma}^{2(b)}_{1,re}\right)$ based on $Y^{(b)}$ and $D_k^{(b)}$'s using REHE (2.4).

end for

Estimate the variance components as $\tilde{\sigma}^2_{k,re} = \frac{1}{B}\sum_{b=1}^{B} \tilde{\sigma}^{2(b)}_{k,re},\ k = 0,\ 1$.

REHE with resampling (reREHE)

Estimates obtained from REHE are non-negative. However, in some applications, a zero estimate for the variance component may still make the interpretation and/or subsequent analyses challenging. To address this issue, we equip REHE with a resampling procedure that provides strictly positive variance component estimates with high probability. The resampling procedure utilizes repeated subsamples, which can further improve the computational efficiency of REHE. The resulting approach is termed reREHE.

The idea of subsampling for REHE is simple: instead of estimating the variance components based on all the observations at the same time, we only use a small subsample of the data. The full-sample-based estimates and inference are usually well approximated by statistics based on subsamples [172]. Similar subsampling techniques are also extensively used in stochastic gradient descent methods [152]. Recently, subsampling has also been used with HE-based estimating equations [253]. Our proposed reREHE procedure described in Algorithm 1 is unique in that it subsamples repeatedly to obtain the estimates. Although at a cost of reduced computational efficiency compared to using a single subsample, this resampling offers considerable advantages. By averaging the estimates from repeated subsamples, the reREHE estimates have much higher accuracy compared to estimates based on a single subsample. At the same time, we obtain strictly positive estimates, unless in extremely rare cases when all subsamples yield zero estimates. Other summaries, such as median, can also be used to summarize estimates from repeated subsamples. When the sampling rate r_s

and the number of subsamples B satisfy $r_s^2 B < 1$, reREHE achieves higher computational efficiency than REHE. On the other hand, choosing larger r_s and B results in more stable results. In the simulation studies and data applications, we chose $B = 50$ and varied r_s within $(0.05,\ 0.1)$ to achieve a balance between accuracy and computational efficiency.

REHE with Fixed Effects

The REHE estimation procedure can be easily modified to accommodate fixed effects. Consider the full model (2.1) where X is the design matrix with p covariates and $\boldsymbol{\beta}$ is the fixed effect coefficient vector. Let $P_X^{\perp} = I_n - X\left(X^\top X\right)^{-1} X^\top$ denote the projection matrix onto the orthogonal complement of the column space of X. We project the outcome Y and the random effects $\boldsymbol{\gamma_k}$'s (including the noise term $\boldsymbol{\gamma_0}$) as

$$Y^\dagger = P_X^{\perp} Y, \quad \boldsymbol{\gamma}_{\boldsymbol{k}}^\dagger = P_X^{\perp} \boldsymbol{\gamma_k}.$$

Recall that each random effect $\boldsymbol{\gamma_k}$ follows a normal distribution with zero mean and covariance D_k. Writing $D_k^\dagger = P_X^{\perp} D_k P_X^{\perp}$, model (2.1) becomes

$$Y^\dagger = \sum_{k=0}^{K} \sigma_k \boldsymbol{\gamma}_{\boldsymbol{k}}^\dagger, \quad \boldsymbol{\gamma}_{\boldsymbol{k}}^\dagger \sim N_n\left(0, D_k^\dagger\right), \quad k = 0, \ldots, K. \tag{2.5}$$

With model (2.5), we can directly apply the REHE approach as introduced in Section 2.2.2 to estimate the variance components. When the sample size n is large, computing the projected correlation matrices $D_k^\dagger$ is time-consuming. In genetic and genomics applications, the number of fixed effect covariates p is much smaller than the sample size n. We also have balanced design in many of these applications. In those settings we are able to obtain good estimates of the fixed effects, and we can directly use the original correlation matrices D_k instead of projected matrices $D_k^\dagger$. In such cases, as in our data applications and simulation studies, we expect results based on $D_k^\dagger$ to be very close to those based on D_k. We thus suggest using D_k for computational efficiency when estimating model (2.5) via REHE or reREHE.

If the fixed effect coefficients β are of interest, they can be estimated using ordinary least

Algorithm 2 Parametric Bootstrap Confidence Interval Construction for REHE.

Compute REHE estimates $\tilde{\sigma}_0^2,\ \tilde{\sigma}_1^2$ from (2.4), based on $\tilde{Y}$, D_0, D_1;
for $b = 1$ to B **do**
(a) Generate outcome vector $\tilde{Y}^{*(b)}$ from $N_n\left(0, \tilde{\sigma}_0^2 D_0 + \tilde{\sigma}_1^2 D_1\right)$;
(b) Compute REHE estimates $\tilde{\sigma}_0^{2(b)},\ \tilde{\sigma}_1^{2(b)}$ from (2.4), based on $\tilde{Y}^{*(b)}$, $D_0,\ D_1$;
end for

squares as $\hat{\boldsymbol{\beta}} = \left(X^\top X\right)^{-1} X^\top Y$, or weighted least squares as $\hat{\boldsymbol{\beta}} = \left(X^\top \hat{\Sigma}^{-1} X\right)^{-1} X^\top \hat{\Sigma}^{-1} Y$, where $\hat{\Sigma} = \sum_{k=0}^{K} \hat{\sigma}_k^2 D_k$ is based on previously estimated variance components $\hat{\sigma}_k^2$'s. While the resulting $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}$, one can iteratively update $\hat{\sigma}_k^2$'s and $\hat{\boldsymbol{\beta}}$ as in [206].

Constructing Confidence Intervals with REHE

To obtain confidence intervals for variance component estimates, we use a parametric bootstrap procedure as summarized in Algorithm 2.

Using the bootstrap samples of REHE estimates, $\left(\tilde{\sigma}_k^{2(b)}\right)_{b=1}^{B}$, for $k = 0,\ 1$, we can construct Wald-type confidence intervals as

$$\left[\tilde{\sigma}_k^2 - z_{\alpha/2} \times \mathrm{SD}\left(\tilde{\sigma}_k^{2(b)}\right), \tilde{\sigma}_k^2 + z_{\alpha/2} \times \mathrm{SD}\left(\tilde{\sigma}_k^{2(b)}\right)\right], \quad k = 0,\ 1,$$

where $z_{\alpha/2}$ is the $(1-\alpha/2) \times 100$th percentile of the standard normal distribution, and $\mathrm{SD}(\cdot)$ denotes the sample standard deviation over bootstrap samples. Wald-type confidence intervals are valid, provided that the estimates are normally distributed. We can also construct quantile bootstrap confidence intervals as

$$\left[\tilde{\sigma}_k^2 - \left(\tilde{\sigma}_k^{2(b)} - \tilde{\sigma}_k^2\right)_{1-\alpha}, \tilde{\sigma}_k^2 - \left(\tilde{\sigma}_k^{2(b)} - \tilde{\sigma}_k^2\right)_{\alpha}\right], \quad k = 0,\ 1,$$

where $\left(\tilde{\sigma}_k^{2(b)} - \tilde{\sigma}_k^2\right)_{\alpha}$ is the $(1-\alpha) \times 100$th empirical quantile of $\left(\tilde{\sigma}_k^{2(b)} - \tilde{\sigma}_k^2\right)_{b=1}^{B}$.

For small sample sizes, the quantile confidence intervals are expected to be more robust than their Wald-type counterparts. When the REHE estimator is close to be normally distributed under large sample size, Wald-type confidence interval might have higher accuracy based on

the same number of bootstrap samples. In simulation studies and real data applications, we chose the number of bootstrap samples $B = 50$ to balance computational time and confidence interval accuracy. Confidence intervals for functions of variance components, such as heritability, can be similarly obtained by transforming the bootstrap samples accordingly.

When the correlation matrix D_1 is dense and the sample size n is large, it can be computationally prohibitive to compute a matrix decomposition (through Cholesky or singular value decomposition) of D_1, which is required for the sampling step (a) in the Algorithm 2. To speed up the inference procedure in this case, we propose using correlation matrix sparsification [75]. The sparsification approximates the dense correlation matrix D_1 by a sparse block-diagonal matrix D_1^*, and largely accelerates matrix decomposition. Such a sparse block-diagonal structure is often approximately satisfied in genetic studies: for example, subjects are highly genetically correlated within the same family, and are remotely correlated across families (see Appendix Figure A.12 for an example of the kinship matrix). Note that a standard GRM calculated from genome-screen data may not be sparse. However, an empirical kinship matrix that reflects close familial relationship (and that can be sparsified) along with principal components for population structure has been proposed as an alternative to using a single GRM in linear mixed models [75, 99]. This approach has been used in various genetic studies with relatedness and population structure, including the HCHS/SOL [41], the Population Architecture using Genomics and Epidemiology (PAGE) study [233], and a recent analysis for the Trans-Omics for Precision Medicine (TOPMed) program [99]. [75] provides a detailed comparison of using a linear mixed model with a GRM, a dense kinship matrix (with PCs), and a sparse kinship matrix (with PCs). The genetic association testing results were shown to be very similar for all three approaches, but with a substantial increase in computational efficiency by using a sparse empirical kinship matrix. In the examples presented in Section 2.3, we use a dense empirical kinship matrix and PCs to model the correlation structure in the sample. We investigate the performance of sparsification in the additional simulation studies in Appendix A.3.1 and Appendix A.3.3. Our application of matrix sparsification for inference of variance components in genetic studies is novel, and is discussed in detail in Appendix A.2.2.

2.3 Applications

2.3.1 GWAS and Heritability Studies with HCHS/SOL Data

To evaluate the performance of REHE and reREHE in genetics applications, we conducted a genome-wide association analysis as well as a heritability analysis using a publicly available data set from the HCHS/SOL [210, 41]. The HCHS/SOL sample survey design consisted of a two-stage probability sample of households at four recruitment centers. For each center, census block groups were selected in defined communities, and households were sampled within census block groups [124, 41]. Among a total of 16,415 subjects enrolled in the study at baseline, 12,803 subjects were genotyped on a genome-wide single nucleotide polymorphisms (SNP) array containing 4,100,028 SNPs.

To perform a genome-wide association analysis, we tested the association between each SNP and the red blood cell count using a linear mixed model [41]. We first fit a null model, which was a linear mixed model without any genotype effect [5]. We included fixed effects covariates age, gender, cigarette use, field center indicator, genetic subgroup indicator, the first five principal components for population stratification effect, and individual sampling weights [41]. We removed subjects that have missing values for the above covariates, and included 12,502 subjects in the analysis. Correlations among subjects was modelled by three random effects: genetic relatedness represented by estimated kinship, membership of household, and membership of community group [41].

We separately applied REHE, reREHE and REML to estimate the null model with household membership matrix, community membership matrix, and estimated dense kinship matrix. For reREHE, we chose sampling rate $r_s = 0.1$ and used mean summary function based on 50 repeated subsamples. With n =12,502 subjects to be analyzed, REML took 23.9 minutes to estimate the null model, while REHE took only 2.4 minutes, a 10-fold improvement compared to REML. reREHE was similarly fast as REHE. Based on each estimated null model, we applied score tests for the association between each SNP and the red blood cell count [41], and compared the resulting p-values. We focus here on the 164 SNPs with p-values no larger than the family-wise error rate (FWER) threshold of 5×10^{-8} by at least one approach [54],

and presented p-values for all SNPs in the Appendix Figure A.11. As shown in Figure 2.1a and 2.1b, results based on REHE and reREHE have negligible differences from those based on REML. The correlations are all higher than 0.99 for the p-values on the $-\log_{10}$ scale between REHE and REML, as well as between reREHE and REML. This concordance among REML, reREHE and REHE is not surprising as the estimated variance components are similar (Figure 2.1c).

For the heritability analysis, we used the same data set and fit the same linear mixed null model as in the above genome-wide association analysis. The model was estimated based on REML and REHE separately. We obtained point estimates and confidence intervals for heritability (corresponding to the estimated dense kinship correlation matrix), proportions of variance explained by household membership, community block membership, and noise. REHE took 18.2 minutes to conduct the inference, compared to 23.9 minutes by REML. Heritability and variance proportions estimates, as well as the confidence intervals obtained by the REHE approach are all very similar compared to those obtained by REML (Figure 2.1c).

We used the R package `GENESIS` (*v2.14.3*, 40) for REML estimation and for conducting the genome-wide association analysis. All analyses were conducted on a computer with 2 $\times$ 6-core Intel Xeon CPU E5-2620 @ 2.00GHz 128GB RAM.

2.3.2 Network-based Pathway Enrichment Analysis with Breast Cancer Data

To further demonstrate that REHE and reREHE facilitate downstream analysis with fast variance component estimation, we performed a network-based pathway enrichment analysis, with a breast cancer data set from The Cancer Genome Atlas (TCGA) [216], preprocessed by [148]. The data set contains RNA-seq measurements for 2,598 genes from 100 genetic pathways, with 403 subjects from the ER positive subtype and 117 from the ER negative subtype.

Network-based pathway enrichment analysis tests for differential gene pathways associated with particular phenotypes under different conditions [147]. It assumes a linear mixed model for the relationship between gene expressions and the phenotype (see Appendix A.2.3 and

147 for details). Here, we compared the activities of 100 genetic pathways between the two ER groups. The ER group-specific gene networks — more specifically, the adjacency and influence matrices supplied to the linear mixed model — were estimated according to [147]. We estimated the variance components using REML, REHE and reREHE. For reREHE, we chose the sampling rate $r_s = 0.1$ for sampling the subjects, and additionally sampled gene entries within each subject with sampling rate 0.5 (see Appendix A.2.3 for details). The reREHE estimate was based on the mean of 50 repeated subsamples. After obtaining the variance components estimates, we tested for differences in the activity of each of the 100 genetic pathways [198].

We observed substantial improvement in computational efficiency of reREHE and REHE compared to REML: reREHE- and REHE-based analyses both took less than 2 minutes, whereas analysis with REML took over 1 hour. Comparing the resulting p-values, REHE and reREHE produce slightly more conservative p-values than REML (Figure 2.2). Moreover, REHE yields a zero estimate for the noise variance component, the reREHE estimate is 0.0120, while the REML estimate is 0.266. The corresponding network variance estimates are also quite different: 0.273 by REML, 0.534 by reREHE and 0.610 by REHE. This may be an evidence that the variation explained by the network is much larger than the variation from noise in the true model. As illustrated in our additional simulation studies in Appendix A.3.4, REML may yield unreliable estimates under similar settings. We should thus take extra caution when interpreting REML-based estimates and test results in this application.

We conducted analyses for NetGSA using the R package `netgsa` (*v3.1.0*, 147) on a computer with 2 × 6-core Intel Xeon X5650 @ 2.67 GHz, 96GB RAM.

2.4 Simulation Studies

2.4.1 Simulation Settings

To benchmark the improvement of REHE and reREHE over HE and REML for variance components and heritability estimation, we generated synthetic data based on the HCHS/SOL design [210, 41]. HE and reREHE approaches were implemented only for point estimation comparison. We truncated negative HE estimates at zero. For reREHE, we used $B = 50$

repeated subsamples, and chose sampling rates $r_s = 0.05$ (reREHE 0.05) and $r_s = 0.1$ (reREHE 0.1). Point estimates were evaluated in terms of the root mean squared error (RMSE). We constructed Wald-type (REHE-Wald) and quantile-type (REHE-quantile) confidence intervals at 95% level for REHE estimates, and compared their performances with REML-based confidence intervals in terms of coverage and interval width.

We simulated data based on the linear mixed model (2.2). We used sample size $n \in \{3{,}000,$ 6,000, 9,000, 12,000$\}$. For each sample size, we set the true values of the variance components to $(\sigma_0^2,\ \sigma_1^2) \in \{(0.1,\ 0.1),\ (0.04,\ 0.1),\ (0.1, 0.04),\ (0.01,\ 0.1),\ (0.1,\ 0.01)\}$. These values were chosen based on previous simulation study settings [206] and estimates from real data applications [62]. For each sample size n selected, we generated the correlation matrix D_1 as a random sub-matrix of the kinship correlation matrix from the HCHS/SOL data set. For example, for n =3,000, we subsampled 3,000 out of 12,803 subjects without replacement, and used the corresponding (subsample) kinship correlation matrix as D_1. Under each scenario, we ran 200 replicates. Computation was performed on a computer with 2 $\times$ 6-core Intel Xeon CPU E5-2620 @ 2.00GHz 128GB RAM.

We also conducted additional simulation studies to compare the approaches under different correlation structures; details of these experiments can be found in Appendix A.3.3.

2.4.2 Simulation Results

Simulation results clearly demonstrate the improvement in computational efficiency by REHE compared to REML. For point estimation, REHE was over 50 times faster than REML (Figure 2.3a). At the same time, REHE does not compromise estimation accuracy. Figure 2.3b and 2.3c show that REHE estimates of both the variance components and heritability are very close to those obtained by REML. Another advantage of REHE is that it corrects negative variance estimates from HE. To quantify this difference, We calculated the proportion of simulation replicates resulting in negative HE estimates (before zero-thresholding). This proportion reaches 23% with n =3,000, $(\sigma_0^2 = 0.01,\ \sigma_1^2 = 0.1)$, but reduces to 1.5% at n =12,000. As pointed out before, REHE automatically corrects the issue of negative estimates without hard-thresholding. Besides, REHE has lower RMSE for

point estimates when HE estimates are likely being negative (Figure 2.3c).

The simulation results confirm that reREHE can provide strictly positive estimates with high probability. By providing a positive variance estimate where REHE gives a zero estimate (up to 23% of the simulation repetitions), reREHE is helpful for interpretation and downstream analysis, especially under small sample sizes. As shown in Figure 2.3b, reREHE based estimates have smaller RMSE than all other methods at sample size n =3,000. With larger samples, the RMSE of reREHE is comparable to other methods under some settings (Figure 2.3b), but is much larger in other settings (Figure 2.3c, Appendix A.3.1 and Appendix A.3.3). Setting a higher subsampling rate (0.1 compared with 0.05) reduces RMSE (Figure 2.3b and 2.3c), but comes at the cost of reduced computational efficiency — reduction in computation time compared to REHE diminishes from 67% to 10% (Figure 2.3a).

Turning to inference of variance components and heritability, both REHE based quantile-type and Wald-type confidence intervals provide reasonably good coverage with comparable interval width to REML confidence intervals (Figure 2.4). The empirical coverage is close to nominal level under most cases (Figure 2.4a and 2.4b), considering a Monte Carlo error of 0.03 based on 200 simulation replicates. REHE quantile-type intervals generally have better coverage than Wald-type intervals when the true variance components are very different (Figure 2.4b). In terms of confidence interval width, quantile-type REHE confidence intervals are generally narrower than Wald-type, and both are comparable to REML-based intervals (Figure 2.4c and 2.4d). Inference based on REHE is more time-consuming than REHE-based point estimation; however, it still achieves 50% reduction of computation time compared to REML (Figure 2.3a).

Finally, in some simulation settings, we noticed that REML confidence intervals may suffer from under-coverage. For instance, with n =3,000 samples, when one variance component is substantially smaller, the coverage of REML confidence intervals drop below 87% (Figure 2.4b). In other settings, REML confidence intervals even have coverage below 60%, and have little improvement with increasing sample size (Appendix A.3.3). Another concern is the numerical stability of REML, which fails to provide a confidence interval if the estimate of any variance component becomes zero during the iterative updates. We noticed

frequent occurrence of this issue when the true variance components are unbalanced and the sample size is small — for n =3,000 and $(\sigma_0^2, \sigma_1^2) = (0.1, 0.01)$, REML is unable to provide a confidence interval in 17.5% of the replicates. This proportion increases to 30.5% in other settings (Appendix A.3.3). We view these two issues as a warning sign for REML-based inference in real applications, especially when the underlying variance components are very different and the sample size is small. In contrast, REHE-based inference is robust across different settings with valid confidence intervals and acceptable empirical coverage.

The above simulation results are supported by evidence from additional simulation studies in Appendix A.3.3.

2.5 Discussion

We proposed REHE for fast estimation of variance components in linear mixed models. This new approach is motivated by large-scale genetic studies, including HCHS/SOL, but is applicable more broadly to a wide range of study designs. Through simulation studies and data applications, we demonstrated its substantial gain in computational efficiency over REML, with little compromise in point estimation accuracy. We also showed in several simulation settings that REHE can be more robust than REML for estimating the variance components (Appendix A.3.4). Compared to HE, REHE corrects the issue of negative estimates, and offers potentially large gains in estimation accuracy. Therefore, REHE can be superior compared to HE and be a good alternative to REML for point estimation of variance components in linear mixed models.

We also proposed reREHE based on the resampling technique to produce strictly positive variance component estimates with high probability. Strictly positive estimates are more interpretable and may be more appealing for downstream analyses. Though reREHE estimates may have lower accuracy than REHE, the magnitude of the increase in RMSE was small in our experiments. We have also seen in the real data application that based on a subsampling rate of 0.1 and 50 subsamples, reREHE-based downstream analysis results are close to REML-based results. With suitably chosen subsampling rate and number of subsampling replicates, reREHE can achieve higher computational efficiency than REHE.

As mentioned previously, one can also use the median of the subsample results as the reREHE estimate. We explored this choice in the Appendix A.3.2 and Appendix A.3.3. When the underlying variance components are very different, median-based reREHE estimates generally have smaller RMSE; otherwise mean-based reREHE performs better. However, median-based reREHE is more likely to yield zero estimates. A post hoc selection of the summary function can be made after observing the distribution of the subsample estimates based on reREHE.

As illustrated in the genome-wide association and pathway enrichment analysis examples in Section 2.3, in many applications, only variance component estimates are needed for downstream analyses. The computational burden of REML-based estimation prohibits these analyses on large data sets. Restricting the analysis to subsets of data reduces reliability and may yield contradictory conclusions. Given the fast and reliable estimates by REHE and reREHE in large data sets, we see great potential for their application in areas that only require point estimation of variance components.

When confidence intervals are also of interest, REHE remains a competitive alternative to REML for its robustness and numerical stability. As illustrated in our simulation studies, when the sample size is small and the true variance components are unbalanced, REML based inference may suffer from numerical instability and/or poor coverage. REHE consistently provides valid inference across all settings. Therefore, when the true variance components are expected to be very different and the sample size is not large, we recommend REHE over REML if inference on variance components is needed, as REML-based inference results may be unreliable.

Constructing confidence intervals for variance components when sample size is large (e.g. n >10,000) is inherently computationally challenging. REHE only offers marginal improvements in computational efficiency over REML when it comes to inference. However, we can improve the computational efficiency for REHE confidence interval by parallelizing the bootstrap procedure. An alternative acceleration approach is to use correlation matrix sparsification (75, Appendix A.2.2). In the Appendix A.3.1 and Appendix A.3.3, we explored the application of sparsification for constructing confidence intervals. Our conclusion is that sparsification improves computational efficiency in large sample settings (n >12,000);

however, it may result in less robust confidence intervals for both REHE and REML. We did not explore application of sparsification to linear mixed models with more than one random effect. We expect a much larger sample size beyond which sparsification would show improvement in computational efficiency.

We did not explore confidence interval construction for reREHE estimates in this paper. Due to the repeated subsampling procedure of reREHE, an analytical expression for the confidence intervals is not trivial. The parametric bootstrap procedure for REHE confidence interval construction is readily extendable to reREHE, which we expect to have similar performance as REHE confidence intervals. However, the computational burden will also be similar to those of REHE confidence intervals. Future research should explore fast inference procedure for REHE and reREHE estimates.

We may consider an alternative strategy for obtaining positive variance component estimates by log-transforming the variance components. While REML can directly estimate variance components on the natural scale, empirical evidence suggests superior performance when estimating them on the log scale, particularly when true variance component values are close to zero [212]. However, a systematic evaluation of the REML approach for estimating variance components on the log scale compared to the natural scale is currently lacking, presenting an intriguing avenue for future research. While log-transformation reparametrization proves effective in addressing the issue of negative estimates, moment-based approaches such as HE and REHE can only provide estimates on their natural scale. The incorporation of log-transformation into the estimating equations poses nontrivial challenges, opening up an intriguing area for further investigation.

It is also important to note that the Wald-type confidence interval produced by the REHE framework may include negative values, which is undesirable. One solution is to employ the proposed quantile-based confidence interval derived from parametric bootstrapping in Section 2.2.2, ensuring coverage limited to non-negative values. Alternatively, we could first log-transform the resulting bootstrap samples of REHE estimates, calculate the standard deviation of the log-scale bootstrap samples, construct the Wald-type confidence interval on the log scale, and subsequently transform it back to the natural scale to guarantee

non-negativity. Subsequent studies may investigate the performance of such log-transformed Wald-type confidence intervals.

Data Availability Statement

The HCHS/SOL data is available under accession numbers dbGaP: phs000880.v1.p1 and phs000810.v1.p1. The TCGA breast cancer data set is publicly available at `https://github.com/drjingma/NetGSAreview`. The proposed methods are implemented with codes written in the R language, which are available at `https://github.com/yuek9/REHE`.

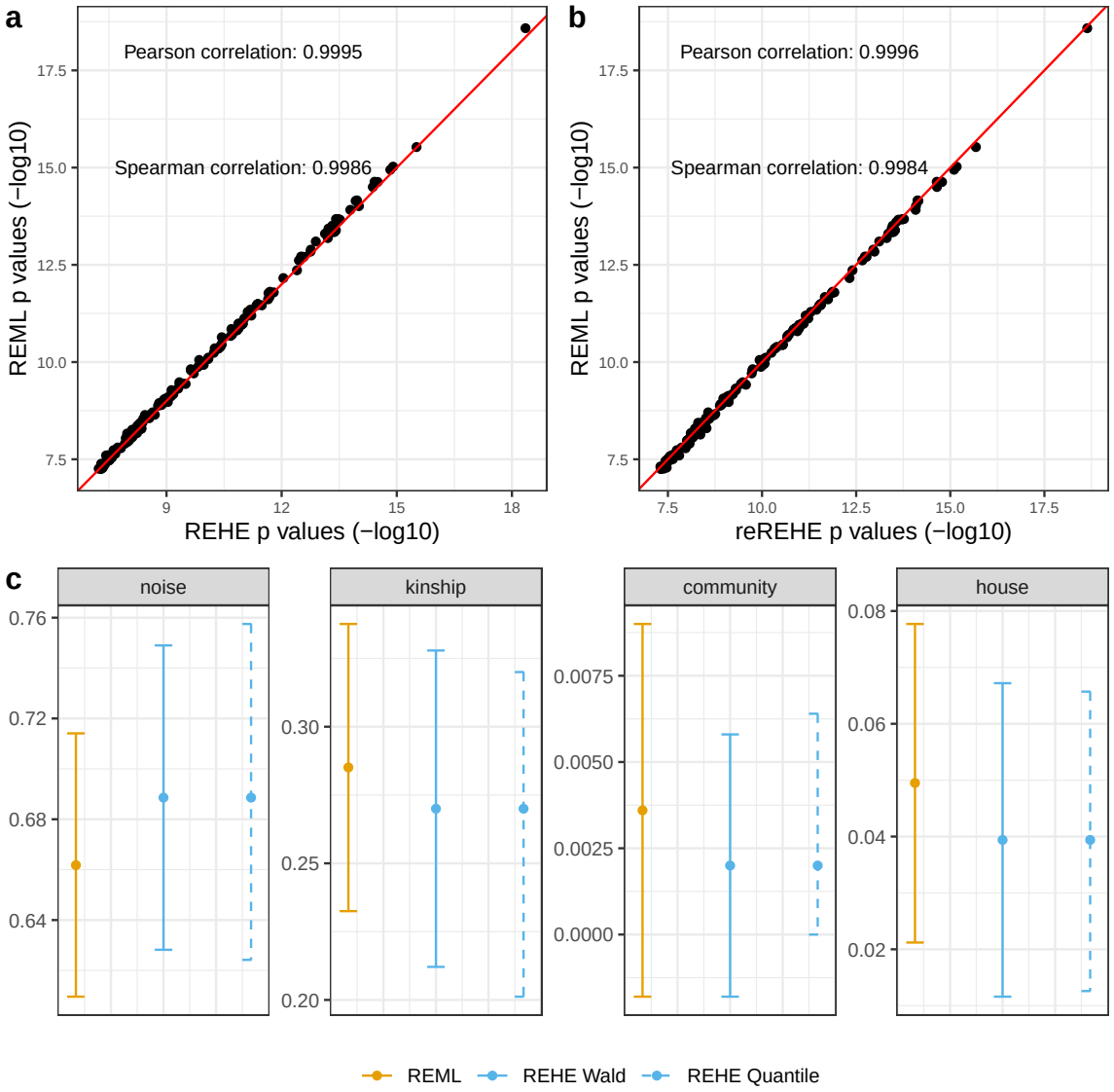

Figure 2.1: Results of genome-wide association testing analysis and heritability analysis with a HCHS/SOL data set. Association score test p-values ($-\log_{10}$ scale) based on: **a** - REML against REHE estimated null models; **b** - REML against reREHE estimated null models. Only 164 SNPs with resulted p-values no larger than 5×10^{-8} by at least one approach were presented. **c** - Dots represent point estimates of proportion of variance attributed to noise, kinship (heritability), community membership and household membership; bars represent corresponding confidence intervals. Results by REHE and REML are displayed. Two types of REHE-based confidence intervals are presented: Wald-type confidence intervals (REHE Wald), and quantile-type confidence intervals (REHE Quantile).

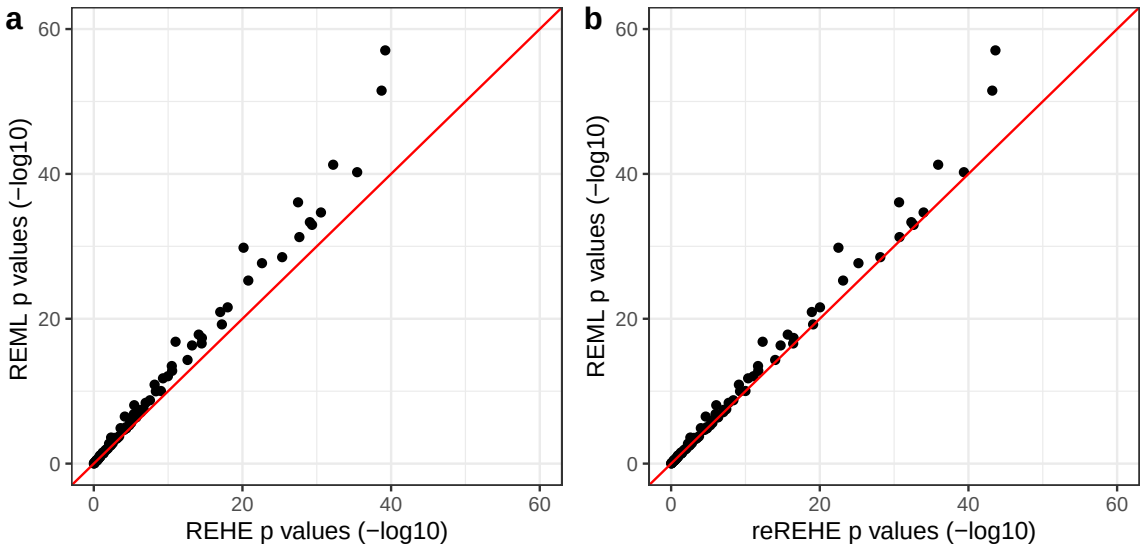

Figure 2.2: Results of network-based pathway enrichment analysis based on a breast cancer data set. p-values ($-\log_{10}$ scale) of t-tests for group difference of each gene pathway: **a** - compare REHE results against REML results; **b** - compare reREHE results against REML results. Two out-of-range data points are omitted from the plots, which correspond to: Glycosphingolipid biosynthesis - lacto and neolacto series pathway, with p-value 1.09×10^{-307} by REML, 5.67×10^{-276} by REHE, 3.74×10^{-304} by reREHE; and Caffeine metabolism pathway with p-value 3.97×10^{-201} by REML, 2.60×10^{-179} by REHE, and 2.13×10^{-198} by reREHE

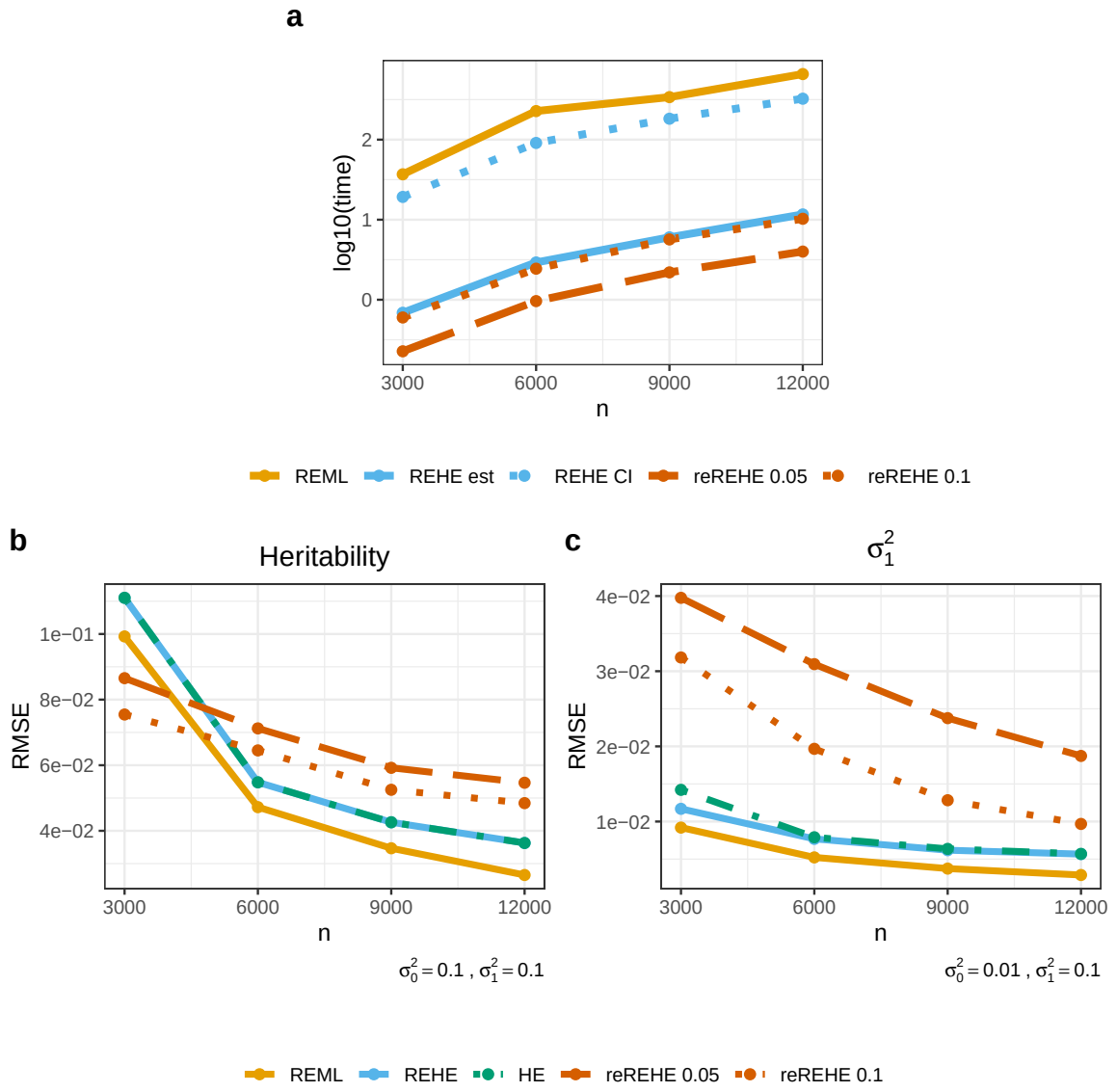

Figure 2.3: Simulation results for REHE and reREHE compared to REML. **a** - CPU time in seconds ($\log_{10}$ scale): time is presented separately for fitting the model using REML (REML), for only computing point estimates by REHE (REHE est), for constructing confidence interval for variance components with REHE (REHE CI), for only computing point estimates by reREHE with subsampling rate $r_s = 0.05$ (reREHE 0.05) and reREHE with subsampling rate $r_s = 0.1$ (reREHE 0.1). **b** - Root mean squared error (RMSE) for heritability estimation; simulation was based on true values $\sigma_0^2 = 0.1$, $\sigma_1^2 = 0.1$, where σ_0^2 is the variance for noise, and σ_1^2 is the variance for the random effect. **c** - RMSE for σ_1^2 estimation; simulation was based on true values $\sigma_0^2 = 0.01$, $\sigma_1^2 = 0.1$.

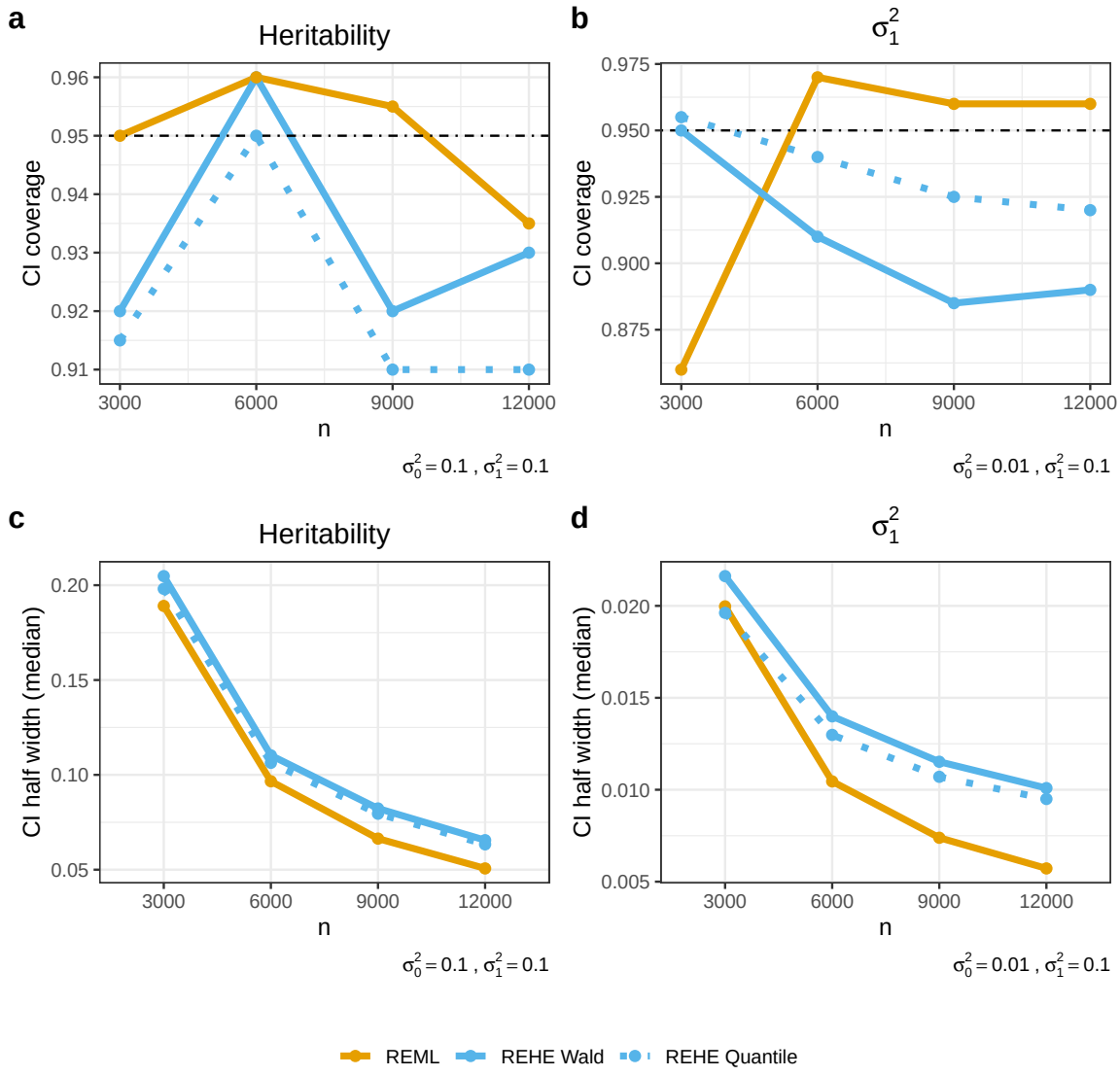

Figure 2.4: Simulation results for confidence interval performance in terms of coverage and width. σ_0^2 is the variance for noise, and σ_1^2 is the variance for the random effect. **a** - Coverage for heritability confidence interval (CI); simulation was based on true values $\sigma_0^2 = 0.1$, $\sigma_1^2 = 0.1$. Monte Carlo error of 0.03 is expected for 200 simulation replications. **b** - Coverage for σ_1^2 CI; simulation was based on true values $\sigma_0^2 = 0.01$, $\sigma_1^2 = 0.1$. Monte Carlo error of 0.03 is expected for 200 simulation replicates. **c** - Line charts of median half width of heritability CI with increasing sample sizes; simulation was based on true values $\sigma_0^2 = 0.1$, $\sigma_1^2 = 0.1$. **d** - Line charts of median half width of σ_1^2 CI with increasing sample sizes; simulation was based on true values $\sigma_0^2 = 0.01$, $\sigma_1^2 = 0.1$.

Chapter 3

INFERENCE FOR HETEROGENEOUS GRAPHICAL MODELS USING DOUBLY HIGH-DIMENSIONAL LINEAR-MIXED MODELS

3.1 Introduction

Gaussian graphical models (GGMs) capture conditional dependence relations among a set of variables, $\{Y_1, Y_2, \ldots, Y_p\}$ via a graph $G = (V, E)$ with node set $V = \{1, 2, \ldots, p\}$ and edge set $E \subset V \times V$. For a mean zero multivariate normal vector $Y = \{Y_j : j \in V\}$ with covariance matrix Σ, the conditional dependence structure, and correspondingly, the edge set E, can be characterized by the nonzero entries of the inverse covariance matrix $\Omega = \Sigma^{-1}$. Specifically, two random variables Y_j and Y_k are conditionally independent if and only if $\Omega_{j,k} = 0$. The value of $\Omega_{j,k}$ can be viewed as the weight of the edge (j, k). Therefore, the problem of inferring the graph structure is effectively an (inverse-)covariance selection problem and has been extensively studied in high-dimensional settings, with applications in neuroscience [162, 245, 154] and genomics [239, 122, 250], among other fields. Two of the most popular approaches for independent observations are the graphical lasso [242, 65] and neighborhood selection [151]. Other graph structure learning methods include greedy search [107, 177, 24], structured regularization [32, 46] and regularized score matching [135]. Recent developments in high-dimensional graphical modeling have also considered non-Gaussian observations [139, 236, 224, 238, 241] and functional data [207, 174].

Estimates of graphical models provide valuable information about the strength of connectivity among variables. However, the uncertainty in these estimates needs to be quantified in order to answer scientific questions of interest — for instance, in brain functional connectivity studies, whether the estimated non-zero dependency between two brain regions indicates a real connection, or if the observed difference in the brain connectivity structures between two patient groups indicates a true population-level difference [196]. As a result, inference

for graphical models has received increasing attention in recent years. Examples include multiple testing with asymptotic control of false discovery rates [141], and direct testing of edge weights based on the asymptotic normality of different (de-biased) ℓ_1-regularized estimators [109, 110, 179]. See [52, 111] for a detailed review.

This chapter is motivated by the problem of inferring the graph structure of functional connectivity networks from multi-level functional magnetic resonance imaging (fMRI) data [204]. A prime example is the resting-state fMRI data from the Human Connectome Project (HCP); one of HCP's main goals is to characterize the functional neural connections in healthy individuals [221], and reliable inference for such connections is paramount to understanding the brain physiology [211]. Figure 3.1 illustrates our application setting: For each subject (i.e., level) $i = 1, \ldots, n$, resting-state whole-brain fMRI signals give an indirect measure of the neuronal activation levels at multiple brain locations over time. Standard pre-processing leads to spatially distributed maps that define a set of brain regions $V = \{j : j \in 1, \ldots, p\}$, with details to be described in Section 3.6.1. Each brain region has an associated fMRI signal describing its activation pattern over time, denoted by Y_j^i. Without loss of generality, we center the observations for each brain region, Y_j^i, at zero.

Learning the functional connectivity graph structure from Y_j^i presents two primary challenges: (i) fMRI observations over time for a single brain node typically exhibit serial correlation; and (ii) the data have a *clustered or multi-level structure*, where each cluster, or level, corresponds to the observations of a specific subject. While the serial correlation can be mitigated through various whitening procedures including model-based pre-whitening procedures [165, 234] or simple down-sampling approaches, the complications due to the clustered structure of the data have not been extensively studied in this setting. Many neuroscience studies ignore the *heterogeneity* inherent in multi-level data and simply infer a single graphical model for all subjects [55]. This assumes a fixed dependence structure for all the subjects, which is contrary to a growing body of evidence that points to considerable subject-level heterogeneity in functional connectivity networks [53, 154, 157]. Such heterogeneity cannot be easily addressed with resampling techniques e.g., [160], which lack theoretical guarantees for type-I error control and are computationally demanding when the number of brain regions

is large. Another popular approach is to employ a two-stage strategy: in the first stage, separate graphical models are inferred for each subject; in the second stage, individual-level summary statistics are used for group-level analysis [160, 10, 47, 155]. P-value aggregation via Fisher's method [47] and t-test based on individual-level statistics [155] are two typical examples. While straightforward, such methods ignore any shared brain network information across subjects, which can lead to inefficient estimation and inference. More importantly, they can lead to conflicting conclusions from different second-stage aggregation choices, as well as erroneous conclusions due to not properly accounting for the uncertainty in first-stage estimates [39]. We illustrate the limitations of these two-stage approaches through a simple toy example depicted in Figure 3.1 and a comprehensive simulation study in Appendix B.5.3. In the toy example, we infer the connectivity between a given node and six other nodes using the neighborhood selection approach from a conditional model perspective (see Figure 3.1 for details). Specifically, as shown in Table 3.1, the fixed GGM method that ignores the heterogeneity can result in false discoveries, even when the population-level average network is of interest. Two-stage approaches may lead to conflicting conclusions, and may result in both inflated type-I errors and/or reduced power. Similar results are observed in the simulation study detailed in Appendix B.5.3.

The analysis of resting-state fMRI signals introduces additional challenges to brain network inference. In contrast to task-based fMRI data, where a shared task pattern enables the alignment of observations across subjects, resting-state fMRI data lacks a clear correspondence between time points across subjects. This absence of alignment renders methods such as functional graphical model approaches [207, 174] impractical, as these methods rely on the assumption of aligned underlying signals or functions. Therefore, to bridge the gap in existing approaches for inferring population-level brain connectivity networks while accounting for subject-level heterogeneity, in Section 3.2 we propose a *mixed effect Gaussian graphical model*. Utilizing a neighborhood-based estimation strategy similar to [160], for each edge, the proposed approach models the subject-level coefficients as random realizations centered around a population mean. The key difference is the estimation and inference approach: we recast the resulting model as a *doubly high-dimensional linear mixed model*, where the number

of fixed and random effects parameters can be larger than the sample size. In addition to the doubly high-dimensional structure, the fixed and random design matrices in the corresponding linear mixed models also have considerable overlap. These factors significantly complicate the theoretical analysis, rendering existing approaches inadequate. We overcome these challenges by utilizing penalized estimation and inference strategies, as well as tools from random matrix theory. We obtain consistent estimators and establish a valid inference framework for the corresponding parameters in Section 3.3. We also provide consistent estimation of the mixed effect variance components in Section 3.4.1 and demonstrate the performance of the proposed approach via extensive simulations in Section 3.5 and analysis of HCP fMRI data in Section 3.6.

The proposed model in Section 3.2 naturally accounts for the heterogeneity in individual-level connectomes. However, heterogeneity also arises in many other applications, including data harmonization [240, 237] and integration of multiple batches of genomic data [249]. The proposed estimation and inference framework for doubly high-dimensional linear mixed models can also be utilized in such problems. Moreover, while in this chapter we focus on estimation of undirected GGMs using data without serial correlation within each level, we show in Section 3.4.2 that our proposed method can be extended to inferring graphs based on a first-order Vector Autoregressive (VAR) model, and is thus able to account for (weak) serial correlations.

3.2 Method

3.2.1 Notations

We denote an $m \times m$ identity matrix by I_m. For a matrix A, we denote by $A_{j,k}$ the (j,k) entry of A, by A_j the jth column of A, and by A_{-j} the sub-matrix of A obtained by dropping the jth column. Similarly, we use index sets S, $\{j,k\}$ and $1:j$ to denote multiple columns/entries in a matrix/vector, and use $-S$, $-\{j,k\}$ and $-\{1:j\}$ to denote a sub-matrix/sub-vector obtained by removing the indicated columns/entries. We denote by $\|A\|_2$ the matrix norm of A, which is the maximum singular value of A. The Frobenius norm of A is denoted by $\|A\|_F$, which is equal to $\sqrt{\mathrm{tr}(AA^\top)}$ with $\mathrm{tr}(A)$ denoting the trace of A. We use $\sigma(A)$,

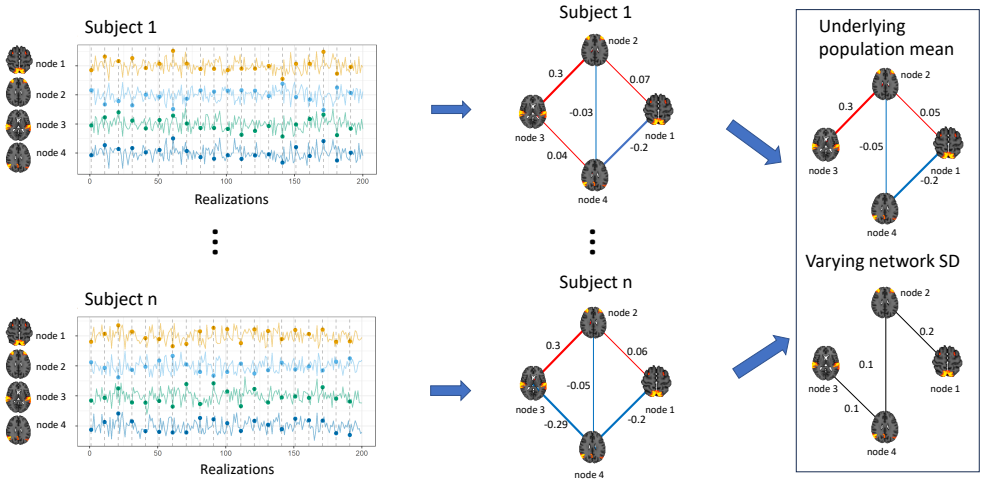

Figure 3.1: Illustration of the proposed functional connectivity brain network analysis. **Left panel**: fMRI signals measuring the activation level at each brain node for individual subjects. **Middle panel**: Subject-level brain networks inferred from the associated fMRI data. The lines between the brain nodes depict connectivity, with the numbers denoting the connectivity strength, red lines representing positive connectivity, and blue lines representing negative connectivity. **Right panel**: The output of our proposed model. This is the underlying population-level brain network and the associated network of standard deviations (SD) summarizing inter-subject variation around the population-level brain network.

$\sigma_{\min}(A)$ and $\sigma_{\max}(A)$ to represent the singular values, the minimum singular value, and the maximum singular value of A. For a set of values $\{u_l\}_{l=1}^K$ and a set of matrices $\{A_l\}_{l=1}^K$, we let $\text{diag}\left(\{u_l\}_{l=1}^K\right)$ be the diagonal matrix with (l,l) entry u_l, and let $\text{diag}\left(\{A_l\}_{l=1}^K\right)$ be the block-diagonal matrix with the lth block A_l.

A random variable U is sub-Gaussian with parameter u if $\forall\, t \in \mathbb{R}$, $\mathbb{E}\left(e^{tU}\right) \leq \exp(ut^2/2)$. We define $\text{SG}(u)$ as the class of all sub-Gaussian random variables with mean 0 and parameter u. A random vector V is sub-Gaussian with parameter v if for any vector x with $\|x\|_2 = 1$, we have $x^\top V \in \text{SG}(v)$. We denote the class of such sub-Gaussian random vectors by $\text{SGV}(v)$.

For two scalars a and b, we write $a \asymp b$ if $c_1|b| \leq |a| \leq c_2|b|$ for some positive constants c_1, c_2. We write $a \vee b = \max(a,b)$, and $a \wedge b = \min(a,b)$. We use $a = O(b)$ to indicate that $a \leq c_1 b$ for some constant $c_1 > 0$, and use $a = o(b)$ to mean that $a/b \to 0$. Throughout the chapter,

we use $c,\ c_0,\ c_1, \ldots$ to represent positive constants, whose values may vary from line to line.

3.2.2 Problem Setup

Suppose that for each subject $i = 1, \ldots, n$ the observed data matrix, $Y^i \in \mathbb{R}^{m \times p}$, includes m observations for each of the p nodes. Without loss of generality, we assume the observations for each node are centered at zero. The assumption of equal number of observations per subject is made for simplicity and our results continue to hold if the ith subject has m_i observations, as long as $c_1 \max_i(m_i) \leq \min_i(m_i) \leq c_2 \max_i(m_i)$ for some constants $c_1, c_2 > 0$. The population-level connectivity network is characterized by the inverse covariance matrix Ω, and subject-specific conditional independent networks are denoted by Ω^i.

We would like to infer the edges in the population-level network, while accounting for subject-level heterogeneity. To this end, we propose a neighborhood-based method where we model the subject-specific edge weight $\Omega^i_{j,k}$ as *random variables* centered at the population-level edge weight $\Omega_{j,k}$. Specifically, we assume the following neighborhood-based model for a node $j \in V$:

$$Y^i_j = \sum_{k=1, k \neq j}^{p} Y^i_k b^i_{j,k} + \epsilon^i_j, \quad i = 1, \ldots, n, \tag{3.1}$$

where given $Y^i_k, k \neq j$, the subject-level connectivity coefficients, $b^i_{j,k}$, have mean $b_{j,k}$ and variance $\sigma^2_{j,k}$. The coefficients $b^i_{j,k}$ and their mean $b_{j,k}$ are proportional to the true edge weights $\Omega^i_{j,k}$ and $\Omega_{j,k}$, respectively. The randomness in $b^i_{j,k}$ captures the subject-level heterogeneity, while its mean $b_{j,k}$ captures the shared dependence structure across subjects. Our main goal is to test whether a pair of nodes (j, k) are functionally connected at a population level, i.e., $H_0 : b_{j,k} = 0$.

The model in equation (3.1) can be seen as a linear mixed model (LMM), formulated as:

$$Y^i_j = Y^i_{-j}\beta_j + Y^i_{-j}\gamma^i_j + \epsilon^i_j, \quad i = 1, \ldots, n. \tag{3.2}$$

Here, we treat node j as the outcome variable and the associated vectors $Y^i_j \in \mathbb{R}^m$ as the

outcome vector of observations. We treat the rest of the $p-1$ nodes as covariates. The matrix $Y^i_{-j} \in \mathbb{R}^{m\times(p-1)}$ serves as both the fixed and random effect design matrices. The fixed effect coefficients $\beta_j \in \mathbb{R}^{p-1}$ correspond to $\{b_{j,k}\}^p_{k=1,k\neq j}$, which represent the (scaled) edge weights $\Omega_{j,-j}$. The random effect coefficients $\gamma^i_j = \left\{b^i_{j,k} - b_{j,k}\right\}^p_{k=1,k\neq j}$ represent the subject-level variation of these $p-1$ (scaled) weights. The hypothesis $H_0 :\ b_{j,k} = 0$ is thus equivalent to $H_0' :\ \beta_{j,k} = 0$, allowing us to recast the graphical selection problem as that of estimating and inferring the fixed effect coefficients in an LMM. Since the number of brain nodes p is typically large in brain connectivity studies, the resulting LMM is *doubly high-dimensional*, i.e., both fixed and random effects are high-dimensional.

Despite its increasing relevance in applications, rigorous estimation and inference procedures for doubly high-dimensional LMMs are lacking. Methods for LMMs with high-dimensional fixed effects and low-dimensional random effects [71, 136, 23] are not readily extendable to this setting. In addition, while Expectation-Maximization has been proposed for estimation [154], the consistency of the resulting estimator has not been thoroughly examined. Likewise, while consistent, valid inference for the estimator of [134] has not been explored. Most related to our setting is the recent work by Li, Cai and Li [132], denoted *LCL* hereafter. The work of *LCL* proposes a de-biased LASSO-based inference framework for doubly high-dimensional linear mixed models; however, as common in the LMM literature, it assumes a form of independence between the fixed and random effect design matrices—the fixed effect design matrix is assumed to have zero mean conditional on the random effect design matrix. This restrictive assumption constitutes a key shortcoming in our graphical modeling application, where the fixed and random effect design matrices are identical, leading to a violation of the zero conditional mean assumption by *LCL*. This assumption can also be restrictive in many other applications, whenever the fixed effects and the random effects share a non-empty set of covariates [134]. Moreover, the *LCL* framework requires the number of random effects to be no larger than the number of observations per subject, which further limits its applicability.

Motivated by the neighborhood-based model in (3.1), in this chapter, we develop a new estimation and inference framework for doubly high-dimensional LMMs in (3.2). To this end, we propose a penalized estimation and inference framework for the fixed effect coefficients.

We make two key extensions to the work of [132] that enable valid inference of heterogeneous GGMs: (i) our approach accommodates a larger number of random effects than the number of observations per subject; and (ii) it does not require conditional independence and only imposes minimal assumptions on the relationship between the fixed and the random effects. Through these extensions, our model provides the first mixed-model inference framework for learning functional connectivity networks in multi-level settings.

3.2.3 Model and Methods

To achieve consistent estimation for the doubly high-dimensional LMM in equation (3.2), we assume the true β_j^* coefficients are sparse, with support S_j and cardinality $s_j = |S_j| \ll p$. Let Σ^i be the $p \times p$ subject-specific covariance matrix for subject i. Throughout this chapter, we assume that, conditional on the covariance matrices Σ^i, the matrices Y^i are independent and follow a matrix normal distribution $Y^i \mid \Sigma^i \sim \mathrm{MN}_{m\times p}(0, I_m, \Sigma^i)$, for $i = 1, \ldots, n$. This implies that within-subject observations are independent conditional on the subject-specific covariance matrix. To allow for subject-level heterogeneity, we assume the subject-specific covariance matrices Σ^i are random, and we denote by Σ the corresponding population-level covariance matrix. Conditional on the matrix Y^i_{-j}, the random effect coefficients γ^i_j and the noise vectors ϵ^i_j are independent with mean zero and covariance $\Psi_j \in \mathbb{R}^{(p-1)\times(p-1)}$ and $R^i_j \in \mathbb{R}^{m\times m}$, respectively. We assume $\gamma^i_j \in \mathrm{SG}(c_1\|\Psi_j\|_2)$ and $\epsilon^i_j \in \mathrm{SG}(c_2\|R^i_j\|_2)$, which implies that $Y^i_j - Y^i_{-j}\beta_j | Y^i_{-j} \in \mathrm{SG}(c_1\|\Psi_j\|_2 + c_2\|R^i_j\|_2)$.

Similar to other specifications of graphical models [see, e.g., 224, 38], the model in equation (3.2) specifies the conditional distribution of Y_j as a function of other variables, Y_{-j}. In this model, the population-level network edge (j, k) is characterized by the $\beta_{j,k}$ coefficient, such that $(j, k) \in E$ if and only if $\beta_{j,k} \neq 0$, providing a convenient specification of the conditional dependence structure while accounting for subject-level variability of the edges. In this work, we focus on developing an inference framework for the model in equation (3.2) and leave the characterization of joint probability distributions that are consistent with the conditionally-specified models [229] to future research. We let p grow with the sample size n and the number of observations per subject m. In the main text, we focus on the setting

where $p > c_0 m$ for some constant $c_0 > 1$, which is most relevant to our application, and defer the case of $p < c_0 m$ for some constant $0 < c_0 < 1$ to Appendix B.

The unknown random effect covariance matrices Ψ_j pose challenges to the estimation and inference of the fixed effect coefficients β_j. Following the quasi-likelihood approach in [61], we use proxy matrices in place of the unknown covariance matrix. Specifically, we use the proxy matrix $\Sigma_a^{i,(j)} = aY_{-j}^i (Y_{-j}^i)^\top + I_m$, with a fixed positive constant a, to approximate the covariance matrix of Y_j^i, which is $\Sigma_\theta^{i,(j)} = Y_{-j}^i \Psi_j (Y_{-j}^i)^\top + R_j^i$, for $i = 1, \ldots, n$. We then use a LASSO estimator for the fixed effect coefficients β_j that leverages the proxy matrix $\Sigma_a^{i,(j)}$ to "decorrelate" the observations. In addition, we propose an inference framework based on the de-biased LASSO method. As we will discuss later, our estimation and inference procedures for the coefficients β_j do not depend on the estimates of the variance components Ψ_j and R_j^i and are, hence, well-defined.

LASSO estimator for β_j

We propose a LASSO estimator for the fixed effect coefficients β_j based on the de-correlated observations. To this end, let $\Sigma_a^{(j)} = \text{diag}\left(\{\Sigma_a^{i,(j)}\}_{i=1}^n\right)$, and Y be the $nm \times p$ matrix obtained by vertically stacking $\{Y^i\}_{i=1}^n$. Our proposed estimator $\hat{\beta}_j$ is given by

$$\hat{\beta}_j = \underset{\beta_j \in \mathbb{R}^{p-1}}{\text{argmin}} \frac{1}{2 \,\text{tr}\left((\Sigma_a^{(j)})^{-1}\right)} \left\| (\Sigma_a^{(j)})^{-1/2} (Y_j - Y_{-j}\beta_j) \right\|_2^2 + \lambda_{a,j} \|\beta_j\|_1, \tag{3.3}$$

with tuning parameter $\lambda_{a,j} > 0$. In Section 3.3, we show that under mild assumptions $\hat{\beta}_j$ is a consistent estimator of β_j, for a suitable choice of $\lambda_{a,j}$ and for any choice of constant a.

Taking $\Sigma_a^{(j)} = I_{nm}$ in (3.3) leads to the classical LASSO estimator. While it is easy to show that this estimator is also consistent, it is known that due to the correlation across observations, this misspecified estimator has sub-optimal rate of convergence [132].

Inference based on de-biased LASSO

We next propose an inference framework for the coefficient $\beta_{j,k}$, $k \in V \backslash \{j\}$ based on the asymptotic normality of the de-biased LASSO estimator. The de-biasing procedure builds on

the idea in [246], which uses regularized regression to estimate the bias term of the LASSO estimator.

For $k \in V \backslash \{j\}$, our de-biased estimator $\hat{\beta}_{j,k}^{(\mathrm{db})}$ is defined as

$$\hat{\beta}_{j,k}^{(\mathrm{db})} = \hat{\beta}_{j,k} + \frac{\sum_{i=1}^{n} \left(\hat{w}_{j,k}^{i}\right)^{\top} \left(\Sigma_b^{i,(j,k)}\right)^{-1/2} \left(Y_j^i - Y_{-j}^i \hat{\beta}_j\right)}{\sum_{i=1}^{n} \left(\hat{w}_{j,k}^{i}\right)^{\top} \left(\Sigma_b^{i,(j,k)}\right)^{-1/2} Y_j^i}, \tag{3.4}$$

where the proxy covariance matrices $\Sigma_b^{i,(j,k)}$ are defined as

$$\Sigma_b^{i,(j,k)} = a Y_{-\{j,k\}}^i (Y_{-\{j,k\}}^i)^{\top} + I_m, \quad \Sigma_b^{(j,k)} = \mathrm{diag}\left(\{\Sigma_b^{i,(j,k)}\}_{i=1}^n\right),$$

with the same constant a used in $\Sigma_a^{(j)}$; and the projection related terms $\hat{\kappa}_{j,k}$, $\hat{w}_{j,k}^i$ are defined as

$$\hat{\kappa}_{j,k} = \underset{\kappa_{j,k} \in \mathbb{R}^{p-2}}{\mathrm{argmin}} \frac{1}{2\,\mathrm{tr}\left(\left(\Sigma_b^{(j,k)}\right)^{-1}\right)} \left\| \left(\Sigma_b^{(j,k)}\right)^{-1/2} \left(Y_k - Y_{-\{j,k\}} \kappa_{j,k}\right) \right\|_2^2 + \lambda_{j,k} \|\kappa_{j,k}\|_1$$

$$\hat{w}_{j,k}^i = \left(\Sigma_b^{(j,k)}\right)^{-1/2} \left(Y_k - Y_{-\{j,k\}} \hat{\kappa}_{j,k}\right),$$

with tuning parameter $\lambda_{j,k} > 0$.

Here, the vector $\hat{w}_j^i$ is approximately the orthogonal complement of the projection of the vector $\left(\Sigma_b^{(j,k)}\right)^{-1/2} Y_j$ onto the space spanned by the columns of $\left(\Sigma_b^{(j,k)}\right)^{-1/2} Y_{-j}$, where the projection vector $\hat{\kappa}_{j,k}$ is computed via LASSO regression. We use the proxy matrix $\Sigma_b^{(j,k)}$ to "decorrelate" observations Y_k and $Y_{-\{j,k\}}$ in the LASSO regression. Note that we define the proxy matrix $\Sigma_b^{(j,k)}$ differently from *LCL* [132]. This modification is crucial to successfully establishing the asymptotic normality of the de-biased estimator $\hat{\beta}_{j,k}^{(\mathrm{db})}$, especially in the setting where the fixed effect design matrix has overlapping columns with the random effect design matrix.

Denoting by z_α the αth quantile of a standard normal distribution, we can construct a two-sided $(1-\alpha) \times 100\%$ confidence interval for the coefficient $\beta_{j,k}$ as $\hat{\beta}_{j,k}^{(\mathrm{db})} \pm z_{\alpha/2} \sqrt{\hat{V}_{j,k}}$,

where $\hat{V}_{j,k}$ is a sandwich-type estimator of the variance of $\hat{\beta}_j^{(\mathrm{db})}$, defined as:

$$\hat{V}_{j,k} = \frac{\sum_{i=1}^{n} \left| (\hat{w}_{j,k}^i)^\top \left(\Sigma_b^{i,(j,k)} \right)^{-1/2} \left(Y_j^i - Y_{-j}^i \hat{\beta}_j \right) \right|^2}{\left| \sum_{i=1}^{n} (\hat{w}_{j,k}^i)^\top \left(\Sigma_b^{i,(j,k)} \right)^{-1/2} Y_j^i \right|^2}. \tag{3.5}$$

3.3 Theoretical Analysis

In this section, we show that, under mild conditions, the proposed LASSO estimator $\hat{\beta}_j$ in (3.3) is consistent, and the proposed de-biased LASSO estimator $\hat{\beta}_{j,k}^{(\mathrm{db})}$ in (3.4) is asymptotically normal. We first state the assumptions under which we establish the consistency of $\hat{\beta}_j$ defined in (3.3). Recall that we use $c, c_0, c_1, \ldots$ to denote generic positive constants whose values may vary line by line.

Assumption 1.

1. *The number of observations per subject m and the number of covariates p satisfy $p > c_0 m$ for some constant $c_0 > 1$, and $p > c_1$ for some suitably large constant $c_1 > 0$. Moreover, $p \log(p)/(mn) = o(1)$.*

2. *$\forall\ i, j$, $\sigma(\Sigma) \asymp \sigma(R_j^i) \asymp \|\Psi_j\|_2 \asymp 1$, and $\|\Sigma^i - \Sigma\|_2 \leq \sigma_{\min}(\Sigma) - c_2$ for some constant $c_2 > 0$.*

In Assumption 1.1, we restrict the growth rate of p relative to m and n. Moreover, we assume p/m is no smaller than a positive constant $c_0 > 1$, which is the situation most relevant to our application. In Appendix B, we discuss the case when $p/m \leq 1/c_0$ for some constant $c_0 > 1$.

By Weyl's theorem, Assumption 1.2 implies $\sigma(\Sigma^i) \asymp 1$ for all $i = 1, \ldots, n$ [15], requiring the singular values of the covariance matrices Σ, Σ^i, and R_j^i to be both lower and upper bounded by positive constants. Note that we do not require independent and identically distributed noise terms for each observation. Consequently, our model is applicable across various settings, including those involving time series outcomes. Specifically, it accommodates scenarios where the noise terms conform to an autoregressive model of order 1, since the covariance matrices R_j^i in such cases are Kac-Murdock-Szego matrices with all eigenvalues

of order $O(1)$ [218].

To establish the consistency of $\hat{\beta}_j$, we only require an upper bound on the singular values of the covariance matrices Ψ_j in Assumption 1.2. However, later we also require one of the following assumptions on Ψ_j to hold in order to establish the asymptotic normality of the de-biased estimator; we state these assumptions here for convenience.

Assumption 2.

1. *(Diagonal structure):* $\forall\ j,\ \Psi_j = \mathrm{diag}(\psi_j)$ *for a vector* $\psi_j \in \mathbb{R}^{p-1}$. *The support of* ψ_j *is* S_{ψ_j} *with cardinality* $s_{\psi_j} < c_1 m \wedge n$ *for some constant* $c_1 > 0$. *Moreover,* $\min(\psi_{S_{\psi_j}}) \asymp \max(\psi_{S_{\psi_j}}) \asymp 1$.
2. *(Bounded eigenvalues):* $\forall\ j,\ \sigma_{\min}(\Psi_j) \asymp 1$.

Assumption 2.1 allows us to cover the settings when the minimum singular values of the covariance matrices Ψ_j are not bounded away from zero. In such a case, we consider a sparse diagonal structure for the matrices Ψ_j, under which our results hold with slightly different sample size assumptions.

The next result establishes the theoretical properties of the proposed estimator $\hat{\beta}_j$ in (3.3).

Theorem 3.3.1 (Fixed effect estimator consistency)**.** *Suppose Assumption 1.1 and Assumption 1.2 hold and that* $\lambda_{a,j} = c_1\sqrt{p\log(p)/(nm)}$ *for a suitably large* $c_1 > 0$. *Then, with probability at least* $1 - 4\exp\{-cn\} - 12\exp\{-c\log(n)\} - 3\exp\{-cmnp^{-1}\}$,

$$\left\|\hat{\beta}_j - \beta_j^*\right\|_\delta = O_p\left(s_j^{1/\delta}\sqrt{\frac{p\log(p)}{mn}}\right), \quad \delta \in \{1,2\},$$
$$\left\|(\Sigma_a^{(j)})^{-1/2} Y(\hat{\beta}_j - \beta_j^*)\right\|_2^2 = O_p\left(s_j \log(p)\right).$$

If, in addition, Assumption 2.1 holds, taking $\lambda_{a,j} = c_2\sqrt{\log(p)\log^2(n)/n}$ *for a suitably large* $c_2 > 0$, *we have with probability at least* $1-4\exp\{-cn\}-12\exp\{-c\log(n)\}-3\exp\{-cmnp^{-1}\}$

that

$$\left\|\hat{\beta}_j - \beta_j^*\right\|_\delta = O_p\left(s_j^{1/\delta}\sqrt{\frac{\log^2(n)\log(p)}{n}}\right), \quad \delta \in \{1,2\},$$
$$\|(\Sigma_a^{(j)})^{-1/2} Y(\hat{\beta}_j - \beta_j^*)\|_2^2 = O_p\left(\frac{s_j m \log^2(n)\log(p)}{p}\right).$$

Theorem 3.3.1 shows that the de-correlated LASSO estimator $\hat{\beta}_j$ achieves ℓ_1, ℓ_2 and prediction consistency. The proof of Theorem 3.3.1 is based on the classical proof of the consistency of the LASSO estimator in regression settings [29] with multiple key modifications to extend it to our setting. One crucial step is to show that the restricted eigenvalue condition [29] holds for the matrix product $X\left(aZZ^\top + I\right)^{-1}X^\top$, where X is the fixed effect design matrix and Z is the random effect design matrix. Since in our setting the fixed and random effect design matrices are identical, we cannot rely on techniques such as those adopted in [132] where the conditional mean independence assumption between X and Z is used to remove the dependence on the matrix Z. Instead, we jointly study the contribution of both the fixed and random effect design matrices, and use results from random matrix theory [219] to directly characterize the eigenvalue distribution of the relevant quantities. In particular, we obtain a set of tight bounds for functions of the non-zero singular values of the matrices Y^i under Assumption 1. This key result, which may be of independent interest, is presented in Lemma B.1.2 in Appendix B.1. We state the other lemmas necessary to prove Theorem 3.3.1, their proofs, and the proof of Theorem 3.3.1 in Appendix B.1.

Next, we state a simplified version of the assumptions under which we establish the asymptotic normality of the de-biased LASSO estimator $\hat{\beta}_{j,k}^{(\mathrm{db})}$ defined in (3.4). In order to allow for such a simplification, we have made some additional mild assumptions such as $\sigma\left(G_k^{(j)}\right) \asymp p$ under Assumption 2.2, $c_1 \leq \sigma_{\min}\left(G_k^{(j)}\right) \leq \sigma_{\max}\left(G_k^{(j)}\right) \leq c_2 m$ under Assumption 2.1, and $|H_k^{(j)}| \asymp s_j \asymp 1$. Recall that model (3.2) implies that given Y_{-j}^i, the vector $Y_j^i - Y_{-j}^i\beta_j$ is sub-Gaussian with covariance matrix $\Sigma_\theta^{i,(j)} = Y_{-j}^i \Psi_j (Y_{-j}^i)^\top + R_j^i$. If we were to assume that given $Y_{-\{j,k\}}^i$, the covariance matrix of Y_k^i has a "sandwich" form akin to $\Sigma_\theta^{i,(j)}$, namely $G_k^{(j)} = Y_{-\{j,k\}}^i \Psi_{j,k} (Y_{-\{j,k\}}^i)^\top + R_{j,k}^i$ for some matrix $\Psi_{j,k} \in \mathbb{R}^{(p-2)\times(p-2)}$ and $R_{j,k}^i \in \mathbb{R}^{m\times m}$,

we could then characterize the rates of $\|G_k^{(j)}\|_2$ and $\sigma_{\min}\left(G_k^{(j)}\right)$ with the above assumed rates (see Lemma B.2.1).

Assumption 3.

1. $\forall\ j,k$, *conditioning on the matrix* $Y^i_{-\{j,k\}}$, *the vector* $Y^i_k - Y^i_{-\{j,k\}}\kappa_k^{(j),*}$ *has mean zero and variance* $G_k^{(j)}$, *and belongs to the class* $\mathrm{SGV}(c_1\|G_k^{(j)}\|_2)$. *The support of* $\kappa_k^{(j),*}$, $H_k^{(j)}$, *has cardinality* $|H_k^{(j)}|$ *satisfying* $\|\kappa_k^{(j),*}\|_1 \le c_1|H_k^{(j)}|$.

2. *If Assumption 2.1 holds, then* $p^2\log(p) \ll nm^3$, $m\log^7(n) \ll n$, $\log(n)\log(p) \ll n$. *If Assumption 2.2 holds, then* $p\log(n)\log(p) \le c_3 mn$.

In Assumption 3.1, we specify an upper bound for $\|\kappa_k^{(j),*}\|_1$. This is not too restrictive, because given that the variance of each node is bounded (implied by Assumption 1.2), it is reasonable to expect that the coefficients $\kappa_k^{(j),*}$ are not too large in absolute value. In the case of no subject-level heterogeneity, that is $Y^i \sim \mathrm{N}(0,\Sigma)$ and $\kappa_k^{(j),*} = \left(\Sigma_{-\{j,k\},-\{j,k\}}\right)^{-1}\Sigma_{-\{j,k\},k}$, it is easy to show that $\sum_{l=1,l\neq k}^{p-1}\left|(\kappa_k^{(j),*})_l\sqrt{(\Omega_j)_{k,k}}/\sqrt{(\Omega_j)_{l,l}}\right| \le |H_k^{(j)}|$, and $\|\kappa_k^{(j),*}\|_1 \le c_1|H_k^{(j)}|$ is satisfied.

In Assumption 3.1, we also specify the conditional distribution of Y_k^i given $Y^i_{-\{j,k\}}$. We do not assume a specific structure for the covariance matrix $G_k^{(j)}$, but only require mild bounds on the minimum and maximum eigenvalues (Appendix B.2).

Assumption 3 indicates different sample size requirements according to different structures of Ψ_j in Assumption 2. Note that Assumption 2 can be relaxed at the cost of stricter sample size requirements: if we only assume $\|\Psi_j\|_2 \le c_1$ and drop Assumption 2, the asymptotic normality property of $\hat\beta_{j,k}^{(\mathrm{db})}$ will hold with some additional sample size assumptions (Appendix B.2.1, Remark B.2.1), which would restrict the growth rate of p to be slower than n.

The next result establishes the asymptotic normality of the estimator $\hat\beta_{j,k}^{(\mathrm{db})}$ in (3.4).

Theorem 3.3.2. *Under Assumptions 1 and 3, we have with probability at least* $1 - c_1\exp\{-cn\} - c_2\exp\{-c\log(n)\} - c_3\exp\{-cmn/p\} - c_4\exp\{-c\log(p)\} - c_5\exp\{-cmn\}$, $(V_{j,k})^{-1/2}\left(\hat\beta_{j,k}^{(\mathrm{db})} - \beta_{j,k}^*\right) = R_{j,k} + o_p(1)$, *where* $R_{j,k} \xrightarrow{d} \mathrm{N}(0,1)$ *and the variance* $V_{j,k}$ *is given*

by

$$V_{j,k} = \frac{\sum_{i=1}^{n} (\hat{w}_{j,k}^{i})^{\top} \left(\Sigma_b^{i,(j)}\right)^{-1/2} \Sigma_\theta^{i,(j)} \left(\Sigma_b^{i,(j)}\right)^{-1/2} \hat{w}_{j,k}^{i}}{\left| \sum_{i=1}^{n} (\hat{w}_{j,k}^{i})^{\top} \left(\Sigma_b^{i,(j)}\right)^{-1/2} Y_j \right|^2}.$$

The proof of Theorem 3.3.2 builds on the properties of the projection vectors $\hat{\kappa}_{j,k}$ and the orthogonal complement of the projection $\hat{w}_{j,k}^{i}$. Specifically, we show that $(V_{j,k})^{-1/2} \left(\hat{\beta}_{j,k}^{(\mathrm{db})} - \beta_{j,k}^{*}\right)$ can be divided into two terms, where one term is $o_p(1)$ and the other term can be shown to be asymptotically normal, thanks to the Lyapunov central limit theorem. As in the proof of Theorem 3.3.1, we extensively use the core Lemma B.1.2 to bound quantities involving both the fixed effect and the random effect design matrices.

A novel feature of our results, compared to those in the literature, including *LCL* [132], is that we establish the unconditional asymptotic normality of $\hat{\beta}_{j,k}^{(\mathrm{db})}$. In contrast, other approaches establish the asymptotic normality only conditionally on the random effect design matrices, even when the design matrices are assumed to be random. This unconditional normality is crucial for our application to brain connectivity network inference, and to the best of our knowledge, this is the first attempt to characterize the unconditional asymptotic properties of the fixed effect coefficients estimator with random design matrices in high-dimensional LMMs.

The lemmas required to prove Theorem 3.3.2 and their proofs are presented in Appendix B.2. Lemma B.2.3 gathers several intermediate results necessary to prove Theorem 3.3.2.

Finally, we show that the sandwich estimator $\hat{V}_{j,k}$, defined in (3.5), is a consistent estimator of $V_{j,k}$ under Assumption 4. We apply the same simplifications as we did for Assumption 3.

Assumption 4.

1. *If Assumption 2.1 holds, then* $m^2 \log^5(n) \log^2(p) \ll n$.
2. *If Assumption 2.2, then* $\log^2(p) \log^4(n) \ll n$.

Theorem 3.3.3. *Under Assumptions 1, 3, and 4, with probability at least* $1 - c_1 \exp\{-cn\} - c_2 \exp\{-c \log(n)\} - c_3 \exp\{-cmn/p\} - c_4 \exp\{-c \log(p)\} - c_5 \exp\{-cmn\}$ *we have* $\hat{V}_{j,k}/V_{j,k} =$

$1 + o_p(1)$.

The proof is presented in Appendix B.3.

3.4 Extensions

3.4.1 Variance Component Estimation

An appealing property of the proposed estimation and inference framework for the fixed effects is that we do not need to estimate the variance components $\theta = (\Psi, \{R^i\}_{i=1}^n)$. This is convenient when only the fixed effects are of interest. However, variance components also contain important information and should not be ignored. In our application of brain network analysis, a non-zero random effect variance component indicates the presence of subject-level heterogeneity in the connectivity between two brain nodes. In other applications, such as heritability analysis [206] and genome-wide association studies [5], the variance component estimates are necessary for downstream analysis, or may be of independent interest.

Unfortunately, estimating the variance components in doubly high-dimensional LMM settings introduces unique challenges that have not been addressed by existing approaches. In particular, the method of [134] assumes bounded m, in order to use a Cholesky decomposition for estimating the random effect covariance matrix. The sample-splitting approach of [132] requires $m > q$ for q random effect covariates, which restricts its applicability in high dimensions. We extend the sample-splitting approach to doubly high-dimensional LMM settings, and propose a penalized moment-based estimator for selecting and estimating the variance components. In particular, we allow for m to be smaller than q and to grow with n. To simplify the problem, in the following, we assume that the noise terms are independent and identically distributed within each subject's observations, i.e., $R^i = \sigma_e^2 I_m$. Moreover, we assume Ψ is a diagonal matrix satisfying Assumption 2.1, such that $\Psi = \text{diag}(\psi)$. To simplify the notation and to broaden the scope of the framework, we will present the proposed estimators and the theoretical results under a more general LMM formulation:

$$y^i = X^i\beta + Z^i\gamma_i + \epsilon_i, \quad i = 1, \ldots, n, \tag{3.6}$$

where each $y^i \in \mathbb{R}^m$ is the observation vector, $X^i \in \mathbb{R}^{m\times p}$ and $Z^i \in \mathbb{R}^{m\times q}$ are the design matrices with $X^i \mid \Sigma_X^i \sim \mathrm{MN}_{m\times p}(0, I_m, \Sigma_X^i)$ and $Z^i \mid \Sigma_Z^i \sim \mathrm{MN}_{m\times q}(0, I_m, \Sigma_Z^i)$. Conditional on X^i, the random effect coefficients $\gamma_i \in \mathbb{R}^q$ and the noise term $\epsilon_i \in \mathbb{R}^m$ are independent with variance Ψ and R^i, and satisfy $\gamma_i \in \mathrm{SGV}(c_1\|\Psi\|_2)$, $\epsilon_i \in \mathrm{SGV}(c_2\|R^i\|_2)$, respectively. The fixed effect coefficient vector β has support S with cardinality s. See Appendix B for additional details on the model definition.

It is known that the variance components become non-identifiable if the random effect covariance matrix is proportional to the noise covariance [227]. In our case, this would happen if $Z^i\Psi(Z^i)^\top = c\sigma_e^2 I_m$, for some constant $c > 0$. However, given that we assume the diagonal of Ψ is sparse, we have $\sigma_{\min}\left(Z^i\Psi(Z^i)^\top\right) = 0$ whereas $\sigma_{\min}(\sigma_e^2 I_m) = \sigma_e^2 > 0$, ensuring identifiability.

Let $\theta = (\psi, \sigma_e^2)$ denote the variance components. To estimate θ, we adopt a sample-splitting approach, and partition the n subjects into three sub-samples of similar sizes with index set S_k, $k = 1, 2, 3$, such that $n_k = |S_k|$ and $n_1 \asymp n_2 \asymp n_3$. We start by using the first n_1 subjects to estimate the fixed effect parameters β, denoting the estimates as $\hat{\beta}$. We then use the second sub-sample with n_2 subjects to estimate the vector ψ. Denoting $\hat{r}_i = y^i - X^i\hat{\beta}$, $i \in S_2$, we define the penalized moment-based estimator for ψ with tuning parameter λ_θ, as:

$$\hat{\psi} = \underset{\psi\in\mathbb{R}^q}{\operatorname{argmin}} \sum_{i\in S_2} \left\| \hat{r}_i\hat{r}_i^\top - \operatorname{diag}(\hat{r}_i)\operatorname{diag}(\hat{r}_i) - Z^i \operatorname{diag}(\psi)(Z^i)^\top + \sum_{l=1}^{q} \psi_l \operatorname{diag}(Z_l^i)\operatorname{diag}(Z_l^i) \right\|_F^2 + \lambda_\theta\|\psi\|_1. \tag{3.7}$$

The estimator for ψ is constructed from the second moment of the residuals $r_i = y^i - X^i\beta^*$ by noticing that $\mathbb{E}\left(\left(r_i r_i^\top\right)_{l,k}\right) = \left(Z^i\Psi(Z^i)^\top\right)_{l,k}$. In the high-dimensional setting here considered, we adopt a LASSO regularization to guarantee the consistency of the estimator and simultaneously perform variable selection.

Finally, we use the third sub-sample with n_3 subjects to estimate the noise variance σ_e^2. To

this end, we propose a simple moment-based estimator defined as

$$\hat{\sigma}_e^2 = \frac{1}{n_3 m} \sum_{i \in S_3} \operatorname{tr}\left(\hat{r}_i \hat{r}_i^\top - Z^i \operatorname{diag}(\hat{\psi}) (Z^i)^\top \right), \tag{3.8}$$

where $\hat{r}_i$ is computed based on the the first n_1 samples, and $\hat{\psi}$ is computed based on the second n_2 samples.

We gather the assumptions needed to prove the consistency of the proposed variance component estimator $(\hat{\psi}, \hat{\sigma}_e^2)$ in Assumption 5 below. The true values are denoted by $\theta^* = (\psi^*, \sigma_e^{2,*})$.

Assumption 5.

1. *The vectors γ_i and ϵ_i are normally distributed with mean zero and variance matrices Ψ, $\sigma_e^2 I_m$, respectively. The covariance matrix Ψ satisfies Assumption 2.1.*
2. *Letting $s_Z = \max_i \max_j \sum_{k=1, k\neq j}^{q} \mathbf{1}\left\{ (\Sigma_Z^i)_{j,k} \gg \log(nq^2)/\sqrt{m} \right\}$, we have $\sqrt{m} \gg \log(nq)$, $nm \gg \max\left\{ q^{3/2} \log(q) \log^2(n), \log(q)\log(n)\log(nq^2)\left(s_Z\sqrt{m} + q\log(nq^2) \right) \right\}$, $nm^3 \gg q^2 \log(q)\log^2(n)$.*
3. *$\sqrt{n} m^2 \gg s s_\psi q \log(p)\log(q)\log(n)\log(nmq)$.*

Theorem 3.4.1. *Under Assumption 1 and 5.1–5.2, with probability at least $1 - c_1\exp\{-c\log(nq)\} - c_2\exp\{-cn\} - c_3\exp\{-c\log(n)\} - c_4\exp\{-cmn/q\} - c_5\exp\{-c\log(q)\}$, we have*

$$\|\hat{\psi} - \psi^*\|_\delta \leq s_\psi^{1/\delta} \frac{q\log(n)\log(p)\log(nmq)}{\sqrt{n}m}, \quad \delta = 1, 2.$$

Theorem 3.4.2. *Under Assumption 1 and 5, we have $|\hat{\sigma}_e^2 - \sigma_e^{2,*}| = o_p(1)$ with probability at least $1 - c_1 s_\psi q \log^3(n)/(nm) - c_2\exp\{-cnm\} - c_3\exp\{-cn\} - c_4\exp\{-c\log(n)\} - c_5\exp\{-cmn/q\} - c_6\exp\left\{-cn^2m^2/(sq^2\log^4(n)\log(p))\right\}$.*

Similar to the proof of Theorem 3.3.1, we leverage random matrix theory [219] to bound the eigenvalues of matrix products and follow the classical proof of the consistency of the LASSO estimator [29]. As previously mentioned, our estimator allows q to be larger than

m, and also applies to the case where q is smaller than m (Appendix B.4). The proof and related lemmas are collected in Appendix B.4.

3.4.2 High-Dimensional Heterogeneous Vector Autoregressive Models

In this section, we show that the proposed estimation and inference framework can be also extended to heterogeneous high-dimensional first-order VAR models. Different from GGMs describing an unconditional functional connectivity brain network, VAR modeling has been a popular approach for inferring the joint *effective connectivity* network [67] among multiple brain regions [74, 35, 104]. Specifically, the VAR coefficients of the fitted model reveal Granger causal relations among brain regions [80, 182, 197]. A subject-specific first-order VAR model is formulated as

$$Y^i(t) = \Phi^i Y^i(t-1) + \epsilon^i(t), \tag{3.9}$$

for observations $\{Y^i(t)\}_{t=1}^T$, VAR coefficient matrix Φ^i and error term $\{\epsilon^i(t)\}_{t=1}^T$ for the i-th subject.

Recent developments have extended the VAR model to high-dimensional settings for stationary [208, 199, 11, 87] and non-stationary [185, 92] time series, and for nonlinear [37, 214, 248, 142], sub-Gaussian [252] and non-Gaussian [215, 228] time series. However, these extensions are limited to modeling a single subject's network. Two-stage approaches are often adopted by neuroscientists for inference in multi-subject settings: in the first stage, separate VAR models are fitted for each subject; in the second stage, individual-level summary statistics are used for group-level analysis [47, 155, 160]. This includes aggregating each subject's p-values for entries of Φ via Fisher's method [47], and conducting a two-sample t-test based on individual-level summary statistics [155]. However, two-stage approaches have significant limitations. Firstly, they often fail to adequately address the uncertainty associated with estimated individual-level statistics in the second-stage analysis [39], which potentially results in inaccurate group-level conclusions. Secondly, subject-level analyses overlook shared structural information among individuals, leading to less efficient estimates for the group-level structure. Lastly, choosing different methodologies for the

second stage can lead to different conclusions. These limitations can be problematic when making inferences in the presence of subject-level heterogeneity.

To overcome the above issues, mixed-effect VAR (MEVAR) models have been proposed for multi-subject brain signal analyses [78, 77]. In MEVAR models, the VAR coefficients Φ^i in (3.9) are *random variables* centered at the population-level matrix Φ such that $\Phi^i = \Phi + \Gamma^i$, where the matrix Φ represents the population-level effective connectivity brain network. However, current applications of this model are limited to low-dimensional fMRI observations [78, 77, 27, 226, 121, 26]. Even though there is an extensive body of work on incorporating mixed effects in VAR models, the majority focus on random coefficient AR models from a Bayesian perspective [140, 133, 159], while the rest are concerned with low-dimensional MEVAR models [163, 222] (see [178] for a detailed overview). Theoretical results on multivariate MEVAR models are rare [178], and, to the best of our knowledge, valid frequentist inference for high-dimensional MEVAR models has not been investigated.

Our proposed doubly high-dimensional LMM framework can be adopted to infer the population structure Φ in a high-dimensional first-order MEVAR model. Let the vector $\mathrm{vec}(A)$ denote the vectorized matrix A through vertical stacking of its columns, and let $A \otimes B$ denote the Kronecker product between two matrices A and B. We can rewrite the model in equation (3.9) as:

$$\tilde{Y}^i = \tilde{X}^i \,\mathrm{vec}(\Phi) + \tilde{X}^i \,\mathrm{vec}(\Gamma^i) + \tilde{\epsilon}^i, \quad i = 1, \ldots, n, \tag{3.10}$$

where

$$\tilde{Y}^i = \begin{pmatrix} Y^i(2) \\ \cdots \\ Y^i(T) \end{pmatrix}, \quad \tilde{X}^i = \begin{pmatrix} \big(Y^i(1)\big)^\top \otimes I_p \\ \cdots \\ \big(Y^i(T-1)\big)^\top \otimes I_p \end{pmatrix}, \quad \tilde{\epsilon}^i = \begin{pmatrix} \epsilon^i(2) \\ \cdots \\ \epsilon^i(T) \end{pmatrix}.$$

In equation (3.10), $\tilde{X}^i$ is the design matrix, $\mathrm{vec}(\Phi)$ is the fixed effect coefficient vector, and

$\text{vec}(\Gamma^i)$ is the random effect coefficient vector. We can thus recast the problem of inferring Φ in a first-order MEVAR as the problem of inferring fixed effect coefficients in a doubly high-dimensional LMM, for which our proposed LMM framework applies. We can also show that inferring the whole matrix Φ is equivalent to inferring each row of Φ separately, which drastically accelerates computation.

We can show that the resulting penalized estimate for Φ is consistent, and the inference framework yields valid confidence intervals. Most of the proof follows the techniques used for the theoretical results in Section 3.3. However, due to the presence of correlated rows in the design matrix $\tilde{X}^i$, we introduce a pivotal lemma that extends the theoretical results in Section 3.3 to the case of a first-order MEVAR mode (Appendix B.6.2, Lemma B.6.1). Using the properties of a stationary first-order VAR process, this lemma connects the singular values of $\tilde{X}^i$ to the singular values of a standard Gaussian matrix and facilitates the remainder of the proof.

We demonstrate through a simulation study in Appendix B.6.1 that our proposed framework works well for inferring the matrix Φ for high-dimensional first-order MEVAR models.

3.5 Simulation Studies

3.5.1 Simulation setting

For ease of exposition, we generate data from a doubly high-dimensional LMM formulated as:

$$y^i = X^i\beta + X^i\gamma_i + \epsilon_i, \quad i = 1, \ldots, n, \tag{3.11}$$

with $X^i \mid \Sigma_X^i \sim \text{MN}_{m\times p}(0, I_m, \Sigma_X^i)$, $\gamma_i \sim \text{N}(0, \text{diag}(\psi))$ and $\epsilon_i \sim \text{N}(0, \sigma_e^2 I_m)$. This is equivalent to analyzing the edges connected to one single node in a GGM.

We compare our approach with two competing approaches: (i) the *LCL* method of [132], and (ii) the de-biased LASSO method [246] (referred to as *dblasso*) as a benchmark approach which ignores the subject-level heterogeneity among observations. We compare the methods in terms of the total mean squared error (total MSE) for the estimates of all β_l coefficients, the power of testing non-zero coefficients, the type-I error for testing zero coefficients at 5%

significance level, and the coverage of 95% confidence intervals. We also compare the method proposed in Section 3.4.1 for estimation of the variance components, $\theta = (\psi, \sigma_e^2)$, with the method of *LCL* by assessing the MSE of the noise variance σ_e^2, the total MSE for estimates of all ψ_l, and the selection consistency of the non-zero random effect variance components ψ_{S_ψ}. The selection consistency performance is evaluated via the Matthews correlation coefficient (MCC) [150]; MCC summarizes true positive and false positive rates, with higher values indicating more accurate identification of the non-zero variance components.

We generated data from the doubly high-dimensional linear mixed model specified in (3.11) with $n \in \{30, 50, 80, 100\}$ and $m \in \{15, 30, 50, 70\}$. For each combination of (n, m), we set $p \in \{20, 60\}$ and replicate 200 independent Monte Carlo simulations. We set $(\beta_1^*, \beta_2^*, \beta_6^*, \beta_7^*, \beta_9^*) = (1, 0.5, 0.2, 0.1, 0.05)$ and $\beta_l^* = 0$ for the remaining l's. The random effects γ_i and the noise terms ϵ_i were independent realizations of two multivariate normal distributions $\mathrm{N}(0, \mathrm{diag}(\psi))$ and $\mathrm{N}(0, \sigma_e^2 I_m)$, respectively. The true values of the variance components $(\psi^*, \sigma_e^{2,*})$ are set as follows: $(\psi_1^*, \psi_4^*, \psi_7^*, \psi_9^*, \psi_{10}^*, \psi_{12}^*, \psi_{16}^*, \psi_{20}^*) = (2, 2, 0.1, 0.1, 4, 0.1, 2, 0.1)$, with the remaining ψ_l^*'s set to 0, and $\sigma_e^{2,*} = 1$. The fixed effect design matrices X^i were independent realizations of a matrix normal distribution $X^i \mid \Sigma_X^i \sim \mathrm{MN}_{m\times p}(0, I_m, \Sigma_X^i)$. To generate Σ_X^i, we first generated a population-level covariance matrix Σ_X, which was set as a sparse random matrix with diagonal entries 1 and off-diagonal entries drawn independently from a mixture distribution: each entry was either set to zero with probability 0.8, or was drawn from a uniform distribution $\mathrm{Unif}(-0.5, 0.5)$. This choice was motivated by the nature of sparsely correlated brain networks. Each Σ_X^i was then generated as a perturbed version of Σ_X by (i) determining varying off-diagonal entries of Σ_X^i by drawing a Bernoulli random variable with success probability 0.2; and (ii) generating variations by adding a mean zero normal perturbation with standard deviation 0.1. To ensure symmetry, only the entries in the upper-diagonal of Σ_X^i were considered as candidates for perturbation. We repeated the above two steps if the generated Σ_X^i was not positive definite. The resulting Σ_X^i represent subject-level heterogeneity in the brain networks.

We used cross-validation with MSE as the error criteria to select the tuning parameters λ_a for $\hat{\beta}$, λ_j's for $\hat{\kappa}_j$ and λ_θ for variance components. We used the R function `cv.glmnet`

from the R package `glmnet` (*v4.1-3*, [66]) to implement the cross-validation algorithm. The constant a in the proxy matrix is also viewed as a tuning parameter. We followed the approach described in [132] to select an optimum a via cross-validation: for each candidate value of a, we let the algorithm select the values for the rest of the tuning parameters, and used cross-validation based on MSE to select an optimal a. We present the results based on the optimal a. We used the authors' publicly available R code for *LCL* [132], and used the `hdi` R-package [50] for *dblasso*.

3.5.2 Simulation results

Results for inference on fixed effect parameters β are summarized in Figure 3.2. The proposed method controls the type-I error rate at the nominal level, whereas *LCL* and *dblasso* show inflated type-I errors in various settings (Figure 3.2a, 3.2b): when the random effect variance is large ($\psi_{10} = 4$ for β_{10}), tests by *dblasso* have type-I errors as high as 0.74; *LCL*'s type-I error reaches 0.28 when the random effect variance is zero (for β_{11}); both methods show higher type-I errors with increasing m. Results for $p = 60$ are presented in Appendix B.5.1 Table B.1. When $p = 60$, the type-I error of *LCL* improves for small m, but is still as high as 0.23 for large m values; *dblasso* has high type-I error regardless of p, which is not surprising. Interestingly, *dblasso* often fails when testing covariates with non-zero random effect variances, and *LCL* often fails when testing covariates with zero random effect variances, especially when m is large.

The confidence intervals constructed by the proposed method provide good coverage, while those constructed by *LCL* and *dblasso* show poor coverage in some settings (Figure 3.2c, 3.2d). The pattern of the confidence interval coverage is similar to the pattern of the type-I error: *LCL* has coverage lower than 0.81 for β_{11} at $p = 20$, $m = 70$, and has insufficient coverage for covariates with zero random effect variance when m is large (Appendix B.5.1 Table B.2); *dblasso*'s coverage is always lower than 0.8 for some β_l's (Figure 3.2c), and is as low as 0.24 for $p = 60$ (Appendix B.5.1 Table B.2); both methods have worse coverage with increasing m.

Since *dblasso* has poor confidence interval coverage and highly inflated type-I error, we do

not include it in the power comparison for testing β_l's. The proposed method in general has comparable power against LCL (Figure 3.2e, 3.2f and Appendix B.5.1 Table B.3).

In terms of estimating the fixed effect coefficients, the proposed method always has smaller total MSE than LCL (Figure 3.3a and Appendix B.5.1 Table B.4). Fixed effect coefficient estimates by *dblasso* always have the smallest total MSE.

Since *dblasso* does not provide variance component estimates, it is not included for comparison in Figure 3.3. When $m < q$, LCL is not able to provide variance component estimates, while our method provides sensible estimates in all settings (Figure 3.3b, 3.3c). When $m > q$, our method's estimate of σ_ϵ^2 are not as good as LCL's (Figure 3.3b), but its total MSEs for the random effect variance components are comparable with LCL (Figure 3.3c and Appendix B.5.1 Table B.4).

We also examined the selection consistency of the random effect variance components evaluated by MCC. The proposed method selects the variance components much more accurately than LCL, and the accuracy improves with increasing n and m (Figure 3.3d). LCL has poor selection accuracy, and is less accurate with increasing n. Examining the proportion of simulations that yield non-zero estimates for each variance component reveals that LCL over selects zero variance components (Appendix B.5.1 Figure B.2).

Finally, in terms of computational time, the proposed method is slightly slower than LCL for fixed effect estimation and inference, but is much faster for variance component estimation. The differences become more pronounced for large p (Appendix B.5.1 Figure B.1).

3.6 Data Application

3.6.1 Brain Networks for Recreational Drug Users and Non-users

We employ the proposed method to learn brain functional connectivity networks using the Human Connectome Project (HCP) resting-state fMRI data [221]. HCP has acquired high-quality functional and structural imaging data from healthy adult subjects in an effort to enhance the understanding of neural connectivity. Here, we focus on analyzing the resting-state fMRI data of 160 unrelated subjects, among which 80 are recreational drug users, and

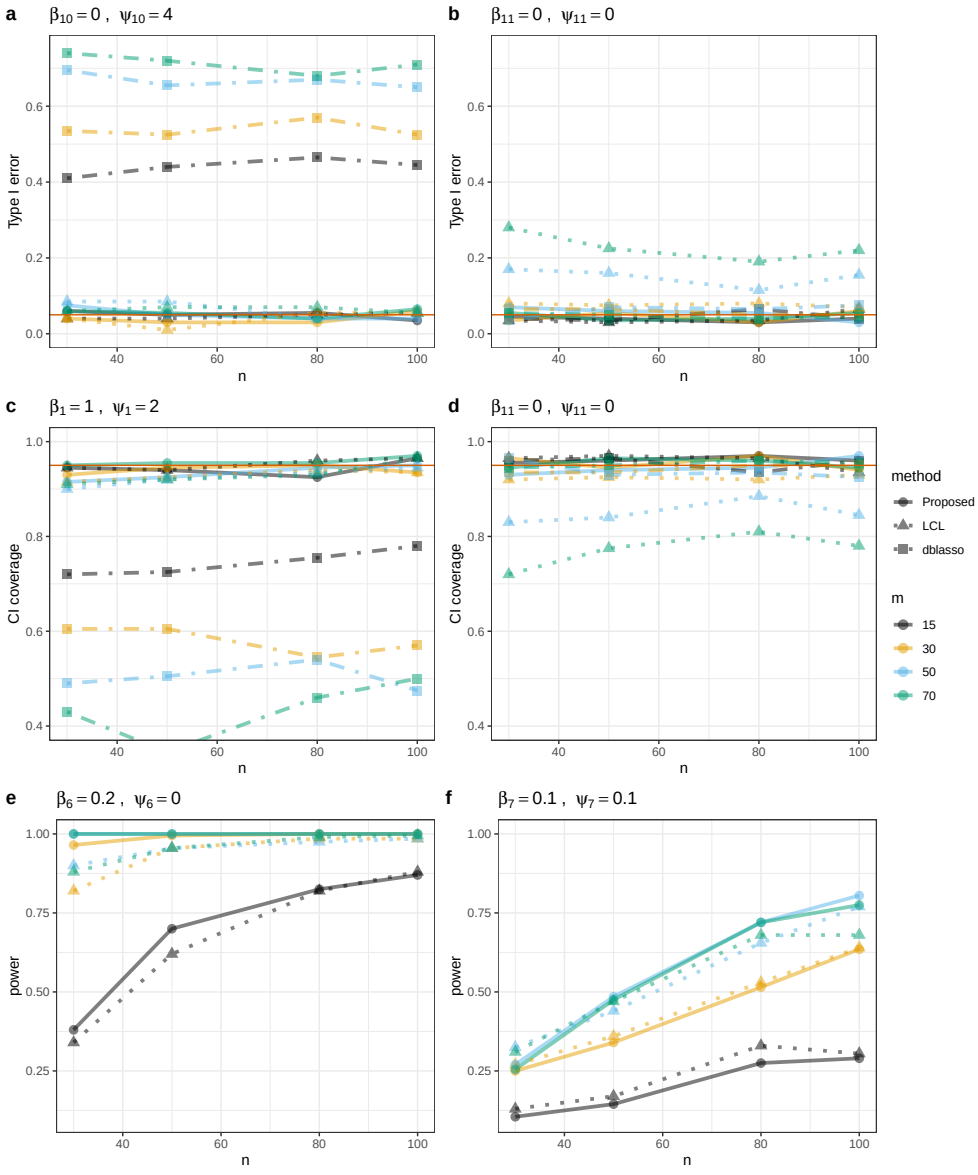

Figure 3.2: **a**,**b**: Type-I error for testing β_l's at the 0.05 significance level (0.05 marked by red solid line), for the proposed method, *LCL* and *dblasso*. **c**,**d**: 95% confidence interval coverage (0.95 marked by red solid line) for fixed effect coefficients, for the proposed method, *LCL* and *dblasso*. **e**,**f**: Power for testing fixed effect coefficients at the 0.05 significance level, for the proposed method and *LCL*. All results are computed at $p = 20$ based on 200 Monte Carlo simulations, and are plotted separately for each method, each m, and against increasing n. The title of each subplot shows the true values of the targeted fixed effect coefficient β_l, and the corresponding random effect variance component ψ_l.

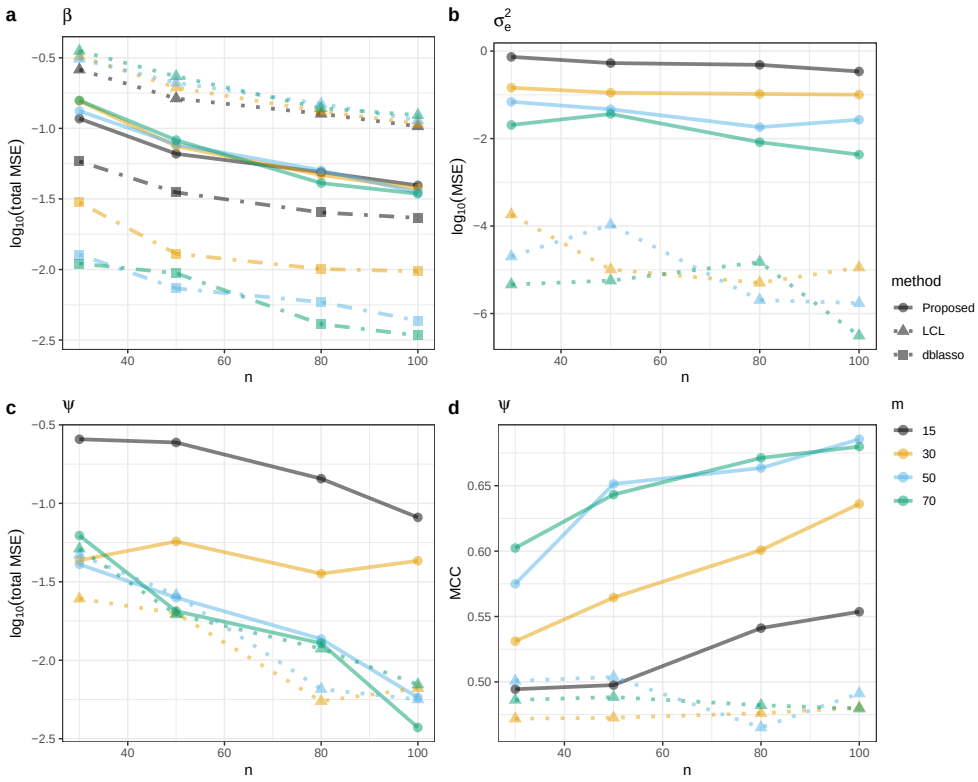

Figure 3.3: **a**: Total MSE on the $\log_{10}$ scale for estimating the fixed effect coefficients β, for the proposed method, *LCL* and *dblasso*. **b**: MSE on the $\log_{10}$ scale for estimating the noise variance σ_e^2, for the proposed method and *LCL*. **c**: Total MSE on the $\log_{10}$ scale for estimating the random effect variance components ψ, for the proposed method and *LCL*. **d**: Average MCC for identifying non-zero variance components for the proposed method and *LCL*, where we use non-zero estimates to select non-zero variance components. Values are computed at $p = 20$ based on 200 Monte Carlo simulations, are plotted separately for each method, each m, and are against increasing n.

the rest are non-users. Each fMRI scan provides a signal with a temporal resolution of 0.73 seconds and a spatial resolution of 2-mm isotropic [205]. Standard pre-processing steps were applied to the fMRI signals [202], including spatial normalization [73] and artifacts removal [86, 188, 205]. Then a Group Independent Component Analysis (ICA) was applied to generate 200 spatially-distributed brain nodes [13, 203]. Using a dual-regression, 200 associated time series of length $1,200$ were obtained, each one representing the activation pattern of a brain node over time [see 205, for more details]. To remove the temporal correlation, we down-sampled each time-series to one of length 120, which we assume consists of independent observations.

We apply the proposed method to the user and the non-user groups separately, estimating and testing the significance of the functional connectivity (network edges) between every pair of brain nodes in each group, and estimating the variance component for each network edge. For the latter, we assume a sparse diagonal covariance matrix for the random effects and we split the samples into three equal-sized sub-samples for estimation. To account for the numerical asymmetry of the neighborhood selection approach, in each group we set the edge $E_{j,k}$ based on $\hat{\beta}_{j\leftrightarrow k} = (\hat{\beta}_{j,k} + \hat{\beta}_{k,j})/2$. The corresponding variance of $\hat{\beta}_{j\leftrightarrow k}$ is computed as $\hat{V}_{j\leftrightarrow k} = (\hat{V}_{j,k} + \hat{V}_{k,j})/2$. The p-value for testing the null hypothesis $H_0 : \beta_{j\leftrightarrow k} = 0$ is then computed based on these averaged quantities.

For comparison, we also apply the *LCL*, the *dblasso* methods and the two-stage approaches using t-test (*two-stage t-test*) or Fisher's method (*two-stage Fisher*) to estimate and test the statistical significance of $\beta_{j\leftrightarrow k}$ for the non-user group. The *dblasso* approach ignores the within-subject correlations and subject-level variations, yielding network strength estimates that are equivalent to applying our model with $a = 0$ in the proxy matrices $\Sigma_a^{(j)}$.

Controlling for the family-wise error rate at 0.05 using Holm's procedure [95], the proposed method detects $1,472$ edges (7.4%) in the non-user group and $1,459$ edges (7.3%) in the recreational drug user group as significantly different from zero. The estimated brain networks in the two groups have $1,028$ edges in common. For better visualization, we only show the results for a sub-network with nodes that are associated with the default mode network, i.e., the brain regions that are active during passive rest and therefore most relevant to resting-state experiments. Also, we only plot the most significant edges (p-values $< 1 \times 10^{-13}$). The estimated sub-networks are shown in Figures 3.4a-b. The proposed method also provides estimates of the variability of each edge. In Figures 3.4c-d, we present the top 25% of edges, selected based on the highest estimated variances, from those depicted in Figures 3.4a-b. The plots show that, for example, the connectivity between node 12 and node 20 presents high subject-level heterogeneity in both groups.

Interestingly, the link between node 3, the posterior cingulate cortex, and node 11, the medial prefrontal cortex, is deemed significant in the non-user group, but not so in the user group. These two areas serve as major hubs within the default mode network [48].

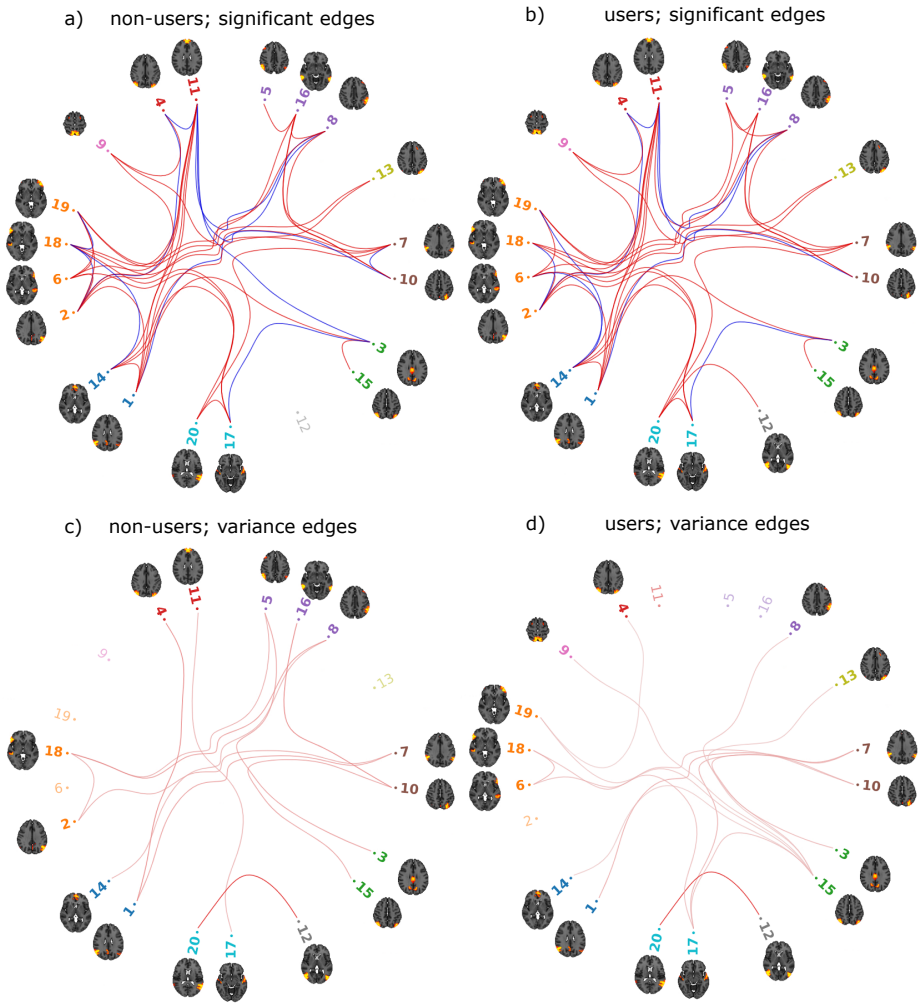

Figure 3.4: Estimated brain sub-networks, and associated variances, in recreational drug users and non-users as estimated by the proposed method. We present the results only for a subset of the 200 nodes that are associated with the default mode network, i.e., the brain regions that are active during passive rest. These were selected based on visual examination of the brain nodes. The node clusters in the networks are defined through hierarchical clustering based on the fixed effect estimates from the proposed method. **a,b**: The most significant edges (adjusted p-values $< 1 \times 10^{-13}$) in each group. Red edges represent positive edge strength, and blue edges represent negative edge strength. **c,d**: The edges in the top 25% based on their variances, selected from those depicted in sub-plots **a** and **b**. The edge width is proportional to the estimate of edge variance, with thicker edges representing higher subject-level heterogeneity.

The disconnection of these regions has been linked to working memory deficits [232]. Even though the primary aim of this work is not to explore this research question, and considering that non-significant p-values merely suggest that the data do not provide sufficient evidence to reject the null hypothesis, these findings corroborate with research that has linked acute cannabis usage to weakening abilities to maintain, manipulate, and recall information [94].

From an estimation accuracy perspective, for the non-user group, *dblasso* estimated 1754 edges (8.8%) as significant connections among 200 brain nodes and *LCL* detects 1326 edges (6.7%); 2648 edges (13.3%) and 1414 edges (7.1%) are detected by *two-stage t-test* and *two-stage Fisher* approaches, respectively. The larger number of significant edges detected by *dblasso* and *two-stage t-test* is likely due to the inflated type-I error, as illustrated in our simulations when subject-specific heterogeneity is present in the data. We illustrate in Figure 3.5 the most significant edges (p-values $< 1 \times 10^{-13}$) detected by *LCL*, *dblasso*, *two-stage t-test* and *two-stage Fisher* among the previously selected brain nodes. While the overall patterns of connectivity are similar among these methods when compared to the proposed method, the stark difference in the number of detected edges by *dblasso* and *two-stage t-test* highlights the importance of modeling the heterogeneity in functional connectivity studies even when the population-level connectivity is of interest. We note that with 200 nodes and 120 observations per subject, *LCL* is not able to estimate the variances of the network edges.

3.6.2 Default Mode Network Alteration Related to Alzheimer's Disease

We apply the proposed method to investigate alterations in brain functional connectivity within the Default Mode Network (DMN) associated with Alzheimer's Disease (AD) using the dataset from the UW Alzheimer's Disease Research Center (ADRC) Imaging and Biomarker Core database. Our focus centers on the analysis of 87 recruited subjects, categorized based on their amyloid beta ($A\beta$) status. Specifically, 62 subjects are classified as healthy controls (referred to as $A\beta-$ subjects), while 25 subjects are identified as AD patients (referred to as $A\beta+$ subjects) (see Section 4.2.1 for detailed classification criteria). Resting-state fMRI scans were acquired for each subject with a temporal resolution of 2.5 seconds. We applied

standard preprocessing steps, the details of which are detailed in Section 4.2.1 in Chapter 4.

For the analysis, we considered eight DMN nodes: left and right superior frontal gyrus (SFG), medial prefrontal cortex (MPFC), posterior cingulate cortex (PCC), left and right parahippocampal cortex (PHC), left and right posterior inferior parietal lobule (pIPL). Utilizing Group ICA (FSL MELODIC, [113]), we generated a brain mask for each node and obtained associated time series observations of length 288. These time series were then down-sampled to a length of 72 to remove temporal correlations.

We followed the same procedures as outlined in Section 3.6.1 to apply the proposed method, independently learning brain functional connectivity networks for the $A\beta-$ and $A\beta+$ groups. Given the small number of edges to estimate relative to the sample size in this application, we also employed the standard LMM approach to estimate model (3.2) as a comparison. Under the standard LMM method, we utilized the standard Wald test to compute p-values for testing each edge.

Controlling the family-wise error rate at 0.05 using Holm's procedure [95], we identified the same 12 significant edges (42.9%) in the $A\beta+$ group with both the proposed method and the standard LMM method. In the $A\beta-$ group, the standard LMM method detected two additional edges beyond the 14 edges (50.0%) identified by the proposed method (see Figure 3.6). These additional edges involve the connection between PCC and PHC right, as well as the connection between PHC left and pIPL right. The consistency between the results obtained by the proposed method and the standard LMM method underscores the validity of our approach in learning brain connectivity networks.

Analysis using the proposed method revealed significant functional connectivity differences between the $A\beta-$ and $A\beta+$ groups. Specifically, the $A\beta+$ group exhibited reduced functional connectivity between SFG left and MPFC, as well as between PHC left and pIPL left, compared to the $A\beta-$ group. These findings align with prior research demonstrating decreased connectivity within the DMN associated with aging and cognitive impairment [115, 173].

We also present the estimated variability of the edges in Figure 3.7. For instance, based

on the proposed method, the connection between PHC left and pIPL left exhibited greater heterogeneity in $A\beta+$ subjects compared to $A\beta-$ subjects. While both the proposed method and the standard LMM method identified a similar set of heterogeneous edges for the $A\beta-$ group, there were notable differences in the estimated edge heterogeneity for the $A\beta+$ group: the proposed approach indicated more pronounced subject-level connectivity variations within the DMN for the $A\beta+$ group.

3.7 Discussion

Motivated by the problem of inferring population-level edges in subject-varying GGMs, we proposed an estimation and inference framework for fixed effect parameters in doubly high-dimensional LMMs. As shown in our numerical studies, ignoring the subject-level variations in GGMs can result in highly-inflated type-I errors and false discoveries, even when the population-level network is of main interest. However, because of correlation and overlap between fixed and random effect covariates, previous works on high-dimensional LMMs, including [132], do not apply to the GGM setting and may suffer from inflated type-I error and insufficient coverage of confidence intervals. Our estimation and inference framework for doubly high-dimensional LMMs addresses these challenges, while also allowing a larger number of random effects than the number of observations per subject. We also proposed a new moment-based penalized variance component estimator, which addresses the challenges of estimating variance components in doubly high-dimensional LMMs.

Similar to [132], we treat the parameter a in the proxy covariance matrix as a tuning parameter and use a cross-validation procedure to select a. However, our theoretical results hold for any positive value of the constant a, and, in practice, we can simply set a to an arbitrary constant. Our additional simulation studies in Appendix B.5.2 explore the effect of a on the finite sample performance of the proposed method and show that, for any constant $a > 0$, the proposed method has correct confidence interval coverage and controls the type-I error. However, the value of a may impact the power of detecting non-zero β_l and the estimation accuracy of β and the variance components.

The concept of a proxy covariance matrix, as previously employed in [132, 61, 23], presents

a significant advancement in the estimation and inference framework by separating the fixed effect coefficients and the variance components. This facilitates extending the classic de-biased LASSO framework to accommodate correlated observations in multi-subject experiments. Beyond serving as a 'de-correlation' matrix, as discussed in Section 3.2.3, the proxy covariance matrix can also be viewed as a weighting matrix in the likelihood function [132], or a working covariance matrix in the Generalized Estimating Equation (GEE) framework. While our proposed framework shares connections with the GEE framework such as adopting working covariance and sandwich-type variance estimator, notable distinctions exist. In the conventional GEE framework, the working covariance matrix commonly remains constant across subjects, often adopting the identity matrix in practice. In contrast, our proxy covariance matrix, inherently a working covariance matrix replacing the true covariance matrix, incorporates subject-level random covariates to construct subject-specific working covariance matrices using the 'sandwich' structure. This approach approximates the eigenvalues of the true covariance matrix (Proposition 2.2 in [132]), which is a crucial aspect of the theoretical analysis. On the other hand, if an identity matrix is adopted as the proxy/working covariance matrix, as discussed in Section 3.2.3, the resulting estimator for β aligns with the classic LASSO estimator. Although this choice forfeits the eigenvalue approximation to the true covariance matrix, the consistency of the fixed effect coefficient estimator persists. However, its convergence rate is slower due to the additional random effect component in a linear mixed model compared to a simple linear regression model [132]. We refrained from exploring the theoretical properties of the proposed framework using an identity proxy covariance matrix, as many of our proof techniques are specifically designed for the sandwich format of the proposed proxy covariance that involves the fixed effect design matrix. Nevertheless, it is worth noting that certain fundamental results, such as Lemma B.1.2, remain applicable and can contribute to the theoretical analysis. Furthermore, our empirical investigations, conducted through an extensive simulation study, offer some insight (detailed in Appendix B.5.3): specifically, when using the identity proxy covariance matrix as in the standard GEE framework, we observed mild under-coverage of confidence intervals and slight inflation of type-I errors, particularly with small sample sizes ($n = 15$). Additionally, we noted a decrease in power compared to the proposed method with the

intended proxy covariance matrix. This power reduction may be attributed to the loss of information resulting from not using random covariate observations in the proxy covariance matrix. We also examine the empirical performance of the naive de-biased LASSO inference framework compared to the proposed framework with an identity proxy covariance. While the two inference frameworks are highly similar, the former fails to perform consistently, whereas the latter demonstrates robust performance in most cases. This discrepancy can be attributed to a key difference in their formulations of the variances of the de-biased estimates. The proposed framework utilizes a sandwich-type estimator for these variances, which can be robust across various specifications of proxy covariance structures and correlations among observations, whereas the naive de-biased LASSO framework is vulnerable to model misspecification.

A novel aspect of the proposed method, compared with the existing procedures, is that it accommodates subject-varying covariance matrices. This is a crucial feature for handling graphical modeling with subject-level heterogeneity. Our theoretical analysis avoids making specific distributional assumptions, allowing for a broad spectrum of potential covariance matrix distributions to be considered. Exploring these diverse choices could be an interesting direction for future research.

Two other extensions of our framework could be of interest. Firstly, motivated by our GGM problem, we focused on the setting where the fixed and random effect design matrices are identical. We also assumed $p/m > c_0 > 1$ for some constant c_0. In the Appendix B.1–B.4, we extend our framework to generic doubly high-dimensional LMMs, where the random effect covariates are a subset of the fixed effect covariates. Assuming fewer random than fixed effect covariates ($q \leq p$), we show that our framework works if $\lim_{q,m\to\infty} q/m \neq 1$. However, these results are based on unified proof techniques for $q/m > c_0 > 1$ and $q/m < c_1 < 1$, with techniques specifically designed for the case $q/m > c_0 > 1$. Therefore, our assumptions for the setting $q/m < c_1 < 1$ could be relaxed and our rates may not be optimal. Secondly, our framework extends beyond independent and identically distributed noise terms. The proposed method readily accommodates observations featuring correlated noise terms and seamlessly extends to mixed-effect vector autoregressive models. This makes the mode a

versatile tool that can be directly employed in the analysis of time series observations in certain settings.

In summary, our framework provides rigorous inference of brain connectivity networks in the presence of subject heterogeneity during multi-subject experiments. It facilitates a systematic exploration of population-level brain network structures, enhancing our understanding of the functionalities associated with various brain regions. It aids in identifying crucial brain regions, such as those highly connected with many others, providing targeted avenues for further investigations. Additionally, it helps detecting AD-induced alterations in brain connectivity patterns [126, 79]. Beyond its utility in resting-state fMRI data, our framework is directly applicable to task-based fMRI signals. This adaptability is particularly beneficial in scenarios where substantial differences in brain functional connectivity are expected between distinct groups [251].

	Edge					
	1—2	1—3	1—4	1—5	1—6	1—7
fixed effect coefficients β	0.50	-0.40	0.20	0.40	0.00	0.00
SD of random effect coefficients	1.50	0.00	0.50	0.75	0.00	0.50

Approach	Power				Type-I error	
Two-stage t-test	0.255	0.795	0.165	0.460	0.050	0.040
Two-stage Fisher's method	1.000	0.380	0.660	0.925	0.075	0.500
Fixed GGM	0.750	0.755	0.310	0.605	0.035	0.150
Mixed effect GGM	0.285	0.980	0.240	0.505	0.025	0.055

Table 3.1: Toy example where we infer the group-level conditional dependence using a neighborhood selection approach, and target the connection between node 1 and the rest of the six nodes. We infer using two-stage approaches, the fixed GGM approach and the mixed effect GGM approach. We generate data from a conditional model following model (3.2), where the fixed effect coefficients and the standard deviation (SD) of the corresponding random effect coefficients are shown in the top table. The noise terms independently and identically follow the standard normal distribution. For each of the 20 subjects, we simulate 10 observations. We then use the neighborhood selection approach to infer the network edges. For the fixed GGM approach, we concatenate observations from all subjects and fit a single GGM using a simple linear regression model. For two-stage approaches, we use simple linear regression in the first stage for each subject, then in the second stage use either Fisher's method to aggregate p-values [47] or use t-test on each subject's coefficient estimates [155]. For the mixed effect GGM approach, we use a linear mixed-effect model with each subject as a cluster (see Section 3.2 for model details). Power/Type-I errors are computed based on 500 replications. The simulations show that when there is high subject-level heterogeneity, two-stage approaches may have highly-inflated type-I error (Fisher's method for edge 1—7), or have lower power to detect dependence (t-test method) compared to the mixed effect GGM approach. The fixed GGM approach shows an inflated type-I error for edge 1—7.

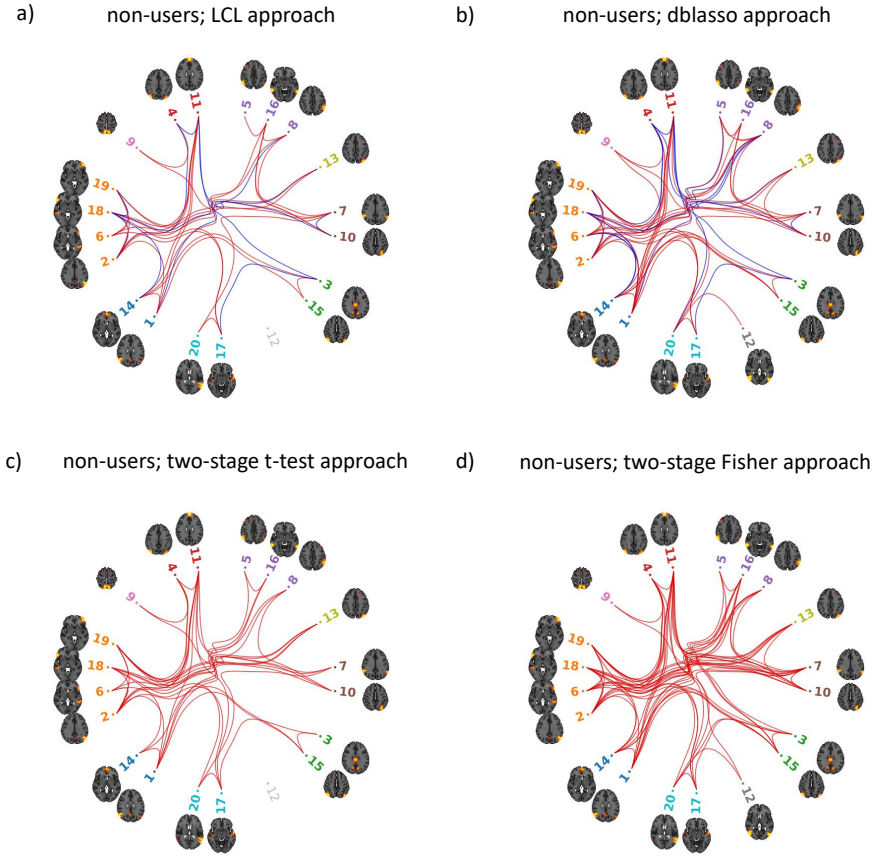

Figure 3.5: Brain sub-networks in non-users as estimated with the **a**: *LCL* approach; **b**: *dblasso* approach; **c** *two-stage t-test* approach; **d** *two-stage Fisher* approach. We present the results only for the nodes we selected in Figure 3.4. Only edges with p-value $< 1 \times 10^{-13}$ are presented. Red edges represent positive edge strength, and blue edges represent negative edge strength. For two-stage approaches, the edges are tested without computing population-level edge estimates, so we plot all the edges in red.

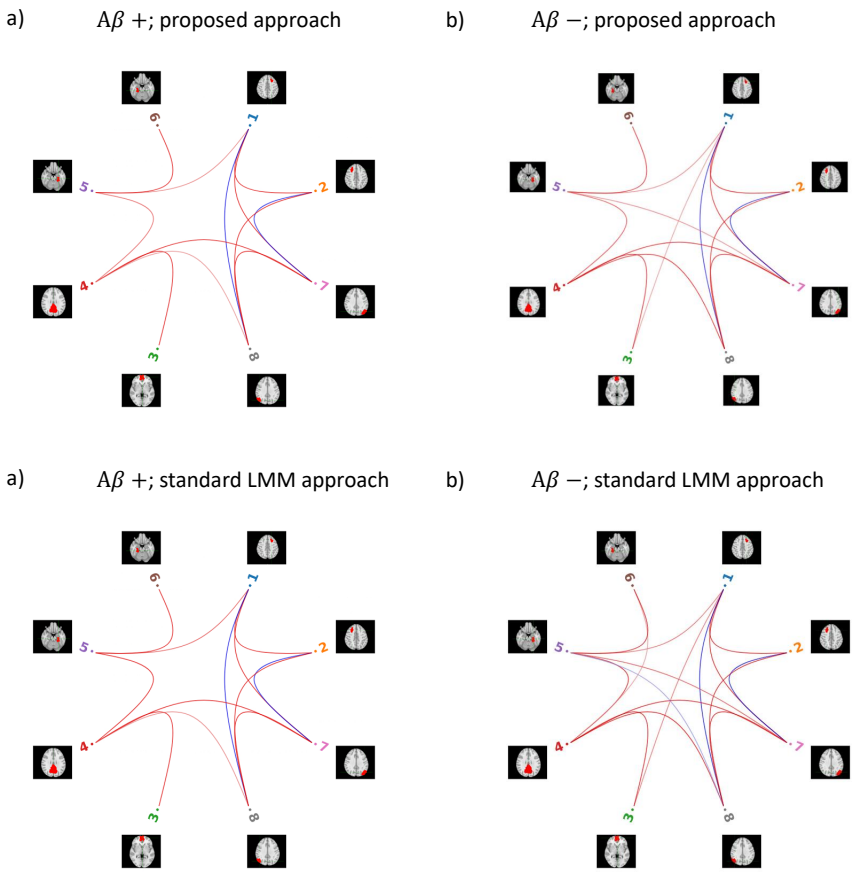

Figure 3.6: Estimated DMN functional connectivity networks in $A\beta+$ and $A\beta-$ groups as estimated by the proposed method and the standard LMM method. Red edges represent positive edge strength, and blue edges represent negative edge strength. Edges with adjusted p-values smaller than 0.05 are plotted.

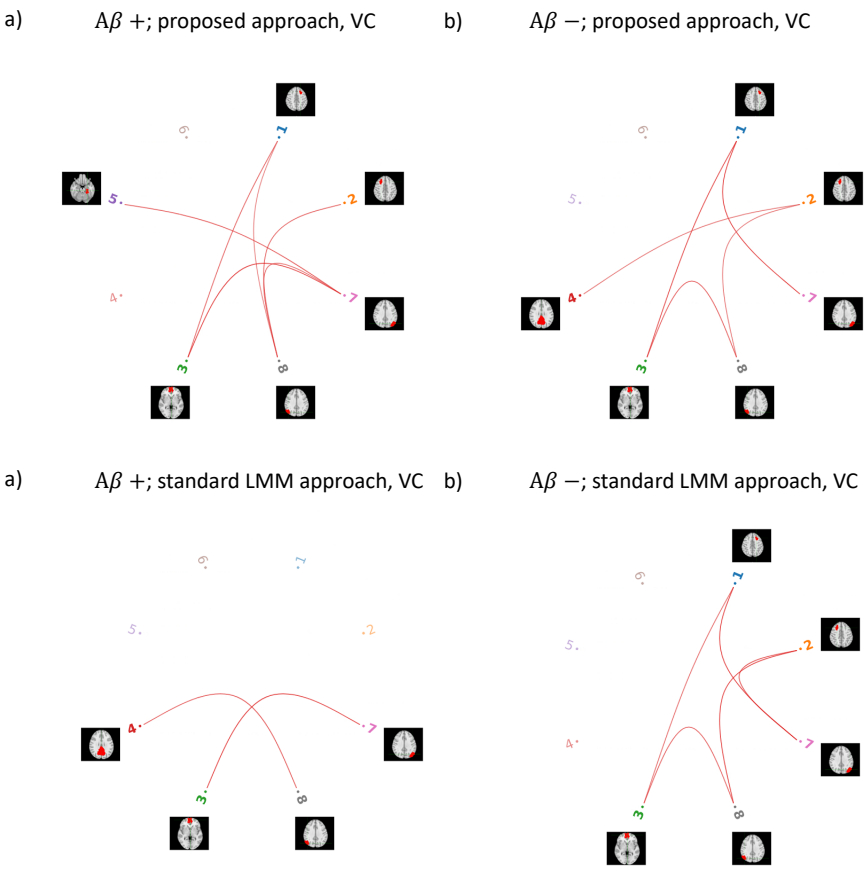

Figure 3.7: Estimated variance associated with the edges in the DMN functional connectivity networks for $A\beta+$ and $A\beta-$ groups, using the proposed method and the standard LMM approach. The edge width is proportional to the estimate of edge variance, with thicker edges representing higher subject-level heterogeneity. Only edges in the top 25% based on their variances are plotted.

Chapter 4

UNRAVELING ALZHEIMER'S DISEASE: INVESTIGATING DYNAMIC FUNCTIONAL CONNECTIVITY IN THE DEFAULT MODE NETWORK THROUGH ADVANCED WHITENING PROCEDURES AND DCC-GARCH MODELING

4.1 Introduction

Alzheimer's disease (AD) is characterized by a prolonged latent phase before noticeable symptoms manifest. This emphasizes the importance of developing sensitive biomarkers capable of identifying the preclinical stages of the disease. Accurate biomarkers that can detect AD in its early phases would enable the implementation of disease-modifying therapies that might prevent cognitive impairment [191].

In 2018, the National Institute on Aging and the Alzheimer's Association (NIA-AA) introduced a groundbreaking research framework for AD, which fundamentally redefines the disease based on its underlying pathological processes, observable through biomarkers [105]. Unlike traditional diagnostic methods relying on clinical consequences or symptoms, this novel framework transforms the definition of AD in living individuals from a syndromal concept to a biological construct. According to this paradigm, the initiation of AD pathological changes in the brain begins with the formation of amyloid beta ($A\beta$) plaques, and subsequent presence of neurofibrillary tangles confirms the transition to AD. This biomarker-centered definition remains independent of clinical symptoms, indicating that biomarkers are poised to hold a central role in future clinical and research settings for AD.

Currently, widely accepted biomarkers of $A\beta$ deposition include cortical amyloid PET ligand binding or low levels of Cerebrospinal fluid (CSF), $A\beta 42$. Normalizing CSF $A\beta 42$ concentration to the level of total $A\beta$ using the $A\beta 42/40$ ratio can enhance the differentiation between AD and controls [7, 131, 130, 108, 6]. Despite their significant success and invaluable contri-

butions to research, these biomarkers may face challenges in population-based and clinical settings due to limited availability, cost, associated radiation exposure, and invasiveness.

More recent work indicates that soluble tau species (181-p-tau and 217-p-tau) also correlate with significant $A\beta$ deposition [9, 153]. However, despite the high cost, fluid markers of Alzheimer's pathophysiology do not give a sense of the neurophysiological impact of the disease. Consequently, there is an urgent need for widely accessible, non-invasive AD biomarkers that can accurately detect the early stages of the physiologic disruptions of AD. Despite substantial efforts in this field, the development of such biomarkers remains a formidable challenge.

The identification of functional connectivity (FC) changes that correlate with Alzheimer's disease (AD) pathology has suggested their potential use as non-invasive biomarkers for $A\beta$ deposition [126, 79, 45, 8]. Reduced FC within the default mode network (DMN) has been observed in the early stages of AD [85, 209, 225, 247]. This reduction is associated with $A\beta$ pathology [100, 146, 28, 93, 195, 156]. Resting-state functional MRI can measure these changes non-invasively. However, despite encouraging results at the group level, current FC measurements do not yet have the sensitivity to act as standalone biomarkers [217].

The limited sensitivity of FC measurements may be partly attributed to the prevailing assumption of stationarity of FC in many studies. This assumption implies constant FC between brain regions throughout the entire scan duration (typically 5 to 10 minutes), represented by a single parameter derived from several minutes of data [16, 194]. However, evidence from electrophysiology suggests that FC undergoes dynamic changes within seconds to minutes [102, 34, 213, 168], exhibiting variations in both strength and direction [89, 187, 117, 2]. In the resting state, where mental activity is unrestricted, these dynamics may be even more pronounced.

The sliding-window approach has been a common approach for studying dynamic temporally-varying brain connectivities [2, 34, 89]. In this approach, the FC is measured by correlation, and a fixed-length window is employed to compute the correlation coefficient using data within the window. By shifting the window incrementally across time, a time-varying

connectivity measurement between brain regions can be obtained. Although simple to implement, this approach has several important limitations. These include the utilization of arbitrary window lengths, the inability to handle abrupt changes, and the uniform weighting assigned to all observations within the selected window while disregarding older observations [137]. In particular, the choice of window length significantly influences the analysis results. Overly long window lengths obscure the true temporal dynamics in FC, diminishing the ability to capture essential information. Excessively short window lengths introduce inflated variance and may confound noise with genuine signals, potentially leading to misleading interpretations [96].

To overcome the limitations associated with the sliding-window approach, we turn to model-based methods for characterizing dynamic FC profiles. The Generalized Autoregressive Conditional Heteroscedastic (GARCH) model family emerges as a versatile choice for modeling time-dependent second moments in time series observations. Originally proposed for dynamic standard deviations in univariate time series [18], GARCH models have been extended to multivariate data to capture the dependency of dynamic covariance and correlation matrices on past information. Despite their flexibility in covariance structure, early models like VEC-GARCH [20] and BEKK [59] face challenges related to computational demands and interpretability [201]. In light of our focus on FC that corresponds to correlation structures, we direct our attention to GARCH models explicitly designed to capture dynamics in correlation matrices. The Constant Conditional Correlation GARCH model [19] and its extensions, such as Varying Correlation GARCH [220] and Dynamic Conditional Correlation (DCC-) GARCH [60, 58], offer parsimonious descriptions of time-varying correlation structures. While subsequent models like the Asymmetric Generalized DCC-GARCH [33] offer enhanced flexibility, the increased number of parameters complicates interpretation. Considering our interest in modeling temporally dynamic correlation structures and our objective of detecting and interpreting alterations in dynamic correlation patterns due to AD, we opt for the application of the DCC-GARCH model. This model, recently introduced in neuroscience studies [137, 180, 127], provides a plausible mechanism to capture temporal variations in brain networks while maintaining interpretability. The DCC-GARCH model's

parameters, particularly α and β (to be explained in Section 4.2.3), offer valuable insights into the dynamic behavior of the FC system over time. We demonstrate that our method provides superior sensitivity compared to stationary FC measurement methods, enabling the detection of subtle pathological changes in AD and facilitating the development of non-invasive, MR-based AD biomarkers.

To meet the DCC-GARCH model's assumptions, the time series observations for each node should exhibit no serial correlation after removing the mean signal [137, 180, 127] (see Section 4.2.3 for more details). To achieve this, we apply a whitening step as a common preprocessing approach to mitigate or eliminate serial correlations in time series observations. This procedure involves estimating the serial correlations from the observed data and utilizing the inverse correlation matrix for de-correlating the observations. While the literature on applying DCC-GARCH models to dynamic brain networks lacks relevant discussions on whitening, the broader fMRI analysis literature extensively discusses available approaches and the necessity of whitening in relevant situations [165, 128, 25, 234, 145, 42]. In fact, the existence of serial correlation gives rise to significant challenges in the analysis of fMRI data. For example, the usual standard error of the sample correlation coefficient between two time series is biased, which could lead to incorrect conclusions in linear regression analysis [1]. These challenges are particularly pertinent when using a generalized linear regression model, the most common method for analyzing fMRI data, to explore the relationship between fMRI observations and task regressors in task-fMRI studies. Eklund et al. [57] have substantiated the considerable role of poorly modeled temporal autocorrelation in the inflation of type-I errors. To address the serial correlations, there are currently many whitening approaches available, among which the autoregressive moving average (ARMA) model with fixed order has been a standard practice for whitening time series observations in popular fMRI software packages. Examples include Analysis of Functional NeuroImaging (AFNI) [43] using a voxel-specific ARMA(1,1) model [36], and Statistical Parametric Mapping (SPM) [170] as well as FMRIB Software Library (FSL) [113] using a global AR(1) model [68]. However, while most of these traditional approaches were developed for fMRI data acquired with longer repetition times (TRs) of approximately 2 seconds, they may not

adequately whiten long TR datasets (see 4.3.1 for details). In addition, recent advancements in imaging technology have enabled the acquisition of fMRI measurements with much shorter TRs ($\leq$ 0.5s), resulting in even higher levels of serial correlations with slower decay rates per time-point. Traditional whitening procedures are not capable of capturing the serial correlation for short-TR observations, primarily due to the utilization of low model order, such as AR(1), as well as the assumption of uniform AR coefficients across the brain [186, 22, 145]. Recent studies suggest that using higher-order ARMA models with spatially-varying coefficients can effectively whiten the time series, even when dealing with sub-second TRs [145, 186]. However, as we will demonstrate in Section 4.3.1, these methods may not sufficiently remove the serial correlations in short-TR datasets. Furthermore, as will be illustrated in Section 4.3.1, different whitening approaches can yield significant variations in resulting signals, subsequently influencing downstream analyses and conclusions. This underscores the pressing concern of replicability within fMRI research.

To address the limitations of current whitening approaches and properly removing the serial correlations for subsequent DCC-GARCH modeling, we propose an iterative whitening procedure, named *IDAR* (for Iterative Data-adaptive AR). IDAR leverages high order AR models with a flexible and data-adaptive order, where the iterative nature can better model the complicated serial correlation structure that is often not adequately captured by a single AR model. IDAR is applicable to both short-TR and long-TR fMRI datasets, providing a universally adaptable whitening approach that is capable of accommodating diverse fMRI dataset characteristics.

In addition to assessing the performance of IDAR through the residual serial correlations post-whitening, we also explore the impact of whitening on type-I errors in simulated task-fMRI using obtained resting-state fMRI data. Previous literature has not concurrently examined both aspects of whitening procedure performance, making our investigation comprehensive in nature. We note that IDAR provides users with the flexibility to correct either serial correlations or type-I errors based on their specific requirements. When the primary concern is controlling the type-I error, the IDAR approach with only one iteration seems most effective. On the other hand, if the focus is on removing serial correlations, the IDAR approach can be

applied with a few iterations. This adaptability empowers users to customize their analyses to align with their particular objectives, effectively addressing the challenges associated with short-TR fMRI data.

4.2 Modeling Dynamic Functional Connectivity

Figure 4.1 illustrates the details of the analysis procedures. The collected resting-state fMRI data undergoes standard preprocessing. Next, we compute time series observations for each node in the default mode network (DMN). These observations are obtained using a brain mask created through group independent component analysis (Group ICA) (FSL MELODIC, [113]). Subsequently, we apply the proposed IDAR model to the node-wise time series observations for whitening. To assess the efficacy of the whitening algorithm, we evaluate it from two perspectives: (1) the extent of residual serial correlations after whitening, and (2) the impact of whitening on power and type-I error within the context of simulated task-fMRI using resting-state fMRI data. Then, the DCC-GARCH model quantifies the dynamic functional connectivity for each subject. Finally, we conduct a group comparison analysis to investigate the differences in dynamic connectivity patterns between AD patients and control subjects.

4.2.1 Data

We employed two datasets in this study. The primary dataset has a long TR at 2.5s. It serves the dual purpose of studying the functional connectivity in AD and assessing the performance of whitening procedures. The secondary dataset, characterized by a short TR at 0.49s, is exclusively employed for the evaluation of whitening approaches.

Main Dataset (Long TR)

The dataset included 87 subjects (mean age = 74 yr, range = 54 to 91 yr). All participants enrolled in UW ADRC (NIH P50 AG005136) Clinical Core, and were recruited with either normal cognition or mild cognitive impairment. All MRI data were acquired using a research-dedicated 3T Philips Achieva scanner equipped with a 32-channel receiver coil. During the

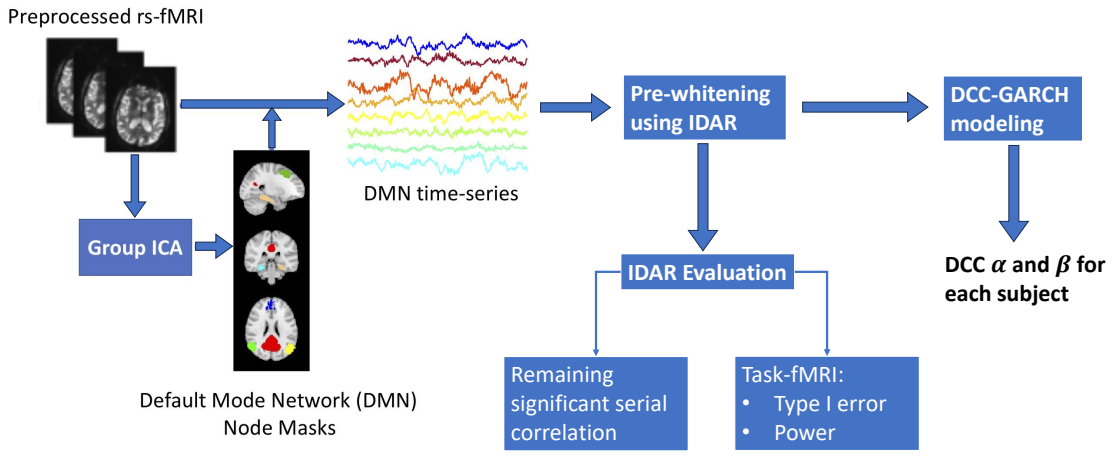

Figure 4.1: Flow chart of the analysis procedure.

scanning session, subjects were instructed to lie in the scanner with their eyes open while wearing a Pearl-Tec® Crania to minimize head motion.

For registration purposes and volumetric measurements, we acquired 3D T1 data from all participants. The 3D structural images were obtained using a conventional MPRAGE sequence with the following parameters: spatial resolution $= 1 \times 1 \times 1$ mm^3, 150 slices, flip angle $= 8°$, TR/TE $= 8.8/4.6$ ms, SENSE acceleration factor $= 2$, and matrix size $= 256 \times 256 \times 176$.

Resting-state fMRI data was collected using axial whole-brain multi-echo T2-weighted sequence over a 10-minute scan period. These images used 3.5 mm isotropic voxels and had TR/TEs of 2,500/9.5, 27.5, and 45.5 ms. Following this, we carried out motion correction using FSL's MCFLIRT software [113]. Non-brain matter was then removed using the FSL Brain Extraction Tool [113]. The multi-echo BOLD data were then processed using AFNI's specialized module TEDANA, for multi-echo planar imaging and analysis with independent component analysis (ME-ICA). TEDANA allows for the differentiation between BOLD (neuronal) and non-BOLD (artifact) components, leveraging the characteristic linear echo-

time dependence of BOLD T2 signals [123]. This approach produced recombined images optimally weighted across the three echo times, along with ME-ICA-denoised time series and spatial component maps. Subsequently, the data were co-registered to T1 images using FSL. After that, high-pass temporal filtering was applied, and the global signal and motion parameters were regressed from the data.

CSF samples were collected via lumbar puncture from all subjects. Levels of $A\beta 42$ and $A\beta 40$ in CSF samples were measured using multiplex bead assays (Luminex xMAP) run on the Luminex 200 instrument according to the manufacturer's instructions [90]. Based on their CSF $A\beta 42/40$ ratio, we divided the subjects into two groups: $A\beta+$ and $A\beta-$, using the internally determined cut-off threshold of 0.11 [63]. 25 subjects with CSF $A\beta 42/40 < 0.11$ are classified as $A\beta+$ AD patients, and the rest 62 subjects are $A\beta-$ control subjects.

Additional Evaluation Dataset (Short TR)

The short-TR data were collected from a cohort of 99 subjects enrolled in Stanford ADRC using a 3T GE Discovery MR750 scanner equipped with a 32-channel head coil (NOVA Medical). Resting-state fMRI data were acquired utilizing a Simultaneous Multi Slice (SMS) EPI with blipped controlled aliasing in parallel imaging (CAIPI) sequence [193]. The imaging parameters were set as follows: TR/TE = 490/30 ms, multiband acceleration factor of 5, CAIPI shift of FOV/3, field of view (FOV) = 24×24 cm^2, voxel size = $3.14 \times 3.14 \times$ 4mm, and scan duration of 10 minutes [106]. Prior to the functional scans, a high order shimming procedure was implemented to reduce B0 inhomogeneity [119]. The SMS EPI images were reconstructed using the SENSE/GRAPPA combination method [17]. K-space calibration data were designed to correspond to a one-dimensional undersampling along the phase encoding direction, with an empirically set interpolation kernel of 7×4 (readout $\times$ phase encoding) acquired points for each missing point.

A high-resolution 3D T1-weighted image was acquired before the resting-state fMRI scans for spatial normalization and anatomical reference, utilizing an IR-SPGR sequence with the following parameters: TR/TE/TI=8.18/3.2/900 ms, matrix=256×256, in-plane resolution = 0.94×0.94mm, and slice thickness/slice spacing = 1/0mm, covering 176 sagittal slices.

The acquired resting-state fMRI images underwent preprocessing and analysis using various tools from FSL, AFNI, and custom MATLAB code. After discarding the first 6 volumes to ensure magnetic stabilization, the images were motion-corrected using least square minimization and normalized to the MNI152 standard space using the subject's high-resolution T1-weighted image with an affine linear registration technique involving 12 degrees of freedom [112]. To remove low-frequency signal drifts (<0.01Hz) from the data, a high-pass filter was applied. Additionally, two band-stop temporal filters targeting respiratory ($[0.25 - 0.35]$ Hz) and cardiac ($[0.8 - 1.02]$ Hz) frequency bands were applied using 5th order Butterworth IIR filters [106]. Various sources of variance including movement [144], cerebrospinal fluid, white matter [44, 230] and global signal [49, 84] were removed using multiple linear regression. Variance associated with movement parameters was estimated using FSL across 6 dimensions (lateral, vertical, and horizontal translation, and yaw, pitch, and roll rotation). Cerebrospinal fluid and white matter confounds were calculated from a 3mm spherical ROI placed in the ventricles and the white matter, respectively.

4.2.2 Whitening: IDAR

The proposed IDAR for whitening fMRI observations is an iterative and data-driven whitening model based on high order AR models. The basic concept involves estimating serial correlations through AR models, with the selection of AR orders guided by the corrected Akaike Information Criterion (AICc) [101]. The iterative procedure is motivated by the fact that persistent residual serial correlations are frequently observed when attempting to de-correlate the data using a single AR model, even when a high order AR model is used. This phenomenon is particularly evident when processing datasets with short TRs. This observation suggest that the serial correlation structure may not be adequately explained by a single AR model. We thus adopt an iterative approach that estimates any remaining serial correlations of the whitened observations from the preceding iteration. The whitening procedure is described in Algorithm 3 for a univariate time series $Y(t)$ with T total observations.

When the algorithm stops at iteration K, the resulting whitened time series, $Y^{(K)}$, has the cor-

Algorithm 3 Algorithm for whitening based on IDAR

Set $Y^{(0)} = Y(t)$, $i = 0$;
while i < max.iteration **do**
Step 1: Fit the following models to the time series $Y^{(i)}$:
(1.1) AR(p) models, with lag $p = 0, \ldots, p_{\max}$.
(1.2) For $l = 1, \ldots, p_{\max}$, we compute the $\rho_l^{(i)}$, which is the l-th lag serial correlation of $Y^{(i)}$. If none of the correlations satisfies $\left|\rho_l^{(i)}\right| > z_{0.975}\sqrt{1/T}$, where $z_{0.975}$ is the 97.5% quantile of the standard normal distribution, we do not fit new models. Otherwise, suppose for $l \in S^{(i)}$ we have $\left|\rho_l^{(i)}\right| \leq z_{0.975}\sqrt{1/T}$, we fit the following two models:
(1.2.1) AR($p_{\max}$) with the l-th lag AR coefficient set to zero, for $l \in S^{(i)}$.
(1.2.2) AR($p_{\max}$) with the l-th lag AR coefficient set to zero, for $l \in \{w \in S^{(i)} : w \leq 0.75p_{\max}\}$.

Step 2: Compute the AICc of all the models fitted in Step 1. Select the model with the smallest AICc value.
Step 3: Based on the selected model, compute the theoretical autocorrelation function $\gamma^{(i)}(h)$, $h = 0, \ldots, T-1$. The estimated correlation matrix at the i-th iteration is $\hat{\Sigma}^{(i)}$: $\hat{\Sigma}^{(i)}_{j,k} = \gamma^{(i)}(|j-k|)$.
Step 4: De-correlate at the i-th iteration: $Y^{(i+1)} = (\hat{\Sigma}^{(i)})^{-1/2} Y^{(i)}$.

if The selected model is AR(0): **then**
BREAK
else
i ← i +1
end if
end while

responding estimated serial correlation matrix $\prod_{i=1}^{K}(\hat{\Sigma}^{(K-i+1)})^{-1/2}\left(\prod_{i=1}^{K}(\hat{\Sigma}^{(K-i+1)})^{-1/2}\right)^{\top}$. We find that setting the maximum number of iterations to 5 is typically sufficient, even for datasets with short TRs. In selecting the best model in each iteration, information criteria such as AIC and Bayesian Information Criterion (BIC) can be employed. For this study, we use AICc due to its relatively superior capability in detecting the appropriate AR orders [192]. Through our investigation, we determine that setting the maximum lag order as $p_{\max} = 10/\text{TR}$ achieves satisfactory performance. This choice assumes that the correlations significantly diminish beyond 10 seconds, which reasonably aligns with the time duration of typical hemodynamic response functions [4, 30, 69, 88]. When dealing with a

short TR and a large number of observations per time series, fitting all $p_{\max}$ models in Step (1.1) would be time-consuming. As a practical solution, we employ the step-wise selection approach (`auto.arima`) implemented in the R package `forecast` [103]. In this approach, we initialize the lag values at $p_{\max}$, $p_{\max}/2$, and $p_{\max}/4$. The inclusion of restricted models in Step (1.2) is motivated by observations made while processing datasets with short TRs. When the TR is small, fitting a full AR($p_{\max}$) model can be both time-consuming and at risk of overfitting. The utilization of restricted models addresses these concerns while still effectively targeting high serial correlations. Specifically, the restricted model employed in Step (1.2.2) incorporates additional higher-order AR lags to ensure the accurate capture of high order serial correlations.

We have observed that the estimated correlation matrices, $\hat{\Sigma}^{(i)}$, can occasionally have extremely large condition numbers, leading to unstable numerical performance. To address this issue, we set a threshold for the maximum condition number at 1×10^8. We progressively set the last 10 non-zero autocorrelations in $\hat{\Sigma}^{(i)}$ to zero until the condition number falls below the specified threshold. Specifically, we will first set $\hat{\Sigma}^{(i)}_{j,k} = 0$ for $|j-k| \geq T-10$, and if the condition number is still above the threshold, continue to set $\hat{\Sigma}^{(i)}_{j,k} = 0$ for $|j-k| \geq T-20$.

Our approach is distinctive in its iterative whitening of the data, stacking multiple AR models to estimate the serial correlation structure. While this necessitates the fitting of additional parameters, we have observed that a singular AR model is often inadequate for whitening, even for long-TR data. Conversely, our algorithm automatically accounts for low-order AR models and single-iteration AR models, rendering it suitable for both short-TR and long-TR data analyses.

Evaluation Approaches

We assess the efficacy of the IDAR-based whitening algorithm from two perspectives: (1) the degree of residual serial correlations after the whitening process, which is especially relevant to resting-state fMRI, and (2) the impact of whitening on power and type-I error in the context of task-fMRI, simulated using resting-state fMRI data. Both the short-TR and the long-TR datasets are used for evaluation.

To evaluate the residual serial correlations, we adopt the evaluation procedure outlined in [42]. After applying the whitening algorithm to a single time series, we utilize the Ljung-Box test [143] at lags ranging from 1 to 20/TR [42]. We control family-wise type-I error using Holm's method [95]. We consider the time series not adequately whitened if the smallest adjusted p-value among the lags is less than 0.05. We calculate the percentage of such inadequately whitened time series, where a successful whitening procedure should yield less than 5% of inadequately whitened time series.

Additionally, whitening procedures are particularly relevant in task-fMRI analyses that study the association between fMRI signals and event paradigms. In such analyses, the presence of serial correlation in the observations diminishes the effective sample size, which potentially leads to inflated type-I error rates in association tests based on the generalized linear regression model (GLM) [170, 145]. Thus, we also examine the applicability of the whitening procedure to task-fMRI analysis. We investigate whether the proposed whitening algorithm effectively controls the type-I error rate. Moreover, we assess whether whitening has any adverse effect on the statistical power. We use simple simulation studies based on available resting-state fMRI data to generate task-fMRI observations. For each subject, we simulate a single task regressor, denoted as X, based on a "boxcar" task paradigm: the time course is divided into blocks of 30 seconds, consisting of a 15-second task epoch followed by a 15-second resting epoch. Each task epoch is randomly assigned as either task A or task B with equal probability, with task A having twice the intensity level of task B. We convolve the boxcar task paradigm with the canonical hemodynamic response function (HRF) to obtain the unscaled task regressor, using the R function `hrfConvolve` from R package `FIAR` (v0.6, [183]) with default double gamma function parameters. Figure 4.2a illustrates the canonical HRF used in the simulation. Considering a signal-to-noise ratio of 0.1, we scale the task regressor such that its mean value equals one-tenth of the standard deviation of the raw resting-state fMRI data [231]. Consequently, the simulated task-fMRI data, denoted as Y_{ts}, are obtained by adding the scaled task regressor X to the raw resting-state fMRI data Y_{rs} (Figure 4.2b). Next, we use the following simple linear model to analyze the association

between the fMRI signal Y and the simulated task regressor X:

$$Y = \beta_0 + \beta_1 X + \epsilon,$$

where β_0, $\beta_1 \in \mathbb{R}$ are the regression coefficients, and ϵ is the error term. In our analysis, we use the simulated task-fMRI data $Y = Y_{\text{ts}}$ as the outcome variable to examine the power of detecting the association between task-fMRI and the task regressor. Additionally, we employ the resting-state fMRI data $Y = Y_{\text{rs}}$ as the null outcome to investigate the type-I error. We use a Wald statistic to test the hypothesis $H_0 : \beta_1 = 0$, where β_1 represents the association between the task regressor X and the outcome Y. The outcome Y will be whitened before fitting the linear regression model, and the regressors will be multiplied with the de-correlation matrix accordingly. Specifically, suppose the whitening procedure estimates a data correlation matrix $\hat{\Sigma}$, we essentially fit the following model to estimate the association between $\hat{\Sigma}^{-1/2}Y$ and $\hat{\Sigma}^{-1/2}X$:

$$\hat{\Sigma}^{-1/2}Y = \beta_0\hat{\Sigma}^{-1/2} + \beta_1\hat{\Sigma}^{-1/2}X + \hat{\Sigma}^{-1/2}\epsilon.$$

In addition to evaluating the proposed IDAR-based whitening procedure, we include standard whitening procedures based on AR(1) model and the ARMA(1,1) models for comparison. Moreover, we investigate the performance of our whitening algorithm with a single iteration, denoted as *IDAR-iter1*. The IDAR-iter1 procedure essentially encompasses the high order AR models proposed in [145, 186]. This comparison provides valuable insights into the performance of existing whitening approaches in the context of both long-TR and short-TR datasets. We also include results obtained from the non-whitened outcome for comparison.

4.2.3 Functional Connectivity Group-level Analysis

We use the primary dataset (long TR) to study the brain functional connectivity networks. We calculated two sets of FC measures from this resting-state fMRI dataset: using static network modeling via a dual regression approach, and dynamic network modeling using DCC-GARCH. Next, we explain the two modeling approaches.

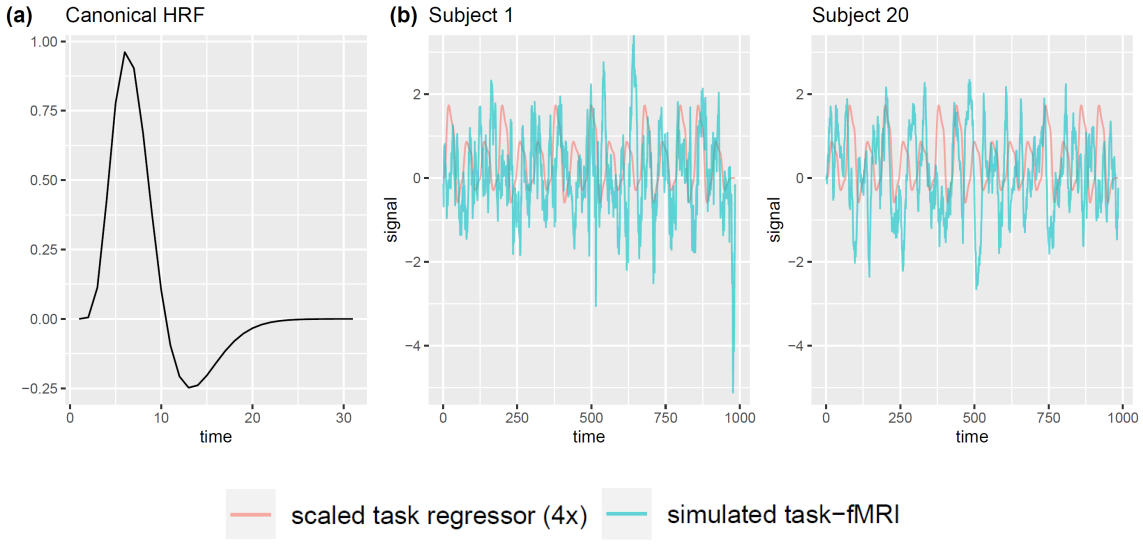

Figure 4.2: Illustration of generating task-fMRI signals from resting-state fMRI observations. **(a)**: The canonical HRF used to convolve with the task paradigm. **(b)**: Two example subjects' simulated task regressor and the corresponding simulated task-fMRI for node PCC.

Static Network: Dual Regression

We measured the static strength of the DMN via the dual regression approach of [164]. We first identified the standard space maps of the group average DMN by independent component analysis (FSL MELODIC) of the group-concatenated time series data. For each subject, the group-average DMN map was then regressed into that subject's 4D space-time dataset using spatial predictors in a multiple regression, yielding a set of subject-specific time series reflecting DMN activity levels. These time series were then regressed back into the same 4D dataset using temporal predictors in another multiple regression, leading to subject-specific DMN maps. For each subject, we computed the average dual regression coefficient of all voxels within the DMN mask (mask obtained from group-average DMN map, thresholded at $p < 0.001$). This average served as the stationary FC measure of the DMN. Higher stationary FC measures imply a stronger average FC throughout the scan

duration.

Dynamic network: DCC-GARCH

We represented the FC of DMN via the conditional correlations among DMN nodes, and applied DCC-GARCH model to describe its temporal variation. The following 8 DMN nodes were considered: left and right superior frontal gyrus (SFG), medial prefrontal cortex (MPFC), posterior cingulate cortex (PCC), left and right parahippocampal cortex (PHC), left and right posterior inferior parietal lobule (pIPL). We extracted the corresponding time-series for each of the 8 nodes by averaging the measured signals over the mask voxels. Before performing the dynamic network analysis, we applied the IDAR-based whitening procedure proposed in Section 4.2.2 to each time series to remove the serial correlations.

Next, we explain the formulation of the DCC-GARCH model. Suppose we have a multivariate time series $\{X(t)\}_{t=1}^{T}$ with $X(t) \in \mathbb{R}^m$. The time series can be decomposed into a mean component $\mu(t)$ and a residual component $\epsilon(t)$:

$$X(t) = \mu(t) + \epsilon(t), \quad t = 1, \ldots, T.$$

The mean process $\mu(t)$ can be modeled as a constant or with a time series model. Here we use a constant mean component. The residual component $\epsilon(t)$ is modeled by $\epsilon(t) = H_t^{1/2} z_t$, where z_t is a white noise process with variance I_m. The conditional covariance matrix H_t conditions on the past information up to time $t-1$. The residual component model indicates $\epsilon(t)$ is serially uncorrelated, even though it is still serially dependent.

The DCC-GARCH model assumes

$$H_t = D_t R_t D_t,$$

where the dynamic conditional covariance matrix $D_t = \text{diag}\left(\{d_{l,t}\}_{l=1}^{m}\right)$ is a diagonal matrix with time-varying standard deviations $d_{l,t}$ along its diagonal, and R_t is the dynamic conditional correlation matrix that represents the dynamic FC among brain nodes. The dynamic

standard deviations $d_{l,t}$ are usually modeled by a GARCH(1,1) process:

$$d_{l,t} = \theta_{l,0} + \theta_{l,1} \left[\epsilon(t)\right]_l^2 + \theta_{l,2} d_{l,t-1}, \quad l = 1, \ldots, m.$$

Higher order GARCH models are also applicable for the dynamic standard deviations, but the GARCH(1,1) model is usually adequate [167, 180].

The dynamic correlation matrix is modeled by

$$R_t = Q_t^{*-1/2} Q_t Q_t^{*-1/2},$$

where $Q_t^* = \mathrm{diag}(Q_t)$ is the diagonal matrix with the diagonal elements of Q_t, and

$$Q_t = (1 - \alpha - \beta)\bar{Q} + \alpha \epsilon_{t-1} \epsilon_{t-1}^\top + \beta Q_{t-1},$$
$$\bar{Q} = \frac{1}{T} \sum_{t=1}^{T} \epsilon_{t-1} \epsilon_{t-1}^\top.$$

This formulation ensures a proper correlation matrix R_t that is positive-definite with diagonal entries equal to 1. The two parameters, α and β, characterize the temporal dynamics in the correlation network: α represents how the correlation matrix dynamically varies as a response to new stimuli, and β quantifies the degree of dependence of the correlations on their past values. In Figure 4.3, we illustrate the dynamic profile of correlations in a simulated 3-node network, highlighting the impact of α and β values. To generate the data from a DCC-GARCH model, we employed the R function `simulateDCC` from the `ccgarch2` package (V0.0.0-42, [158]). For the simulations shown in Figure 4.3, we set the residual covariance matrix as

$$\begin{pmatrix} 1 & 0.8 & 0.3 \\ 0.8 & 1 & 0.1 \\ 0.3 & 0.1 & 1 \end{pmatrix}.$$

Additionally, for the GARCH(1,1) model of the dynamic standard deviations, we set all the coefficient parameters to 0.5. As shown in the plots, small α and β values lead to stable

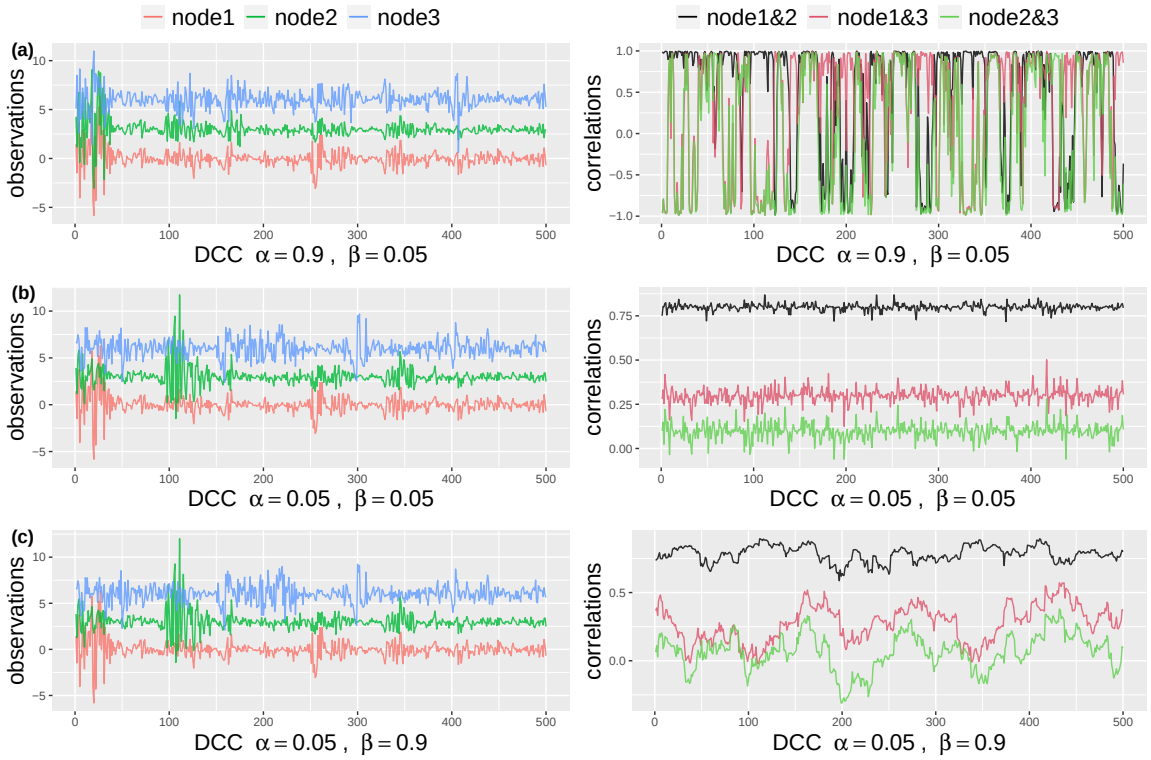

Figure 4.3: Illustration of the simulated dynamic profile of correlations with respect to different α and β values. We simulate three sets of time series observations for a 3-node network based on DCC-GARCH model. Each set has different parameter values for (α, β). The left side plots show the simulated signals from three nodes, and the right side plots show the dynamic profile of the corresponding pairwise correlations.

correlations that resemble white noise (Figure 4.3b). Conversely, when α is small and β is large, the correlation pattern becomes "sticky", changing slowly over time (Figure 4.3c). On the other hand, large values of α and small values of β generate rapid changes in correlations over time (Figure 4.3a). It is crucial to highlight that the dynamic profile of the correlations cannot be identified through simple visual examination of the observed signals. This underscores the necessity of employing appropriate models to accurately capture the temporal dynamics.

To analyze the dynamic correlation profiles from the fMRI data, the DCC-GARCH model

was estimated via maximizing the quasi-likelihood, as implemented using the R package `rmgarch` [70]. After obtaining the subject-level estimates for the dynamic characterization parameters α and β, we collected the estimates and conduct a two-sample t-test for a group-level comparison between $A\beta+$ patients and $A\beta-$ control subjects. We tested for group differences at 0.05 significance level.

4.3 Results

We first demonstrate the superior performance of the IDAR approach compared to other whitening procedures in both short-TR and long-TR datasets. We then show that in the primary dataset (long-TR), dynamic FC measurements modeled via DCC-GARCH show sensitivity to the $A\beta$ status.

4.3.1 Performance Evaluation of Whitening Approaches

We begin by illustrating the raw and whitened resting-state fMRI signals of an example subject (Figure 4.4). For clarity, we only showcase data from three nodes within the DMN. The application of various whitening techniques did not significantly alter the resting-state fMRI profiles, with the resulting signals demonstrating visual consistency across the different whitening methods (Figure 4.4a). For a simple and non-rigorous visualization of the correlation profiles, we show the sliding-window-based correlations based on the raw signals and whitened signals (Figure 4.4b). We notice changes in the sliding-window-based correlation profile after applying whitening procedures to the raw signal. Nevertheless, these differences remain visually inconspicuous across the various whitening methods, as depicted in Figure 4.4b. The serial correlations unveil distinct patterns. Specifically, the raw signals were highly serially correlated, while the application of standard whitening methods such as AR(1) and ARMA(1,1) resulted in the persistence of relatively large serial correlations at lag 4 and lag 6 for node 1 (Figure 4.4c). Meanwhile, the IDAR-based whitening approaches successfully removed serial correlations in the selected signals.

Figure 4.5 provides a comprehensive comparison of the performance of various whitening algorithms. The proposed IDAR approach effectively removes serial correlations in nearly all

analyzed time series, irrespective of whether short-TR or long-TR datasets were analyzed, or whether resting-state fMRI or simulated task-fMRI signals were considered. In instances of long-TR datasets, where serial correlations diminish rapidly, the IDAR-iter1 strategy proficiently removes serial correlations. However, this method falls short in addressing the short-TR datasets, as a considerable 88% of the provided time series still display significant residual serial correlations. In contrast, conventional procedures based on AR(1) or ARMA(1,1) models proved ineffective in achieving optimal whitening, particularly in the context of short-TR datasets. Specifically, with the long-TR dataset, around 17.24% of time series did not attain adequate whitening, while with the short-TR dataset, none of the time series achieved satisfactory whitening (Figure 4.5a, 4.5b).

We subsequently investigated the power and type-I error concerning simulated task-fMRI data analysis. We observed a decrease in the statistical power of testing the associations between the task paradigm and the fMRI signal following the application of the whitening procedures (Figure 4.5c). Figure 4.5d presents the corresponding type-I error rates. In the absence of any whitening procedure, the type-I error could reach as high as 0.56. Whitening procedures lead to a substantial decrease in type-I error. Notably, the IDAR-iter1 method effectively maintained the type-I error at the nominal level for both long-TR and short-TR datasets, whereas the IDAR approach demonstrated an inflated type-I error of 0.11 within the short-TR dataset. The conventional AR(1) and ARMA(1,1) approaches exhibit inflated type-I error rates across both datasets.

4.3.2 Dynamic Functional Connectivity

We begin by visualizing the dynamic correlation profile of an example subject based on the raw and whitened fMRI data using the sliding window approach with a simple rectangular window function. For clarity, we focused on visualizing the correlations between specific pairs of brain nodes, namely PCC and SFG left, PCC and MPFC, PCC and PHC left, and PCC and pIPL left. As suggested in [129], the choice of a suitable minimum window length is influenced by the minimum frequency of the fMRI measurement. In our dataset, a window length of 100 seconds (equivalent to 50 data points) seemed appropriate. For comparison,

we also included window lengths of 50s, 200s, and 400s. As depicted in Figure 4.6, the observed correlations exhibited temporal variability. The variability could get substantial, resulting in correlations that not only change sign, but also exhibite a large amplitude that can be more than twice the magnitude of the static correlation measurement. However, the magnitude of this variation was highly dependent on the window length, making it challenging to distinguish between genuine underlying dynamics and spurious fluctuations. Comparing the correlation profiles estimated from raw and whitened data, we observed that the dynamic pattern was not significantly altered by the whitening procedure. However, there was a notable decrease in the correlation between PCC and MPFC for this subject, with the static correlation measurement decreasing from 0.51 to 0.33.

Figure 4.7 compares the sliding-window-based correlations with the correlation profiles based on the estimated DCC-GARCH model. Three example subjects whose correlation profiles vary are shown. The correlations calculated with the DCC-GARCH model show more noticeable differences among the subjects. Notably, the estimated correlation trajectories capture certain variation patterns identified by the sliding window approach. For instance, in subject 1, large drops in the correlations between PCC and SFG left were observed towards the end of the observation period.

By fitting separate DCC-GARCH models to each subject in the dataset, we observed significant group differences at 0.05 level in terms of the subject-level dynamic characterization parameter α between the healthy controls and the AD subjects (Figure 4.8a). The group difference in dynamic characterization parameter β was marginally significant at 0.1 level. The AD subjects in general had lower α values and higher β values, indicating less responsiveness to outside stimuli and more dependence on the previous brain's connectivity status in the DMN compared to the healthy subjects. In contrast, the stationary measurement based on standard dual-regression ICA analysis (Figure 4.8b) did not show significant group differences.

We then assessed the sensitivity and specificity of the estimated DCC-GARCH parameters in identifying AD pathology (Figure 4.9). To classify a subject as $A\beta+$, we employed varying threshold values for the estimated DCC-GARCH parameters α and/or β. With the

thresholds ranging from 0 to 1, we generated ROC curves and calculated their corresponding AUC. In the "combined" version, where both α and β values were used for classification, we employed the Support Vector Machine (SVM) approach with the radial kernel, where the tuning parameters (γ and regularization constant C) were selected to maximize AUC based on 10-fold cross-validation. When used separately for classification, the α and β parameters had moderate power to identify $A\beta+$ patients, each with an AUC of 0.6. We observed higher discriminative power for patient diagnosis when using both α and β values for classification, yielding an AUC of 0.7. On the contrary, the stationary functional connectivity measurement of average DMN strength computed from dual-regression ICA analysis shows no classification power for patient diagnosis.

4.4 Discussion

Our study contributes to understanding the dynamics of functional connectivity in AD and offers insights into the development of non-invasive and widely available biomarkers for the early detection of the disease. It sheds light on the importance of considering dynamic network configurations and the limitations of stationary measures of functional connectivity when searching for biomarkers of AD. To capture the temporal dynamics of brain network connectivity, we proposed using a DCC-GARCH model. The results of our study indicate that the proposed DCC-GARCH modeling approach for dynamic functional connectivity in the default mode network (DMN) is sensitive to the CSF $A\beta$ status. Specifically, the FC measurements of $A\beta+$ patients seem to exhibit higher lagging dependency and reduced responsiveness to new information compared to $A\beta-$ controls. The resulting DCC-GARCH model parameters show reasonable power in identifying $A\beta+$ patients in the population, indicating its potential as a sensitive biomarker for early detection of AD based on $A\beta$ deposition. We anticipate that with a larger number of subjects, the differences between the two groups would become more significant. Additionally, we expect that an increased sample size would lead to better predictability of our model. Future research should focus on validating our findings in larger cohorts of independent subjects.

It is important to note that the DCC-GARCH model provides a plausible description of the

temporal dynamics in the DMN, considering the data from the perspective of autoregressive dependency. While it may not represent the true or best model for describing the brain mechanism, it effectively captures important variation patterns identified by the sliding-window-based approach. The estimated dynamic functional correlation profiles using the DCC-GARCH model tend to be more stable and more concentrated around the static correlation measurement compared to the sliding-window-based approach, reflected in the smaller standard deviations over the time course. Further investigation should explore alternative perspectives, such as network topology or identifying change points in correlation profiles, to gain deeper insights into potential group differences. By examining multiple aspects, we can enhance our understanding of the dynamic functional connectivity alterations in AD and uncover more sensitive biomarkers for early detection.

The DCC-GARCH model comprises a mean component and a serially-uncorrelated residual component. In this study, we advocate for the application of the DCC-GARCH model with a constant mean to fMRI observations, where serial correlations have been removed by the whitening procedure. Alternatively, we may skip the whitening procedure and account for the serial correlations using ARMA models for the mean component, resulting in *ARMA-GARCH* models. Parameter estimation for ARMA-GARCH models typically involves maximizing the joint (quasi-)likelihood [21, 12, 138]. However, this approach encounters computational challenges, particularly with high order ARMA models [12]. Another strategy, the two-step estimation, involves an initial ARMA parameter estimation, followed by fitting a GARCH model to the residuals in the second stage [12, 171]. While straightforward in implementation, there is a lack of comprehensive investigations into the properties of the resulting estimates. Notably, theoretical discussions of ARMA-GARCH models have predominantly focused on univariate observations. The only multivariate version with theoretical analysis is the VARMA-GARCH model, which assumes a constant correlation matrix and is not of interest for our application [138]. Despite some applications of the ARMA-DCC-GARCH model [254, 31, 200], there exists a research gap in the theoretical properties of ARMA-DCC-GARCH models. This represents a promising avenue for future exploration. In this context, we explored an IDAR-DCC-GARCH model using a two-step approach: the mean component

was estimated using the IDAR model where we did not de-correlate the observations using estimated correlation matrices, but calculated the residuals from the AR models; then we fit a DCC-GARCH model to the IDAR model residuals. We obtained consistent results with those from the proposed DCC-GARCH approach with pre-whitened observations. The examination of connections and distinctions between the (ID)AR-DCC-GARCH approach and the proposed method involving pre-whitening observations, presents an intriguing direction for future research.

Our study provides an effective whitening approach as well as a universally adaptable method capable of accommodating the diverse characteristics of fMRI data sets. An effective whitening procedure that eliminates serial correlations is necessary to ensure the validity of the DCC-GARCH model. This necessity extends beyond the confines of the DCC-GARCH model; a proper whitening method is pivotal in various applications, including task-fMRI analysis, and becomes even more critical with the widespread use of short-TR data sets. Conventional whitening techniques prove to be inadequate for fMRI data, especially for data collected with short TR. Furthermore, as highlighted in Section 4.3.1, the utilization of various whitening methods can introduce substantial disparities in the resulting signals, consequently impacting subsequent analyses and conclusions. This underscores the issue of replicability in the field of fMRI research and the need for a unified whitening approach that accommodates the diverse characteristics inherent in fMRI data sets. In response to the challenge of whitening fMRI signals, we introduced a novel iterative data-driven whitening procedure, termed IDAR. This approach offers versatility by effectively addressing either serial correlations and/or the type-I error rate in task-fMRI analysis. IDAR is grounded in AR models and stands out for its iterative nature, designed to capture the intricate serial correlation structures in fMRI datasets.

No existing whitening method satisfactorily addresses both residual serial correlations and type-I error in task analysis simultaneously. For datasets with longer TR, standard approaches employing low-order AR models have been extensively studied. However, these models, such as AR(1) and ARMA(1,1), often lack the requisite order to adequately capture serial correlation structures, leading to incomplete removal of serial correlations. These

standard models encounter further challenges in the context of short-TR datasets due to the presence of slow-decaying serial correlations, as our results in Section 4.3.1 confirm. Although prior literature has suggested high order AR models as effective for sub-second TR datasets [145, 186], our findings in Section 4.3.1 reveal that even the flexible high order AR model equivalent (IDAR-iter1) struggles with short-TR datasets. Such unaddressed serial correlations can introduce biases in estimation and distort variance computations [3].

We also investigated the performance of the whitening procedures in the context of task-fMRI analysis using simulated task-fMRI data. The standard low-order models showed inflated type-I error rates that lead to false discovery, potentially impacting clinical applications like brain tumor surgery or epilepsy treatment [166]. Interestingly, while the high order AR model (IDAR-iter1) could not fully mitigate serial correlations, it demonstrated a capacity to control type-I error at the nominal level for both short-TR and long-TR datasets. The IDAR approach with multiple iterations, however, fell short of fully controlling type-I error for short-TR datasets. Thus, complexities arise when comparing the performance of whitening procedures across different evaluation metrics. Notably, the evaluation of whitening procedures lacks a consensus on proper metrics, and our chosen criteria are not exhaustive. Our analysis reveals instances where methods excel in one metric but fall short in another. The relationship between these metrics remains unclear, underscoring the need for exploring appropriate evaluation measures for whitening techniques. Such endeavors could significantly contribute to advancing the field.

Our proposed IDAR approach demonstrates efficacy in tackling challenges associated with long-TR datasets, while also exhibiting enhanced performance compared with currently available methods for short-TR datasets. In the context of long-TR data, the adaptive order-selection mechanism inherent in IDAR encompasses low-order models automatically. Notably, IDAR-iter1 is often sufficient for long-TR signals, and this aligns with the prevailing understanding that long-TR datasets typically exhibit fast-decaying serial correlations, making moderate-order time series models usually sufficient. Conversely, for short-TR datasets, the IDAR approach offers a more comprehensive solution than existing methods. This is attributed to its incorporation of high order AR models and its capacity to address

intricate serial correlation structures through iterative procedures.

In the comparison of various whitening procedures, our proposed IDAR and IDAR-iter1 approaches leverage more comprehensive models and exhibit superior control over type-I error rates. However, it is worth noting that compared to simpler models, they tend to yield lower statistical power when testing for associations in task-fMRI analysis. The optimization of statistical power while maintaining stringent control over type-I error rates represents a promising avenue for future research in the field.

Although neither IDAR nor IDAR-iter1 simultaneously addresses remaining serial correlations and inflated type-I error, these approaches provide partial solutions. Researchers can utilize multiple iterations with IDAR to address residual correlations in resting-state fMRI data, or can opt for IDAR-iter1 to mitigate inflated type-I error rates in task-fMRI analysis. Nonetheless, it remains a challenge to develop methods that effectively meet both evaluation criteria for fully whitening short-TR datasets. The intricacies of short-TR data introduce complexities that necessitate further exploration and innovative solutions to achieve the dual goals of reduced serial correlations and controlled type-I error rates.

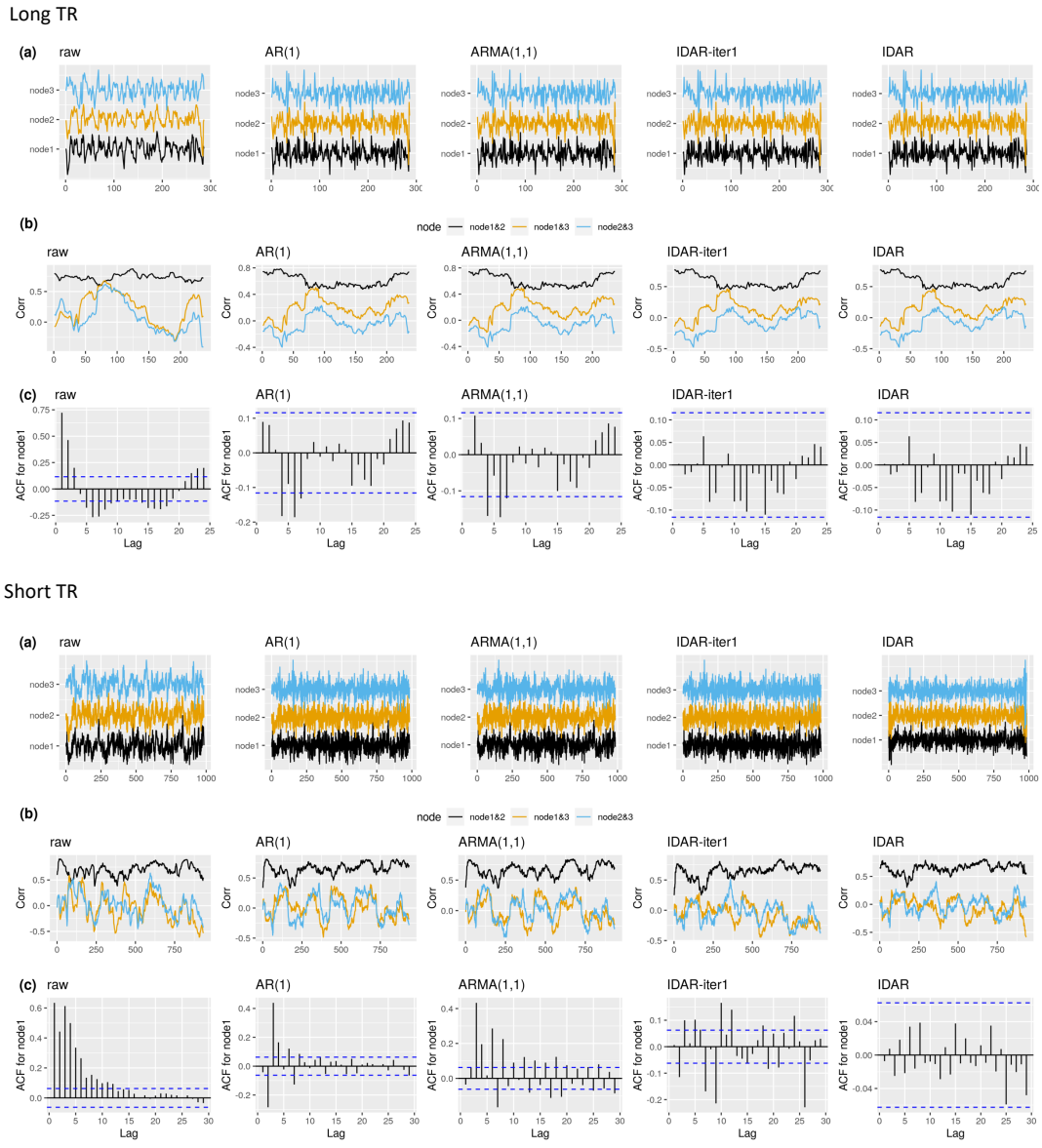

Figure 4.4: The resting-state fMRI signal with different whitening procedures for two example subjects. Top panel shows an example subject from long-TR dataset; bottom panel shows an example subject from short-TR dataset. We select to present the signals from node 1 – dPFC left, node 2 – dPFC right, node 3 – mPFC. We show the raw resting-state fMRI, as well as the whitened signals after applying the IDAR, IDAR-iter1, AR(1) and ARMA(1,1) based whitening procedures. **(a)**: The raw and whitened resting-state fMRI observations. **(b)**: The sliding-window-based correlation profile, computed from raw and whitened observations. A window length of 100s is selected for illustration purpose. **(c)**: The serial correlations at different lags for observations from node 1. The dashed lines represent the threshold, beyond which the serial correlation is considered significantly different from zero at 0.05 significance level.

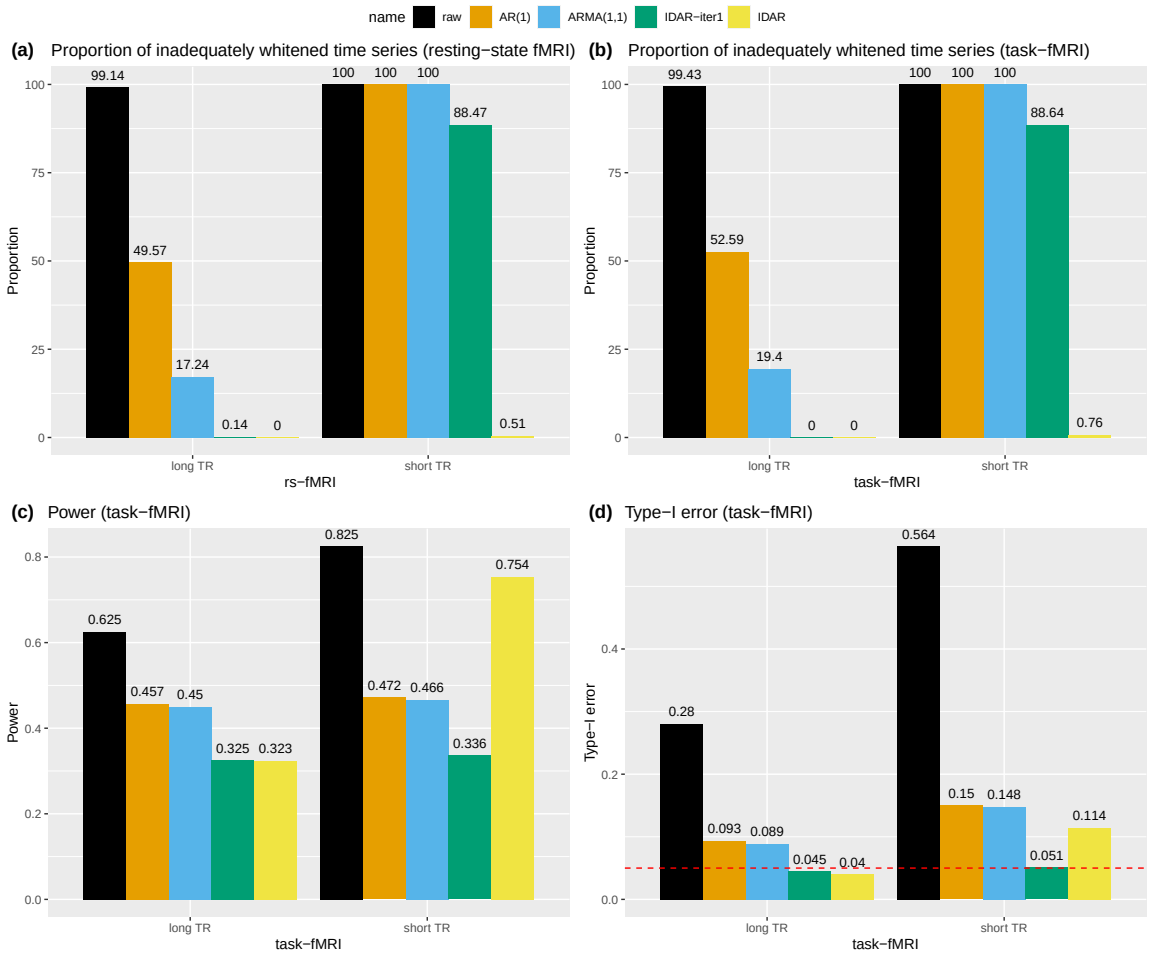

Figure 4.5: Bar charts for comparing the performance of different whitening approaches in long-TR and short-TR datasets. **(a)**, **(b)**: Proportion of time series that are not adequately whitened for resting-state fMRI and task-fMRI. **(c)**: Power for detecting the association between signals and task paradigm in task-fMRI analysis. **(d)**: Type-I error for testing the association between signals and task paradigm in task-fMRI analysis. The red dashed line represents the 0.05 nominal level.

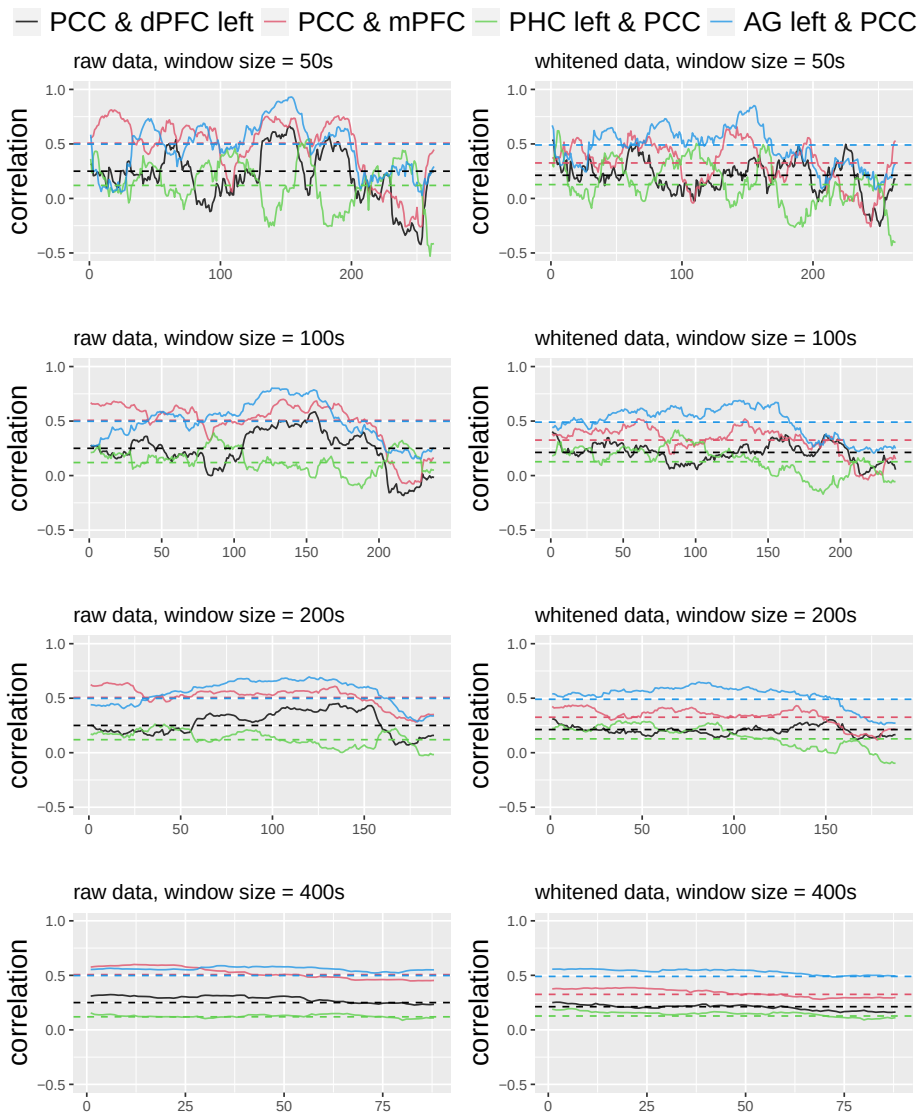

Figure 4.6: The dynamic correlation profile obtained via the sliding window approach, using the raw data and the whitened data, for an example subject. We visualize the correlation profile of four chosen brain node pairs. Correlations are computed using the rectangular window function at window length = 50s, 100s, 200s, 400s. The dashed lines represent the standard static correlation coefficient for each node pair, computed using data from the full time course.

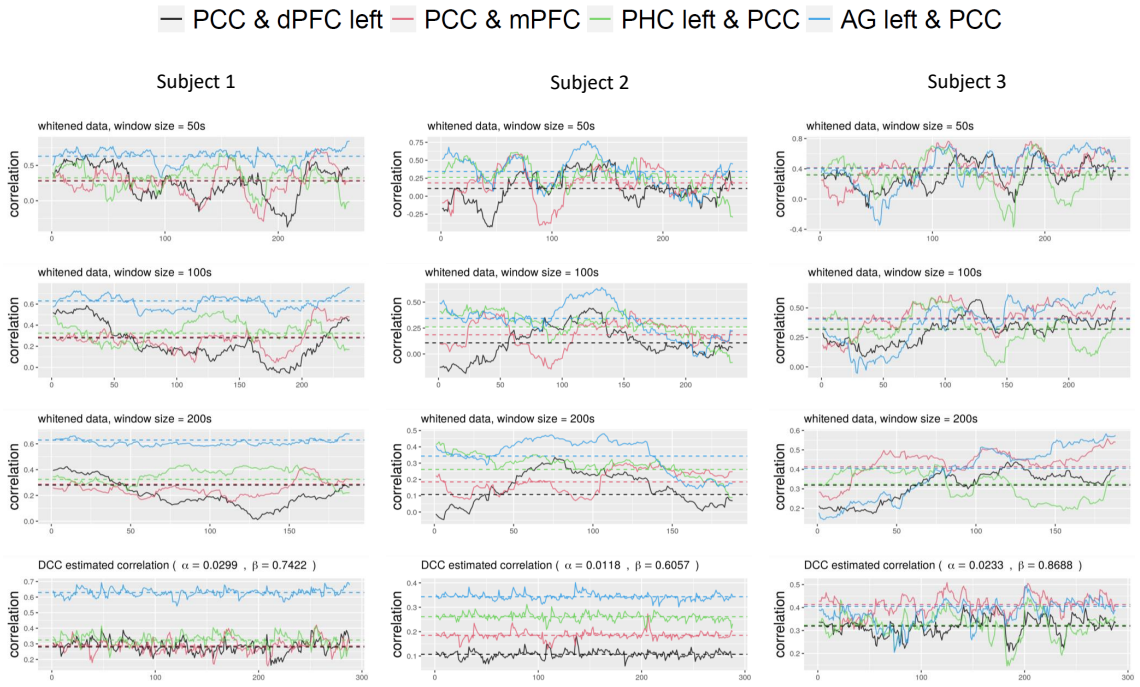

Figure 4.7: The dynamic correlation profile of three example subjects, using the whitened data, based on the sliding window approaches (top three rows) and the DCC-GARCH approach (last row). The sliding window approach applies the rectangular window function at window length = 50s, 100s, 200s. We visualize the correlation profile of four chosen brain node pairs. The dashed lines represent the standard static correlation coefficient for each node pair, computed using data from the full time course.

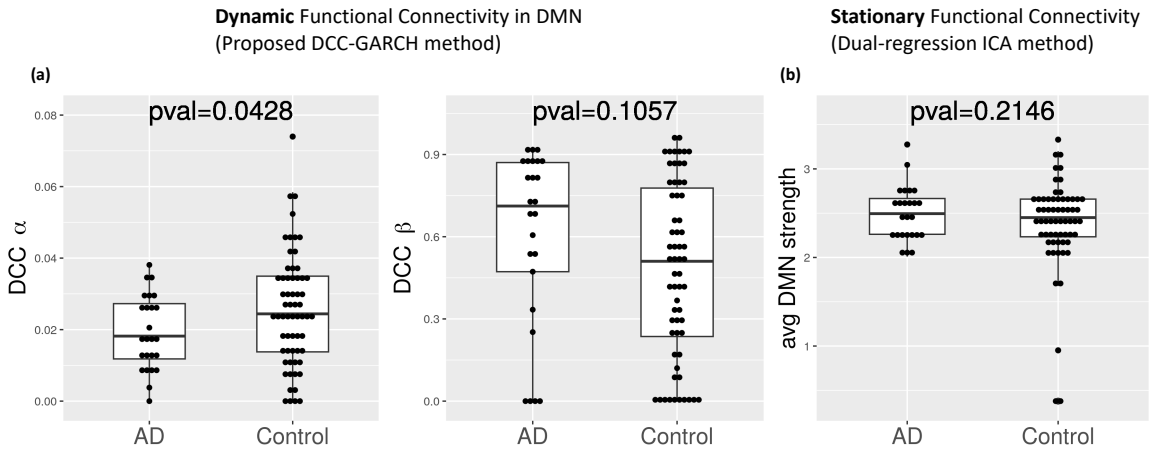

Figure 4.8: Group difference between AD patients and healthy controls in estimated DCC-GARCH parameters, and in dual-regression yielded functional connectivity measurements. (a) Boxplots of estimated DCC-GARCH α and β parameters calculated using the proposed method for all subjects. Each dot represents the parameter estimate for a single subject. Subjects were divided into two groups based on their CSF $A\beta$ profile: 1) AD patients (CSF $A\beta 42/\ A\beta 40 < 0.11$), and 2) control (CSF $A\beta 42/\ A\beta 40 \geq 0.11$). p-values of two-sample t-test for group differences in parameter estimates are reported. (b) Boxplots of average stationary functional connectivity in DMN, calculated using dual-regression ICA analysis.

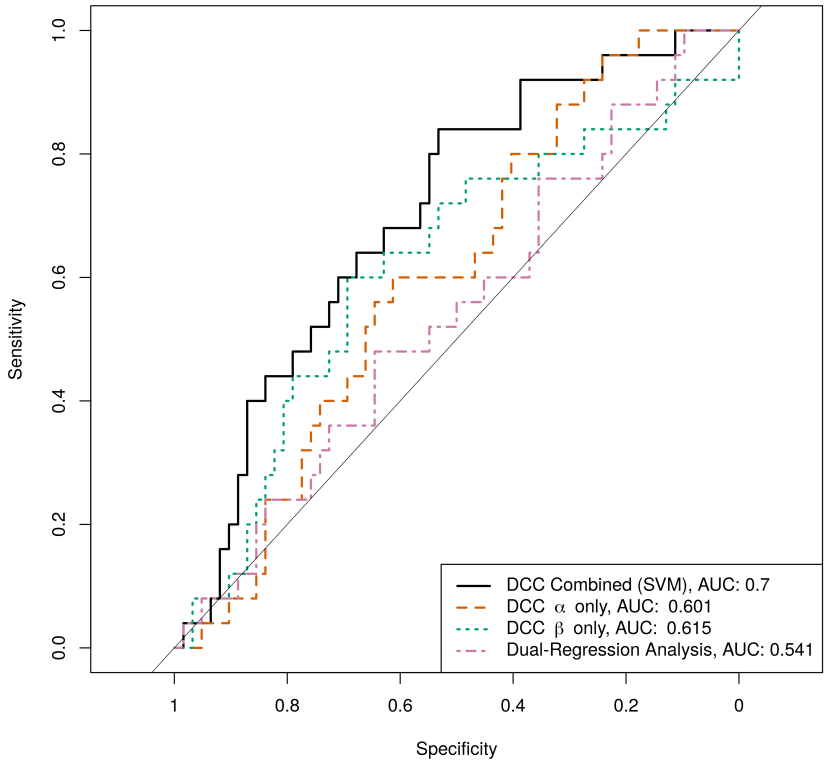

Figure 4.9: ROC curves and AUC for identifying AD patients based on estimated DCC-GARCH parameters. The 'DCC α only' and 'DCC β only' curves are based on thresholding the estimated α or β values separately, the 'DCC Combined' ROC curve is yielded from thresholding both α and β values. The 'Dual-Regression Analysis' ROC curve is based on average DMN strength computed from dual-regression ICA analysis.

Chapter 5

CONCLUSION

In this dissertation, we have focused on addressing the challenges associated with analyzing correlated data, drawing inspiration from diverse examples in genetic and neuroscience studies. Our primary focus has been on advancing statistical methodologies tailored for large-scale linear mixed models and network analysis.

In Chapter 2, we developed a computationally efficient and robust estimator for variance components in large-scale genetic studies. While the proposed REHE and reREHE estimates demonstrated efficacy, the computational expense of constructing the corresponding confidence intervals remains a hurdle, particularly with non-sparse correlation matrices. The path forward involves enhancing the computational efficiency of this procedure. Moreover, exploring the extension of the REHE approach to the estimation of covariance matrices rather than scalar variance components opens up intriguing possibilities for future research.

In Chapter 3, we were motivated by exploring brain connectivity structure within a multi-subject experiment. To address the inherent heterogeneity in this setting, we took a neighborhood selection approach with doubly high-dimensional linear mixed models, proposing a novel estimation and inference framework specifically designed for network inference applications. Our theoretical analysis concentrated on scenarios where $q/m > 1$, guiding our proof techniques. Future research could explore alternative proof techniques for situations where $q/m < 1$, allowing for the relaxation of assumptions and potentially improved convergence rates. Furthermore, we showcased the versatility of our framework by demonstrating its applicability to time series data, including extensions to linear mixed models with correlated noise components and to mixed-effect VAR model of order 1 (MEVAR(1)). Subsequent investigations could delve into extending this framework beyond the MEVAR(1) model to

higher-order MEVAR models, potentially requiring distinct proof techniques. In Chapter 3, our approach to inferring the network structure adopted a conditional modeling perspective, specifying the conditional distribution at node j as a function of other nodes' variables. However, characterizing the corresponding joint distribution of all nodes appears to be nontrivial. Future research could focus on exploring these aspects in greater depth.

The focus of Chapter 4 was on understanding dynamic brain functional connectivity alterations in AD using the DCC-GARCH model. Our observations reveal significant group disparities in DCC-GARCH parameters between healthy subjects and those with AD, and showcase the model's superior predictive ability for $A\beta$ status. Future investigations should prioritize validating these findings in larger cohorts, which may lead to heightened significance in group differences and enhance model predictability. Despite its merits, the DCC-GARCH model provides just one plausible depiction of the dynamics in functional connectivity profiles. To deepen insights into AD-induced brain connectivity alterations and potentially identify more sensitive biomarkers for early detection, future research should explore alternative perspectives. This could involve examining network topology or pinpointing change points in correlation profiles. In addition to the DCC-GARCH model, we introduced the IDAR whitening procedure to ensure proper fitting of the DCC-GARCH model. Instead of whitening the observations, an alternative strategy involves employing GARCH models with a time-series mean component to capture serial correlations in the observations. Nevertheless, there is a lack of GARCH models for describing dynamic correlation profiles in multivariate serially correlated observations, presenting an intriguing avenue for future research. As for the IDAR whitening approach, it provides a partial solution to simultaneously addressing both residual serial correlations and type-I errors in task-fMRI analysis. Moreover, the evaluation of whitening procedures lacks consensus on proper metrics. Subsequent work should focus on exploring suitable and comprehensive evaluation approaches for whitening techniques, unraveling the relationship among different metrics. The field could benefit significantly from efforts dedicated to refining the whitening procedure, aiming for universal approaches that excel across diverse evaluation metrics.

In conclusion, this dissertation lays the groundwork for advanced statistical methods in

handling correlated data within the realms of genetic studies and neuroscience. The identified areas for future research pave the way for a more comprehensive understanding of these complex systems.

BIBLIOGRAPHY

[1] Soroosh Afyouni, Stephen M Smith, and Thomas E Nichols. Effective degrees of freedom of the pearson's correlation coefficient under autocorrelation. *NeuroImage*, 199:609–625, 2019.

[2] Elena A Allen, Eswar Damaraju, Sergey M Plis, Erik B Erhardt, Tom Eichele, and Vince D Calhoun. Tracking whole-brain connectivity dynamics in the resting state. *Cerebral cortex*, 24(3):663–676, 2014.

[3] Mohammad R Arbabshirani, Eswar Damaraju, Ronald Phlypo, Sergey Plis, Elena Allen, Sai Ma, Daniel Mathalon, Adrian Preda, Jatin G Vaidya, Tülay Adali, et al. Impact of autocorrelation on functional connectivity. *Neuroimage*, 102:294–308, 2014.

[4] Tomoki Arichi, Gianlorenzo Fagiolo, Marta Varela, Alejandro Melendez-Calderon, Alessandro Allievi, Nazakat Merchant, Nora Tusor, Serena J Counsell, Etienne Burdet, Christian F Beckmann, et al. Development of bold signal hemodynamic responses in the human brain. *Neuroimage*, 63(2):663–673, 2012.

[5] Yurii S Aulchenko, Dirk-Jan De Koning, and Chris Haley. Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics*, 177(1):577–585, 2007.

[6] Inês Baldeiras, Isabel Santana, Maria João Leitão, Helena Gens, Rui Pascoal, Miguel Tábuas-Pereira, José Beato-Coelho, Diana Duro, Maria Rosário Almeida, and Catarina Resende Oliveira. Addition of the aβ42/40 ratio to the cerebrospinal fluid biomarker profile increases the predictive value for underlying alzheimer's disease dementia in mild cognitive impairment. *Alzheimer's research & therapy*, 10(1):1–15, 2018.

[7] Inês Baldeiras, Isabel Santana, Maria João Leitão, Maria Helena Ribeiro, Rui Pascoal, Diana Duro, Raquel Lemos, Beatriz Santiago, Maria Rosário Almeida, and Catarina Resende Oliveira. Cerebrospinal fluid aβ40 is similarly reduced in patients with frontotemporal lobar degeneration and alzheimer's disease. *Journal of the neurological sciences*, 358(1-2):308–316, 2015.

[8] Marcio Luiz Figueredo Balthazar, Brunno Machado de Campos, Alexandre Rosa Franco, Benito Pereira Damasceno, and Fernando Cendes. Whole cortical and default mode network mean functional connectivity as potential biomarkers for mild alzheimer's disease. *Psychiatry Research: Neuroimaging*, 221(1):37–42, 2014.

[9] Nicolas R Barthélemy, Kanta Horie, Chihiro Sato, and Randall J Bateman. Blood plasma phosphorylated-tau isoforms track cns change in alzheimer's disease. *Journal of Experimental Medicine*, 217(11):e20200861, 2020.

[10] Danielle S Bassett and Edward T Bullmore. Human brain networks in health and disease. *Current opinion in neurology*, 22(4):340, 2009.

[11] Sumanta Basu and George Michailidis. Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43(4):1535–1567, 2015.

[12] Luc Bauwens, Sébastien Laurent, and Jeroen VK Rombouts. Multivariate garch models: a survey. *Journal of applied econometrics*, 21(1):79–109, 2006.

[13] Christian F Beckmann and Stephen M Smith. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 23(2):137–152, 2004.

[14] Andreas Weingessel Berwin A. Turlach. *quadprog: Functions to Solve Quadratic Programming Problems*, 2019. R package version 1.5-7.

[15] Rajendra Bhatia. *Perturbation bounds for matrix eigenvalues*. SIAM, 2007.

[16] Bharat Biswal, F Zerrin Yetkin, Victor M Haughton, and James S Hyde. Functional

connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, 34(4):537–541, 1995.

[17] Martin Blaimer, Felix A Breuer, Nicole Seiberlich, Matthias F Mueller, Robin M Heidemann, Vladimir Jellus, Graham Wiggins, Lawrence L Wald, Mark A Griswold, and Peter M Jakob. Accelerated volumetric mri with a sense/grappa combination. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 24(2):444–450, 2006.

[18] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[19] Tim Bollerslev. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized arch model. *The review of economics and statistics*, pages 498–505, 1990.

[20] Tim Bollerslev, Robert F Engle, and Jeffrey M Wooldridge. A capital asset pricing model with time-varying covariances. *Journal of political Economy*, 96(1):116–131, 1988.

[21] Tim Bollerslev and Jeffrey M Wooldridge. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews*, 11(2):143–172, 1992.

[22] Saskia Bollmann, Alexander M Puckett, Ross Cunnington, and Markus Barth. Serial correlations in single-subject fmri with sub-second tr. *NeuroImage*, 166:152–166, 2018.

[23] Jelena Bradic, Gerda Claeskens, and Thomas Gueuning. Fixed effects testing in high-dimensional linear mixed models. *Journal of the American Statistical Association*, 115(532):1835–1850, 2020.

[24] Guy Bresler. Efficiently learning ising models on arbitrary graphs. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 771–782, 2015.

[25] Molly G Bright, Christopher R Tench, and Kevin Murphy. Potential pitfalls when denoising resting state fmri data using nuisance regression. *Neuroimage*, 154:159–168, 2017.

[26] Laura F Bringmann, Nathalie Vissers, Marieke Wichers, Nicole Geschwind, Peter Kuppens, Frenk Peeters, Denny Borsboom, and Francis Tuerlinckx. A network approach to psychopathology: new insights into clinical longitudinal data. *PloS one*, 8(4):e60188, 2013.

[27] Annette Brose, Florian Schmiedek, Peter Koval, and Peter Kuppens. Emotional inertia contributes to depressive symptoms beyond perseverative thinking. *Cognition and Emotion*, 29(3):527–538, 2015.

[28] Randy L Buckner, Abraham Z Snyder, Benjamin J Shannon, Gina LaRossa, Rimmon Sachs, Anthony F Fotenos, Yvette I Sheline, William E Klunk, Chester A Mathis, John C Morris, et al. Molecular, structural, and functional characterization of alzheimer's disease: evidence for a relationship between default activity, amyloid, and memory. *Journal of neuroscience*, 25(34):7709–7717, 2005.

[29] Peter Bühlmann and Sara Van De Geer. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.

[30] Richard B Buxton, Kâmil Uludağ, David J Dubowitz, and Thomas T Liu. Modeling the hemodynamic response to brain activation. *Neuroimage*, 23:S220–S233, 2004.

[31] Göknur Büyükkara, Onur Enginar, and Hüseyin Temiz. Volatility spillovers between oil and stock market returns in g7 countries: A var-dcc-garch model. *Regulations in the Energy Industry: Financial, Economic and Legal Implications*, pages 169–186, 2020.

[32] Tony Cai, Weidong Liu, and Xi Luo. A constrained $\ell 1$ minimization approach to sparse precision matrix estimation. *Journal of the American Statistical Association*, 106(494):594–607, 2011.

[33] Lorenzo Cappiello, Robert F Engle, and Kevin Sheppard. Asymmetric dynamics in

the correlations of global equity and bond returns. *Journal of Financial econometrics*, 4(4):537–572, 2006.

[34] Catie Chang and Gary H Glover. Time–frequency dynamics of resting-state brain connectivity measured with fmri. *Neuroimage*, 50(1):81–98, 2010.

[35] Gang Chen, Daniel R Glen, Ziad S Saad, J Paul Hamilton, Moriah E Thomason, Ian H Gotlib, and Robert W Cox. Vector autoregression, structural equation modeling, and their synthesis in neuroimaging data analysis. *Computers in biology and medicine*, 41(12):1142–1155, 2011.

[36] Gang Chen, Ziad S Saad, Audrey R Nath, Michael S Beauchamp, and Robert W Cox. Fmri group analysis combining effect estimates and their variances. *Neuroimage*, 60(1):747–765, 2012.

[37] Shizhe Chen, Ali Shojaie, and Daniela M Witten. Network reconstruction from high-dimensional ordinary differential equations. *Journal of the American Statistical Association*, 112(520):1697–1707, 2017.

[38] Shizhe Chen, Daniela M Witten, and Ali Shojaie. Selection and estimation for mixed graphical models. *Biometrika*, 102(1):47–64, 2015.

[39] Sharon Chiang, Michele Guindani, Hsiang J Yeh, Zulfi Haneef, John M Stern, and Marina Vannucci. Bayesian vector autoregressive model for multi-subject effective connectivity inference using multi-modal neuroimaging data. *Human brain mapping*, 38(3):1311–1332, 2017.

[40] Matthew P. Conomos, Stephanie M. Gogarten, Lisa Brown, Han Chen, Ken Rice, Tamar Sofer, Timothy Thornton, and Chaoyu Yu. *GENESIS: GENetic EStimation and Inference in Structured samples (GENESIS): Statistical methods for analyzing genetic data from samples with population structure and/or relatedness*, 2019. R package version 2.14.3.

[41] Matthew P Conomos, Cecelia A Laurie, Adrienne M Stilp, Stephanie M Gogarten, Caitlin P McHugh, Sarah C Nelson, Tamar Sofer, Lindsay Fernández-Rhodes,

Anne E Justice, Mariaelisa Graff, et al. Genetic diversity and association studies in US Hispanic/Latino populations: applications in the Hispanic Community Health Study/Study of Latinos. *The American Journal of Human Genetics*, 98(1):165–184, 2016.

[42] Nadège Corbin, Nick Todd, Karl J Friston, and Martina F Callaghan. Accurate modeling of temporal correlations in rapidly sampled fmri time series. *Human brain mapping*, 39(10):3884–3897, 2018.

[43] Robert W Cox. Afni: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, 29(3):162–173, 1996.

[44] Mandeep S Dagli, John E Ingeholm, and James V Haxby. Localization of cardiac-induced signal change in fmri. *Neuroimage*, 9(4):407–415, 1999.

[45] Jessica S Damoiseaux, Katherine E Prater, Bruce L Miller, and Michael D Greicius. Functional connectivity tracks clinical deterioration in alzheimer's disease. *Neurobiology of aging*, 33(4):828–e19, 2012.

[46] Aaron Defazio and Tibério Caetano. A convex formulation for learning scale-free networks via submodular relaxation. *Advances in neural information processing systems*, 25, 2012.

[47] Gopikrishna Deshpande, Stephan LaConte, George Andrew James, Scott Peltier, and Xiaoping Hu. Multivariate granger causality analysis of fmri data. *Human brain mapping*, 30(4):1361–1373, 2009.

[48] Gopikrishna Deshpande, Priya Santhanam, and Xiaoping Hu. Instantaneous and causal connectivity in resting state brain networks derived from functional mri data. *Neuroimage*, 54(2):1043–1052, 2011.

[49] Adrien E Desjardins, Kent A Kiehl, and Peter F Liddle. Removal of confounding effects of global signal in functional mri analyses. *Neuroimage*, 13(4):751–758, 2001.

[50] Ruben Dezeure, Peter Bühlmann, Lukas Meier, and Nicolai Meinshausen. High-dimensional inference: Confidence intervals, p-values and R-software hdi. *Statistical Science*, 30(4):533–558, 2015.

[51] Ruben Dezeure, Peter Bühlmann, Lukas Meier, and Nicolai Meinshausen. High-dimensional inference: confidence intervals, p-values and r-software hdi. *Statistical science*, pages 533–558, 2015.

[52] Mathias Drton and Marloes H Maathuis. Structure learning in graphical modeling. *arXiv preprint arXiv:1606.02359*, 2016.

[53] Julien Dubois and Ralph Adolphs. Building a science of individual differences from fmri. *Trends in Cognitive Sciences*, 20(6):425–443, 2016.

[54] Frank Dudbridge and Arief Gusnanto. Estimation of significance thresholds for genome wide association scans. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, 32(3):227–234, 2008.

[55] Martin Dyrba, Reza Mohammadi, Michel J Grothe, Thomas Kirste, and Stefan J Teipel. Gaussian graphical models reveal inter-modal and inter-regional conditional dependencies of brain alterations in alzheimer's disease. *Frontiers in aging neuroscience*, 12:99, 2020.

[56] Morris L Eaton. Multivariate statistics : a vector space approach, 2007.

[57] Anders Eklund, Thomas E Nichols, and Hans Knutsson. Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates. *Proceedings of the national academy of sciences*, 113(28):7900–7905, 2016.

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

[59] Robert F Engle and Kenneth F Kroner. Multivariate simultaneous generalized arch. *Econometric theory*, 11(1):122–150, 1995.

[60] Robert F Engle III and Kevin Sheppard. Theoretical and empirical properties of dynamic conditional correlation multivariate garch, 2001.

[61] Yingying Fan and Runze Li. Variable selection in linear mixed effects models. *Annals of Statistics*, 40(4):2043, 2012.

[62] Mogens Fenger. Heritability and genetics of lipid metabolism. *Future Lipidology*, 2(4):433–444, 2007.

[63] Orestes V Forlenza, Marcia Radanovic, Leda L Talib, Ivan Aprahamian, Breno S Diniz, Henrik Zetterberg, and Wagner F Gattaz. Cerebrospinal fluid biomarkers in alzheimer's disease: diagnostic accuracy and prediction of dementia. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 1(4):455–463, 2015.

[64] Vojtěch Franc, Václav Hlaváč, and Mirko Navara. Sequential coordinate-wise algorithm for the non-negative least squares problem. In *International Conference on Computer Analysis of Images and Patterns*, pages 407–414. Springer, 2005.

[65] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[66] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010.

[67] Karl J Friston. Functional and effective connectivity: a review. *Brain connectivity*, 1(1):13–36, 2011.

[68] Karl J Friston, Daniel E Glaser, Richard NA Henson, S Kiebel, Christophe Phillips, and John Ashburner. Classical and bayesian inference in neuroimaging: applications. *Neuroimage*, 16(2):484–512, 2002.

[69] Karl J Friston, Andrea Mechelli, Robert Turner, and Cathy J Price. Nonlinear responses in fmri: the balloon model, volterra kernels, and other hemodynamics. *NeuroImage*, 12(4):466–477, 2000.

[70] Alexios Galanos. *rmgarch: Multivariate GARCH models.*, 2022. R package version 1.3-9.

[71] Abhik Ghosh and Magne Thoresen. Non-concave penalization in linear mixed-effect models and regularized selection of fixed effects. *AStA Advances in Statistical Analysis*, 102(2):179–210, 2018.

[72] Arthur R Gilmour, Robin Thompson, and Brian R Cullis. Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics*, 55:1440–1450, 1995.

[73] Matthew F Glasser, Stamatios N Sotiropoulos, J Anthony Wilson, Timothy S Coalson, Bruce Fischl, Jesper L Andersson, Junqian Xu, Saad Jbabdi, Matthew Webster, Jonathan R Polimeni, et al. The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, 80:105–124, 2013.

[74] Rainer Goebel, Alard Roebroeck, Dae-Shik Kim, and Elia Formisano. Investigating directed cortical interactions in time-resolved fmri data using vector autoregressive modeling and granger causality mapping. *Magnetic resonance imaging*, 21(10):1251–1261, 2003.

[75] Stephanie M Gogarten, Tamar Sofer, Han Chen, Chaoyu Yu, Jennifer A Brody, Timothy A Thornton, Kenneth M Rice, and Matthew P Conomos. Genetic association testing using the genesis r/bioconductor package. *Bioinformatics*, 35(24):5346–5348, 2019.

[76] Donald Goldfarb and Ashok Idnani. A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming*, 27(1):1–33, 1983.

[77] Cristina Gorrostieta, Mark Fiecas, Hernando Ombao, Erin Burke, and Steven Cramer. Hierarchical vector auto-regressive models and their applications to multi-subject effective connectivity. *Frontiers in computational neuroscience*, 7:159, 2013.

[78] Cristina Gorrostieta, Hernando Ombao, Patrick Bédard, and Jerome N Sanes. Investigating brain connectivity using mixed effects vector autoregressive models. *NeuroImage*, 59(4):3347–3355, 2012.

[79] Kamil A Grajski, Steven L Bressler, Alzheimer's Disease Neuroimaging Initiative, et al. Differential medial temporal lobe and default-mode network functional connectivity and morphometric changes in alzheimer's disease. *NeuroImage: Clinical*, 23:101860, 2019.

[80] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.

[81] Franklin A Graybill. On quadratic estimates of variance components. *The Annals of Mathematical Statistics*, 25(2):367–372, 1954.

[82] Franklin A Graybill and Robert A Hultquist. Theorems concerning eisenhart's model ii. *The Annals of Mathematical Statistics*, 32(1):261–269, 1961.

[83] Franklin A Graybill and AW Wortham. A note on uniformly best unbiased estimators for variance components. *Journal of the American Statistical Association*, 51(274):266–268, 1956.

[84] Michael D Greicius, Ben Krasnow, Allan L Reiss, and Vinod Menon. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the national academy of sciences*, 100(1):253–258, 2003.

[85] Michael D Greicius, Gaurav Srivastava, Allan L Reiss, and Vinod Menon. Default-mode network activity distinguishes alzheimer's disease from healthy aging: evidence from functional mri. *Proceedings of the National Academy of Sciences*, 101(13):4637–4642, 2004.

[86] Ludovica Griffanti, Gholamreza Salimi-Khorshidi, Christian F Beckmann, Edward J Auerbach, Gwenaëlle Douaud, Claire E Sexton, Enikő Zsoldos, Klaus P Ebmeier,

Nicola Filippini, Clare E Mackay, et al. Ica-based artefact removal and accelerated fmri acquisition for improved resting state network imaging. *Neuroimage*, 95:232–247, 2014.

[87] Fang Han, Huanran Lu, and Han Liu. A direct estimation of high dimensional stationary vector autoregressions. *Journal of Machine Learning Research*, 2015.

[88] Daniel A Handwerker, John M Ollinger, and Mark D'Esposito. Variation of bold hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage*, 21(4):1639–1651, 2004.

[89] Daniel A Handwerker, Vinai Roopchansingh, Javier Gonzalez-Castillo, and Peter A Bandettini. Periodic changes in fmri connectivity. *Neuroimage*, 63(3):1712–1719, 2012.

[90] Erika Oliveira Hansen, Natalia Silva Dias, Ivonne Carolina Bolaños Burgos, Monica Vieira Costa, Andréa Teixeira Carvalho, Antonio Lucio Teixeira, Izabela Guimarães Barbosa, Lorena Aline Valu Santos, Daniela Valadão Freitas Rosa, Aloisio Joaquim Freitas Ribeiro, et al. Millipore xmap® luminex (hatmag-68k): An accurate and cost-effective method for evaluating alzheimer's biomarkers in cerebrospinal fluid. *Frontiers in Psychiatry*, 12:716686, 2021.

[91] JK Haseman and RC Elston. The investigation of linkage between a quantitative trait and a marker locus. *Behavior Genetics*, 2(1):3–19, 1972.

[92] Alain Hecq, Luca Margaritella, and Stephan Smeekes. Inference in non-stationary high-dimensional vars. *arXiv preprint arXiv:2302.01434*, 2023.

[93] Trey Hedden, Koene RA Van Dijk, J Alex Becker, Angel Mehta, Reisa A Sperling, Keith A Johnson, and Randy L Buckner. Disruption of functional connectivity in clinically normal older adults harboring amyloid burden. *Journal of Neuroscience*, 29(40):12686–12694, 2009.

[94] Stephen J Heishman, Kamyar Arasteh, and Maxine L Stitzer. Comparative effects of alcohol and marijuana on mood, memory, and performance. *Pharmacology Biochemistry and Behavior*, 58(1):93–101, 1997.

[95] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70, 1979.

[96] Hamed Honari, Ann S Choe, James J Pekar, and Martin A Lindquist. Investigating the impact of autocorrelation on time-varying connectivity. *NeuroImage*, 197:37–48, 2019.

[97] Jean Honorio and Tommi Jaakkola. Tight bounds for the expected risk of linear classifiers and pac-bayes finite-sample guarantees. In *Artificial Intelligence and Statistics*, pages 384–392. PMLR, 2014.

[98] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[99] Yao Hu, Adrienne M Stilp, Caitlin P McHugh, Shuquan Rao, Deepti Jain, Xiuwen Zheng, John Lane, Sébastian Méric de Bellefon, Laura M Raffield, Ming-Huei Chen, et al. Whole-genome sequencing association analysis of quantitative red blood cell phenotypes: The nhlbi topmed program. *The American Journal of Human Genetics*, 108(5):874–893, 2021.

[100] Willem Huijbers, Elizabeth C Mormino, Sarah E Wigman, Andrew M Ward, Patrizia Vannini, Donald G McLaren, J Alex Becker, Aaron P Schultz, Trey Hedden, Keith A Johnson, et al. Amyloid deposition is linked to aberrant entorhinal activity among cognitively normal older adults. *Journal of Neuroscience*, 34(15):5200–5210, 2014.

[101] Clifford M Hurvich and Chih-Ling Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297–307, 1989.

[102] R Matthew Hutchison, Thilo Womelsdorf, Elena A Allen, Peter A Bandettini, Vince D Calhoun, Maurizio Corbetta, Stefania Della Penna, Jeff H Duyn, Gary H Glover, Javier Gonzalez-Castillo, et al. Dynamic functional connectivity: promise, issues, and interpretations. *Neuroimage*, 80:360–378, 2013.

[103] Rob Hyndman, George Athanasopoulos, Christoph Bergmeir, Gabriel Caceres, Leanne Chhay, Mitchell O'Hara-Wild, Fotios Petropoulos, Slava Razbash, Earo Wang, and

Farah Yasmeen. *forecast: Forecasting functions for time series and linear models*, 2023. R package version 8.21.

[104] Aapo Hyvärinen, Kun Zhang, Shohei Shimizu, and Patrik O Hoyer. Estimation of a structural vector autoregression model using non-gaussianity. *Journal of Machine Learning Research*, 11(5), 2010.

[105] Clifford R Jack Jr, David A Bennett, Kaj Blennow, Maria C Carrillo, Billy Dunn, Samantha Budd Haeberlein, David M Holtzman, William Jagust, Frank Jessen, Jason Karlawish, et al. Nia-aa research framework: toward a biological definition of alzheimer's disease. *Alzheimer's & Dementia*, 14(4):535–562, 2018.

[106] Hesamoddin Jahanian, Samantha Holdsworth, Thomas Christen, Hua Wu, Kangrong Zhu, Adam B Kerr, Matthew J Middione, Robert F Dougherty, Michael Moseley, and Greg Zaharchuk. Advantages of short repetition time resting-state functional mri enabled by simultaneous multi-slice imaging. *Journal of Neuroscience Methods*, 311:122–132, 2019.

[107] Ali Jalali, Christopher Johnson, and Pradeep Ravikumar. On learning discrete graphical models using greedy methods. *Advances in Neural Information Processing Systems*, 24, 2011.

[108] Shorena Janelidze, Henrik Zetterberg, Niklas Mattsson, Sebastian Palmqvist, Hugo Vanderstichele, Olof Lindberg, Danielle van Westen, Erik Stomrud, Lennart Minthon, Kaj Blennow, et al. Csf aβ42/aβ40 and aβ42/aβ38 ratios: better diagnostic markers of alzheimer disease. *Annals of clinical and translational neurology*, 3(3):154–165, 2016.

[109] Jana Jankova and Sara Van De Geer. Confidence intervals for high-dimensional inverse covariance estimation. *Electronic Journal of Statistics*, 9(1):1205–1229, 2015.

[110] Jana Janková and Sara van de Geer. Honest confidence regions and optimality in high-dimensional precision matrix estimation. *Test*, 26(1):143–162, 2017.

[111] Jana Janková and Sara van de Geer. Inference in high-dimensional graphical models. *arXiv preprint arXiv:1801.08512*, 2018.

[112] Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2):825–841, 2002.

[113] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith. Fsl. *Neuroimage*, 62(2):782–790, 2012.

[114] Longda Jiang, Zhili Zheng, Ting Qi, Kathryn E Kemper, Naomi R Wray, Peter M Visscher, and Jian Yang. A resource-efficient tool for mixed model association analysis of large-scale data. *Nature Genetics*, 51(12):1749–1755, 2019.

[115] Dan D Jobson, Yoshiki Hase, Andrew N Clarkson, and Rajesh N Kalaria. The role of the medial prefrontal cortex in cognition, ageing and dementia. *Brain communications*, 3(3):fcab125, 2021.

[116] Charles R Johnson and Herbert A Robinson. Eigenvalue inequalities for principal submatrices. *Linear Algebra and its Applications*, 37:11–22, 1981.

[117] David T Jones, Prashanthi Vemuri, Matthew C Murphy, Jeffrey L Gunter, Matthew L Senjem, Mary M Machulda, Scott A Przybelski, Brian E Gregg, Kejal Kantarci, David S Knopman, et al. Non-stationarity in the "resting brain's" modular architecture. *PloS one*, 7(6):e39731, 2012.

[118] Hyun Min Kang, Jae Hoon Sul, Susan K Service, Noah A Zaitlen, Sit-yee Kong, Nelson B Freimer, Chiara Sabatti, Eleazar Eskin, et al. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, 42(4):348–354, 2010.

[119] Dong-Hyun Kim, Elfar Adalsteinsson, Gary H Glover, and Daniel M Spielman. Regularized higher-order in vivo shimming. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 48(4):715–722, 2002.

[120] Dongmin Kim, Suvrit Sra, and Inderjit S Dhillon. A new projected quasi-newton

approach for the nonnegative least squares problem. Technical Report TR-06-54, Computer Science Department, University of Texas at Austin Austin, 2006.

[121] Peter Koval and Peter Kuppens. Changing emotion dynamics: individual differences in the effect of anticipatory social stress on emotional inertia. *Emotion*, 12(2):256, 2012.

[122] Jan Krumsiek, Karsten Suhre, Thomas Illig, Jerzy Adamski, and Fabian J Theis. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC systems biology*, 5(1):1–16, 2011.

[123] Prantik Kundu, Souheil J Inati, Jennifer W Evans, Wen-Ming Luh, and Peter A Bandettini. Differentiating bold and non-bold signals in fmri time series using multi-echo epi. *Neuroimage*, 60(3):1759–1770, 2012.

[124] Lisa M LaVange, William D Kalsbeek, Paul D Sorlie, Larissa M Avilés-Santa, Robert C Kaplan, Janice Barnhart, Kiang Liu, Aida Giachello, David J Lee, John Ryan, et al. Sample design and cohort selection in the hispanic community health study/study of latinos. *Annals of Epidemiology*, 20(8):642–649, 2010.

[125] Charles L Lawson and Richard J Hanson. *Solving least squares problems*, volume 15. Siam, 1995.

[126] Eek-Sung Lee, Kwangsun Yoo, Young-Beom Lee, Jinyong Chung, Ji-Eun Lim, Bora Yoon, and Yong Jeong. Default mode network functional connectivity in early and late mild cognitive impairment. *Alzheimer Disease & Associated Disorders*, 30(4):289–296, 2016.

[127] Namgil Lee and Jong-Min Kim. Dynamic functional connectivity analysis based on time-varying partial correlation with a copula-dcc-garch model. *Neuroscience Research*, 169:27–39, 2021.

[128] Brian Lenoski, Leslie C Baxter, Lina J Karam, José Maisog, and Josef Debbins. On the performance of autocorrelation estimation algorithms for fmri analysis. *IEEE Journal of Selected Topics in Signal Processing*, 2(6):828–838, 2008.

[129] Nora Leonardi and Dimitri Van De Ville. On spurious and real fluctuations of dynamic functional connectivity during rest. *Neuroimage*, 104:430–436, 2015.

[130] Piotr Lewczuk, Hermann Esselmann, Markus Otto, Juan Manuel Maler, Andreas Wolfram Henkel, Maria Kerstin Henkel, Oliver Eikenberg, Christof Antz, Wolf-Rainer Krause, Udo Reulbach, et al. Neurochemical diagnosis of alzheimer's dementia by csf aβ42, aβ42/aβ40 ratio and total tau. *Neurobiology of aging*, 25(3):273–281, 2004.

[131] Piotr Lewczuk, Natalia Lelental, Philipp Spitzer, Juan Manuel Maler, and Johannes Kornhuber. Amyloid-β 42/40 cerebrospinal fluid concentration ratio in the diagnostics of alzheimer's disease: validation of two novel assays. *Journal of Alzheimer's disease*, 43(1):183–191, 2015.

[132] Sai Li, T Tony Cai, and Hongzhe Li. Inference for high-dimensional linear mixed-effects models: A quasi-likelihood approach. *Journal of the American Statistical Association*, pages 1–33, 2021.

[133] Wai Keung Li and YV Hui. Estimation of random coefficient autoregressive process: an empirical bayes approach. *Journal of Time Series Analysis*, 4(2):89–94, 1983.

[134] Yun Li, Sijian Wang, Peter X-K Song, Naisyin Wang, Ling Zhou, and Ji Zhu. Doubly regularized estimation and selection in linear mixed-effects models for high-dimensional longitudinal data. *Statistics and its interface*, 11(4):721, 2018.

[135] Lina Lin, Mathias Drton, and Ali Shojaie. Estimation of high-dimensional graphical models using regularized score matching. *Electronic journal of statistics*, 10(1):806, 2016.

[136] Lina Lin, Mathias Drton, and Ali Shojaie. Statistical significance in high-dimensional linear mixed models. In *Proceedings of the 2020 ACM-IMS on Foundations of Data Science Conference*, pages 171–181, 2020.

[137] Martin A Lindquist, Yuting Xu, Mary Beth Nebel, and Brain S Caffo. Evaluating dynamic bivariate correlations in resting-state fmri: a comparison study and a new approach. *NeuroImage*, 101:531–546, 2014.

[138] Shiqing Ling and Michael McAleer. Asymptotic theory for a vector arma-garch model. *Econometric theory*, 19(2):280–310, 2003.

[139] Han Liu, Fang Han, Ming Yuan, John Lafferty, and Larry Wasserman. High-dimensional semiparametric gaussian copula graphical models. *The Annals of Statistics*, 40(4):2293–2326, 2012.

[140] Lon-Mu Liu and George C Tiao. Random coefficient first-order autoregressive models. *Journal of Econometrics*, 13(3):305–325, 1980.

[141] Weidong Liu. Gaussian graphical model estimation with false discovery rate control. *The Annals of Statistics*, 41(6):2948–2978, 2013.

[142] Xialu Liu and Rong Chen. Threshold factor models for high-dimensional time series. *Journal of Econometrics*, 216(1):53–70, 2020.

[143] Greta M Ljung and George EP Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

[144] Torben E Lund, Minna D Nørgaard, Egill Rostrup, James B Rowe, and Olaf B Paulson. Motion or activity: their role in intra-and inter-subject variation in fmri. *Neuroimage*, 26(3):960–964, 2005.

[145] Qingfei Luo, Masaya Misaki, Ben Mulyana, Chung-Ki Wong, and Jerzy Bodurka. Improved autoregressive model for correction of noise serial correlation in fast fmri. *Magnetic resonance in medicine*, 84(3):1293–1305, 2020.

[146] Cindy Lustig, Abraham Z Snyder, Mehul Bhakta, Katherine C O'Brien, Mark McAvoy, Marcus E Raichle, John C Morris, and Randy L Buckner. Functional deactivations: change with age and dementia of the alzheimer type. *Proceedings of the National Academy of Sciences*, 100(24):14504–14509, 2003.

[147] Jing Ma, Ali Shojaie, and George Michailidis. Network-based pathway enrichment analysis with incomplete network information. *Bioinformatics*, 32(20):3165–3174, 2016.

[148] Jing Ma, Ali Shojaie, and George Michailidis. A comparative study of topology-based pathway enrichment analysis methods. *BMC bioinformatics*, 20(1):546, 2019.

[149] Kaarina Matilainen, Esa A Mäntysaari, Martin H Lidauer, Ismo Strandén, and Robin Thompson. Employing a monte carlo algorithm in newton-type methods for restricted maximum likelihood estimation of genetic parameters. *PloS one*, 8(12):e80821–e80821, 2013.

[150] Brian W Matthews. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, 405(2):442–451, 1975.

[151] Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):1436–1462, 2006.

[152] Michael R Metel. Mini-batch stochastic gradient descent with dynamic sample sizes. *arXiv preprint arXiv:1708.00555*, 2017.

[153] Michelle M Mielke, Clinton E Hagen, Jing Xu, Xiyun Chai, Prashanthi Vemuri, Val J Lowe, David C Airey, David S Knopman, Rosebud O Roberts, Mary M Machulda, et al. Plasma phospho-tau181 increases with alzheimer's disease clinical severity and is associated with tau-and amyloid-positron emission tomography. *Alzheimer's & Dementia*, 14(8):989–997, 2018.

[154] Ricardo Pio Monti, Christoforos Anagnostopoulos, and Giovanni Montana. Learning population and subject-specific brain connectivity networks via mixed neighborhood selection. *The Annals of Applied Statistics*, pages 2142–2164, 2017.

[155] Victoria L Morgan, Baxter P Rogers, Hasan H Sonmezturk, John C Gore, and Bassel Abou-Khalil. Cross hippocampal influence in mesial temporal lobe epilepsy measured with high temporal resolution functional magnetic resonance imaging. *Epilepsia*, 52(9):1741–1749, 2011.

[156] Elizabeth C Mormino, Andre Smiljic, Amynta O Hayenga, Susan H. Onami, Michael D Greicius, Gil D Rabinovici, Mustafa Janabi, Suzanne L Baker, Irene V. Yen, Cindee M

Madison, et al. Relationships between beta-amyloid and functional connectivity in different components of the default mode network in aging. *Cerebral cortex*, 21(10):2399–2407, 2011.

[157] Jeanette A Mumford and Thomas Nichols. Modeling and inference of multisubject fmri data. *IEEE Engineering in Medicine and Biology Magazine*, 25(2):42–51, 2006.

[158] Tomoaki Nakatani. *ccgarch2: An R Package for Modelling Multivariate GARCH Models with Conditional Correlations*, 2016. R package version 0.0.0-42, URL ¡http://CRAN.r-project/package=ccgarch¿.

[159] Balgobin Nandram and Joseph D Petruccelli. A bayesian analysis of autoregressive time series panel data. *Journal of Business & Economic Statistics*, 15(3):328–334, 1997.

[160] Manjari Narayan and Genevera I Allen. Mixed effects models for resampled network statistics improves statistical power to find differences in multi-subject functional connectivity. *Frontiers in neuroscience*, 10:108, 2016.

[161] Matey Neykov, Yang Ning, Jun S. Liu, and Han Liu. A Unified Theory of Confidence Regions and Testing for High-Dimensional Estimating Equations. *Statistical Science*, 33(3):427 – 443, 2018.

[162] Bernard Ng, Gaël Varoquaux, Jean Baptiste Poline, and Bertrand Thirion. A novel sparse group gaussian graphical model for functional connectivity estimation. In *International Conference on Information Processing in Medical Imaging*, pages 256–267. Springer, 2013.

[163] DF Nicholls and BG Quinn. The estimation of multivariate random coefficient autoregressive models. *Journal of Multivariate Analysis*, 11(4):544–555, 1981.

[164] Lisa D Nickerson, Stephen M Smith, Döst Öngür, and Christian F Beckmann. Using dual regression to investigate network shape and amplitude in functional connectivity analyses. *Frontiers in neuroscience*, 11:115, 2017.

[165] Wiktor Olszowy, John Aston, Catarina Rua, and Guy B Williams. Accurate autocorrelation modeling substantially improves fmri reliability. *Nature communications*, 10(1):1220, 2019.

[166] Daniel A Orringer, David R Vago, and Alexandra J Golby. Clinical applications and future directions of functional mri. In *Seminars in neurology*, volume 32, pages 466–475. Thieme Medical Publishers, 2012.

[167] Elisabeth Orskaug. Multivariate dcc-garch model:-with various error distributions. Master's thesis, Institutt for matematiske fag, 2009.

[168] George C O'Neill, Prejaas K Tewarie, Giles L Colclough, Lauren E Gascoyne, Benjamin AE Hunt, Peter G Morris, Mark W Woolrich, and Matthew J Brookes. Measurement of dynamic task related functional networks using meg. *NeuroImage*, 146:667–678, 2017.

[169] H Desmond Patterson and Robin Thompson. Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3):545–554, 1971.

[170] William D Penny, Karl J Friston, John T Ashburner, Stefan J Kiebel, and Thomas E Nichols. *Statistical parametric mapping: the analysis of functional brain images*. Elsevier, 2011.

[171] Hong Thom Pham and Bo-Suk Yang. Estimation and forecasting of machine health condition using arma/garch model. *Mechanical systems and signal processing*, 24(2):546–558, 2010.

[172] Dimitris N Politis, Joseph P Romano, and Michael Wolf. *Subsampling*. Springer Science & Business Media, 1999.

[173] Huihui Qi, Hao Liu, Haimeng Hu, Huijin He, and Xiaohu Zhao. Primary disruption of the memory-related subsystems of the default mode network in alzheimer's disease: resting-state functional connectivity mri study. *Frontiers in aging neuroscience*, 10:344, 2018.

[174] Xinghao Qiao, Shaojun Guo, and Gareth M James. Functional graphical models. *Journal of the American Statistical Association*, 114(525):211–222, 2019.

[175] C Radhakrishna Rao. Estimation of heteroscedastic variances in linear models. *Journal of the American Statistical Association*, 65(329):161–172, 1970.

[176] Dieter Rasch and O Masata. Methods of variance component estimation. *Czech Journal of Animal Science*, 51(6):227, 2006.

[177] Avik Ray, Sujay Sanghavi, and Sanjay Shakkottai. Improved greedy algorithms for learning graphical models. *IEEE Transactions on Information Theory*, 61(6):3457–3468, 2015.

[178] Marta Regis, Paulo Serra, and Edwin R van den Heuvel. Random autoregressive models: A structured overview. *Econometric Reviews*, 41(2):207–230, 2022.

[179] Zhao Ren, Tingni Sun, Cun-Hui Zhang, and Harrison H Zhou. Asymptotic normality and optimalities in estimation of large gaussian graphical models. *The Annals of Statistics*, 43(3):991–1026, 2015.

[180] R Riccelli, Luca Passamonti, Andrea Duggento, Maria Guerrisi, Iole Indovina, A Terracciano, and Nicola Toschi. Dynamical brain connectivity estimation using garch models: An application to personality neuroscience. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 3305–3308. IEEE, 2017.

[181] Omar Rivasplata. Subgaussian random variables: An expository note. *Internet publication, PDF*, 5, 2012.

[182] Alard Roebroeck, Elia Formisano, and Rainer Goebel. Mapping directed influence over the brain using granger causality and fmri. *Neuroimage*, 25(1):230–242, 2005.

[183] Bjorn Roelstraete and Yves Rosseel. Fiar: an r package for analyzing functional integration in the brain. *Journal of Statistical Software*, 44:1–32, 2011.

[184] Mark Rudelson and Roman Vershynin. Hanson-wright inequality and sub-gaussian concentration. *Electronic Communications in Probability*, 18:1–9, 2013.

[185] Abolfazl Safikhani and Ali Shojaie. Joint structural break detection and parameter estimation in high-dimensional nonstationary var models. *Journal of the American Statistical Association*, 117(537):251–264, 2022.

[186] Ashish Kaul Sahib, Klaus Mathiak, Michael Erb, Adham Elshahabi, Silke Klamer, Klaus Scheffler, Niels K Focke, and Thomas Ethofer. Effect of temporal resolution and serial autocorrelations in event-related functional mri. *Magnetic resonance in medicine*, 76(6):1805–1813, 2016.

[187] Ünal Sakoğlu, Godfrey D Pearlson, Kent A Kiehl, Y Michelle Wang, Andrew M Michael, and Vince D Calhoun. A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 23:351–366, 2010.

[188] Gholamreza Salimi-Khorshidi, Gwenaëlle Douaud, Christian F Beckmann, Matthew F Glasser, Ludovica Griffanti, and Stephen M Smith. Automatic denoising of functional mri data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage*, 90:449–468, 2014.

[189] Kathrin Schacke. On the kronecker product. *Master's thesis, University of Waterloo*, 2004.

[190] Shayle R Searle. An overview of variance component estimation. *Metrika*, 42(1):215–230, 1995.

[191] Dennis J Selkoe. Presenilin, notch, and the genesis and treatment of alzheimer's disease. *Proceedings of the National Academy of Sciences*, 98(20):11039–11041, 2001.

[192] Liew Khim Sen, Mahnendran Shitan, et al. The performance of aicc as an order selection criterion in arma time series models. *Pertanika Journal of Science and Technology*, 10(1):25–33, 2002.

[193] Kawin Setsompop, Borjan A Gagoski, Jonathan R Polimeni, Thomas Witzel, Van J Wedeen, and Lawrence L Wald. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging with reduced g-factor penalty. *Magnetic resonance in medicine*, 67(5):1210–1224, 2012.

[194] Zarrar Shehzad, AM Clare Kelly, Philip T Reiss, Dylan G Gee, Kristin Gotimer, Lucina Q Uddin, Sang Han Lee, Daniel S Margulies, Amy Krain Roy, Bharat B Biswal, et al. The resting brain: unconstrained yet reliable. *Cerebral cortex*, 19(10):2209–2229, 2009.

[195] Yvette I Sheline, Marcus E Raichle, Abraham Z Snyder, John C Morris, Denise Head, Suzhi Wang, and Mark A Mintun. Amyloid plaques disrupt resting state default mode network connectivity in cognitively normal elderly. *Biological psychiatry*, 67(6):584–587, 2010.

[196] Ali Shojaie. Differential network analysis: A statistical perspective. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2020.

[197] Ali Shojaie and Emily B Fox. Granger causality: A review and recent advances. *Annual Review of Statistics and Its Application*, 9:289–319, 2022.

[198] Ali Shojaie and George Michailidis. Analysis of gene sets based on the underlying regulatory network. *Journal of Computational Biology*, 16(3):407–426, 2009.

[199] Ali Shojaie and George Michailidis. Discovering graphical granger causality using the truncating lasso penalty. *Bioinformatics*, 26(18):i517–i523, 2010.

[200] Richmond Siaw, Eric Ofosu-Hene, and Evans Tee. Investment portfolio optimization with garch models. *ELK Asia Pacific Journal of Finance and Risk Management*, 8(2), 2017.

[201] Annastiina Silvennoinen and Timo Teräsvirta. Multivariate garch models. In *Handbook of financial time series*, pages 201–229. Springer, 2009.

[202] Stephen M Smith, Christian F Beckmann, Jesper Andersson, Edward J Auerbach, Janine Bijsterbosch, Gwenaëlle Douaud, Eugene Duff, David A Feinberg, Ludovica Griffanti, Michael P Harms, et al. Resting-state fmri in the human connectome project. *Neuroimage*, 80:144–168, 2013.

[203] Stephen M Smith, Aapo Hyvärinen, Gaël Varoquaux, Karla L Miller, and Christian F Beckmann. Group-pca for very large fmri datasets. *Neuroimage*, 101:738–749, 2014.

[204] Stephen M Smith, Karla L Miller, Gholamreza Salimi-Khorshidi, Matthew Webster, Christian F Beckmann, Thomas E Nichols, Joseph D Ramsey, and Mark W Woolrich. Network modelling methods for fmri. *Neuroimage*, 54(2):875–891, 2011.

[205] Stephen M Smith, Thomas E Nichols, Diego Vidaurre, Anderson M Winkler, Timothy EJ Behrens, Matthew F Glasser, Kamil Ugurbil, Deanna M Barch, David C Van Essen, and Karla L Miller. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nature Neuroscience*, 18(11):1565–1567, 2015.

[206] Tamar Sofer. Confidence intervals for heritability via haseman-elston regression. *Statistical Applications in Genetics and Molecular Biology*, 16(4):259–273, 2017.

[207] Eftychia Solea and Bing Li. Copula gaussian graphical models for functional data. *Journal of the American Statistical Association*, pages 1–13, 2020.

[208] Song Song and Peter J Bickel. Large vector auto regressions. *arXiv preprint arXiv:1106.3915*, 2011.

[209] Christian Sorg, Valentin Riedl, Mark Muhlau, Vince D Calhoun, Tom Eichele, Leonhard Läer, Alexander Drzezga, Hans Förstl, Alexander Kurz, Claus Zimmer, et al. Selective changes of resting-state networks in individuals at risk for alzheimer's disease. *Proceedings of the National Academy of Sciences*, 104(47):18760–18765, 2007.

[210] Paul D Sorlie, Larissa M Avilés-Santa, Sylvia Wassertheil-Smoller, Robert C Kaplan, Martha L Daviglus, Aida L Giachello, Neil Schneiderman, Leopoldo Raij, Gregory

Talavera, Matthew Allison, Lisa Lavange, Lloyd E Chambless, and Gerardo Heiss. Design and implementation of the hispanic community health study/study of latinos. *Annals of Epidemiology*, 20(8):629–41, Aug 2010.

[211] O. Sporns. Brain connectivity. *Scholarpedia*, 2(10):4695, 2007. revision #91084.

[212] Adam A Szpiro, Paul D Sampson, Lianne Sheppard, Thomas Lumley, Sara D Adar, and Joel D Kaufman. Predicting intra-urban variation in air pollution concentrations with complex spatio-temporal dependencies. *Environmetrics*, 21(6):606–631, 2010.

[213] Enzo Tagliazucchi, Frederic Von Wegner, Astrid Morzelewski, Verena Brodbeck, and Helmut Laufs. Dynamic bold functional connectivity in humans and its electrophysiological correlates. *Frontiers in human neuroscience*, 6:339, 2012.

[214] Alex Tank, Ian Covert, Nicholas Foti, Ali Shojaie, and Emily B Fox. Neural granger causality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(8):4267–4279, 2021.

[215] Alex Tank, Xiudi Li, Emily B Fox, and Ali Shojaie. The convex mixture distribution: Granger causality for categorical time series. *SIAM Journal on Mathematics of Data Science*, 3(1):83–112, 2021.

[216] TCGA. Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418):61, 2012.

[217] Stefan J Teipel, Coraline D Metzger, Frederic Brosseron, Katharina Buerger, Katharina Brueggen, Cihan Catak, Dominik Diesing, Laura Dobisch, Klaus Fliebach, Christiana Franke, et al. Multicenter resting state functional connectivity in prodromal and dementia stages of alzheimer's disease. *Journal of Alzheimer's Disease*, 64(3):801–813, 2018.

[218] William F Trench. Asymptotic distribution of the spectra of a class of generalized kac–murdock–szegö matrices. *Linear algebra and its applications*, 294(1-3):181–192, 1999.

[219] Joel A Tropp. An introduction to matrix concentration inequalities. *arXiv preprint arXiv:1501.01571*, 2015.

[220] Yiu Kuen Tse and Albert K C Tsui. A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business & Economic Statistics*, 20(3):351–362, 2002.

[221] David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79, 2013.

[222] Pavel Vaněček. Estimators of random coefficient autoregressive models. 2008.

[223] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

[224] Arend Voorman, Ali Shojaie, and Daniela Witten. Graph estimation with joint additive models. *Biometrika*, 101(1):85–101, 2014.

[225] Liang Wang, Yufeng Zang, Yong He, Meng Liang, Xinqing Zhang, Lixia Tian, Tao Wu, Tianzi Jiang, and Kuncheng Li. Changes in hippocampal connectivity in the early stages of alzheimer's disease: evidence from resting state fmri. *Neuroimage*, 31(2):496–504, 2006.

[226] Lijuan Peggy Wang, Ellen Hamaker, and CS22924600 Bergeman. Investigating inter-individual differences in short-term intra-individual variability. *Psychological methods*, 17(4):567, 2012.

[227] Wei Wang. Identifiability of linear mixed effects models. *Electronic Journal of Statistics*, 7:244–263, 2013.

[228] Xu Wang, Mladen Kolar, and Ali Shojaie. Statistical inference for networks of high-dimensional point processes. *arXiv preprint arXiv:2007.07448*, 2020.

[229] Yuchung J Wang and Edward H Ip. Conditionally specified continuous distributions. *Biometrika*, 95(3):735–746, 2008.

[230] Andreas Weissenbacher, Christian Kasess, Florian Gerstl, Rupert Lanzenberger, Ewald Moser, and Christian Windischberger. Correlations and anticorrelations in resting-state functional connectivity mri: a quantitative comparison of preprocessing strategies. *Neuroimage*, 47(4):1408–1416, 2009.

[231] Marijke Welvaert and Yves Rosseel. On the definition of signal-to-noise ratio and contrast-to-noise ratio for fmri data. *PloS one*, 8(11):e77089, 2013.

[232] Susan Whitfield-Gabrieli and Judith M Ford. Default mode network activity and connectivity in psychopathology. *Annual review of clinical psychology*, 8:49–76, 2012.

[233] Genevieve L Wojcik, Mariaelisa Graff, Katherine K Nishimura, Ran Tao, Jeffrey Haessler, Christopher R Gignoux, Heather M Highland, Yesha M Patel, Elena P Sorokin, Christy L Avery, et al. Genetic analyses of diverse populations improves discovery for complex traits. *Nature*, 570(7762):514–518, 2019.

[234] Mark W Woolrich, Brian D Ripley, Michael Brady, and Stephen M Smith. Temporal autocorrelation in univariate linear modeling of fmri data. *Neuroimage*, 14(6):1370–1386, 2001.

[235] Minge Xie and Yaning Yang. Asymptotics for generalized estimating equations with large cluster sizes. *The Annals of Statistics*, 31(1):310–347, 2003.

[236] Lingzhou Xue and Hui Zou. Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics*, 40(5):2541–2571, 2012.

[237] Ayumu Yamashita, Noriaki Yahata, Takashi Itahashi, Giuseppe Lisi, Takashi Yamada, Naho Ichikawa, Masahiro Takamura, Yujiro Yoshihara, Akira Kunimatsu, Naohiro Okada, et al. Harmonization of resting-state functional mri data across multiple imaging sites via the separation of site differences into sampling bias and measurement bias. *PLoS biology*, 17(4):e3000042, 2019.

[238] Eunho Yang, Pradeep Ravikumar, Genevera I Allen, and Zhandong Liu. Graphical models via univariate exponential family distributions. *The Journal of Machine Learning Research*, 16(1):3813–3847, 2015.

[239] Jianxin Yin and Hongzhe Li. A sparse conditional gaussian graphical model for analysis of genetical genomics data. *The annals of applied statistics*, 5(4):2630, 2011.

[240] Meichen Yu, Kristin A Linn, Philip A Cook, Mary L Phillips, Melvin McInnis, Maurizio Fava, Madhukar H Trivedi, Myrna M Weissman, Russell T Shinohara, and Yvette I Sheline. Statistical harmonization corrects site effects in functional connectivity measurements from multi-site fmri data. *Human brain mapping*, 39(11):4213–4227, 2018.

[241] Shiqing Yu, Mathias Drton, and Ali Shojaie. Generalized score matching for non-negative data. *The Journal of Machine Learning Research*, 20(1):2779–2848, 2019.

[242] Ming Yuan and Yi Lin. Model selection and estimation in the gaussian graphical model. *Biometrika*, 94(1):19–35, 2007.

[243] Samuel Zahl. Bounds for the central limit theorem error. *SIAM Journal on Applied Mathematics*, 14(6):1225–1245, 1966.

[244] Krzysztof Zajkowski. Bounds on tail probabilities for quadratic forms in dependent sub-gaussian random variables. *Statistics & Probability Letters*, 167:108898, 2020.

[245] Aiying Zhang, Jian Fang, Faming Liang, Vince D Calhoun, and Yu-Ping Wang. Aberrant brain connectivity in schizophrenia detected via a fast gaussian graphical model. *IEEE journal of biomedical and health informatics*, 23(4):1479–1489, 2018.

[246] Cun-Hui Zhang and Stephanie S Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242, 2014.

[247] Hong-Ying Zhang, Shi-Jie Wang, Jiong Xing, Bin Liu, Zhan-Long Ma, Ming Yang, Zhi-Jun Zhang, and Gao-Jun Teng. Detection of pcc functional connectivity characteristics in resting-state fmri in mild alzheimer's disease. *Behavioural brain research*, 197(1):103–108, 2009.

[248] Kunhui Zhang, Abolfazl Safikhani, Alex Tank, and Ali Shojaie. Penalized estimation of threshold auto-regressive models with many components and thresholds. *Electronic Journal of Statistics*, 16(1):1891–1951, 2022.

[249] Yuqing Zhang, Giovanni Parmigiani, and W Evan Johnson. Combat-seq: batch effect adjustment for rna-seq count data. *NAR genomics and bioinformatics*, 2(3):lqaa078, 2020.

[250] Haitao Zhao and Zhong-Hui Duan. Cancer genetic network inference using gaussian graphical models. *Bioinformatics and biology insights*, 13:1177932219839402, 2019.

[251] Weiqi Zhao, Carolina Makowski, Donald J Hagler, Hugh P Garavan, Wesley K Thompson, Deanna J Greene, Terry L Jernigan, and Anders M Dale. Task fmri paradigms may capture more behaviorally relevant information than resting-state functional connectivity. *Neuroimage*, 270:119946, 2023.

[252] Lili Zheng and Garvesh Raskutti. Testing for high-dimensional network parameters in auto-regressive models. *Electronic Journal of Statistics*, 13(2):4977–5043, 2019.

[253] Xiang Zhou. A unified framework for variance component estimation with summary statistics in genome-wide association studies. *The Annals of Applied Statistics*, 11(4):2027, 2017.

[254] Marek Zinecker, Adam P Balcerzak, Marcin Faldzinski, Michal Bernad Pietrzak, and Tomáš Meluzin. Application of dcc-garch model for analysis of interrelations among capital markets of poland, czech republic and germany. Technical report, Institute of Economic Research Working Papers, 2016.

Appendix A

APPENDIX FOR CHAPTER 2

A.1 HE and REHE Properties

A.1.1 HE Unbiasedness

The HE estimator is unbiased. The unbiasedness holds even if we use only a subset of the n^2 estimating equations, i.e., ignoring some entries of the relatedness matrices D_k's.

Recall n^2 estimating equations are constructed:

$$\mathrm{E}[Y_i Y_j] = \sigma_0^2 D_0^{ij} + \sigma_1^2 D_1^{ij}, \quad i, j = 1,\ 2, \dots\ n.$$

Suppose we use S out of the n^2 estimating equations without duplicates to compute the HE estimates, $S \leq n^2$. Let $(i_1, j_1), \dots, (i_S, j_S)$ denote the selected indices. Similar to Section 2.2.1, we recast the problem of solving the S estimating equations as the following linear regression problem. Let $\tilde{Y}_S = (Y_{i_1} Y_{j_1}, \dots, Y_{i_S} Y_{j_S})$ denote the linear regression outcome vector, and let $\tilde{X}_S$ denote the design matrix by stacking the tuples $(D_0^{i_s j_s}, D_1^{i_s, j_s})$, $s = 1, \dots, S$. Note that we have

$$\mathrm{E}[\tilde{Y}_S] = \tilde{X}_S \boldsymbol{\sigma}^2,$$

where the equality holds regardless of the higher moments of Y or the positions of the S indices. The closed form expression of the HE estimates in this case is $\hat{\boldsymbol{\sigma}}^2 = (\tilde{X}_S^\top \tilde{X}_S)^{-1} \tilde{X}_S^\top \tilde{Y}_S$. We thus have:

$$\mathrm{E}(\hat{\boldsymbol{\sigma}}^2) = (\tilde{X}_S^\top \tilde{X}_S)^{-1} \tilde{X}_S^\top \mathrm{E}(\tilde{Y}_S) = (\tilde{X}_S^\top \tilde{X}_S)^{-1} \tilde{X}_S^\top \tilde{X}_S \boldsymbol{\sigma}^2 = \boldsymbol{\sigma}^2.$$

Therefore, the HE estimates is unbiased, even if a subset of the n^2 estimating equations is used.

A.1.2 REHE Consistency and Asymptotic Normality

The REHE estimator is consistent under mild conditions, and is asymptotically normal when the random effect correlation matrix is sparse and block-diagonal.

For expositional clarity, we assume the diagonal entries of the kinship matrix D_1 are scaled to 1. We proceed with first establishing consistency of the HE estimator. By (2.3), the HE estimate is the ordinary least squares solution to a linear regression problem. We write out the model for this linear regression as

$$\tilde{Y} = \tilde{X}\boldsymbol{\sigma^2} + \tilde{\epsilon},$$

for the length n^2 outcome vector $\tilde{Y} = \text{vec}(YY^\top)$, the $n^2 \times 2$ fixed effect design matrix $\tilde{X} = (\text{vec}(D_0), \text{vec}(D_1))$ and the length n^2 error vector $\tilde{\epsilon}$ with covariance matrix R. Note that $R^{ij} = \text{Cov}(\tilde{Y}_i, \tilde{Y}_j)$, which is related to the forth moment of Y. Thus R is not a diagonal matrix. Under this framework, the HE estimator is a special case of the generalized estimating equations estimator, with I_n as the *working correlation matrix*, the total number of clusters bounded (equals to 1 in this case), and the cluster size n^2 going to infinity. Such as estimator has been studied in [235], see Example 2.1 and Example 5.1. The consistency of HE estimates holds under the condition (I_ω) or (I^*_ω) in [235], which in our case corresponds to

$$(I_\omega) : \lambda_{\min}(H_n M_n^{-1} H_n) \xrightarrow{n\to\infty} \infty$$
$$(I^*_\omega) : \lambda_{\min}(H_n)/\lambda_{\max}(R) \xrightarrow{n\to\infty} \infty$$

where

$$H_n = \tilde{X}^\top \tilde{X}$$
$$M_n = \tilde{X}^\top R \tilde{X},$$

and $\lambda_{\min}(X)$ ($\lambda_{\max}(X)$) denotes the smallest (largest) eigenvalue of the symmetric matrix X. As discussed in [235], for (I_ω) or (I^*_ω) to hold it suffices to bound the correlation of

$\tilde{\epsilon}$ by $|R^{ij}/\sqrt{R^{ii}R^{jj}}| \leq \rho_h$, where $h = |i-j|$ and $\{\rho_h\}_{h=1}^{\infty}$ is a sequence of values with $\lim_{h\to\infty} \rho_h = 0$. One can then verify that Assumption 6 is sufficient to guarantee that the correlations of $\tilde{\epsilon}$ are properly bounded, based on $\mathrm{Cov}(Y_iY_j, Y_{i'}Y_{j'}) = V_Y^{ii'}V_Y^{jj'} + V_Y^{ij'}V_Y^{ji'}$, for $V_Y = \mathrm{E}(YY^\top) = \sigma_0^2 D_0 + \sigma_1^2 D_1$.

Assumption 6 (Weak Condition)**.** *The covariance matrix D_1 satisfies $\left|D_1^{ij}\right| \leq \rho_h^*$, $h = |i-j|$, for some sequence of values $\{\rho_h^*\}_{h=1}^{\infty}$ with $\lim_{h\to\infty} \rho_h^* = 0$.*

Assumption 6 is natural in many genetic studies as it is expected that off-diagonal correlation entries would decrease to zero after properly rearranging the rows and columns of the correlation matrix.

The asymptotic normality of the HE estimator requires an stronger assumption on the structure of D_1. Note that we do not need this strong assumption to show the consistency of REHE or to construct bootstrap confidence intervals for the REHE estimator. Suppose Assumption 7 holds for a sparse and block-diagonal kinship matrix D_1.

Assumption 7 (Strong Condition)**.** *D_1 is sparse and block-diagonal such that*

$$D_1 = \begin{pmatrix} D_1^{(1)} & 0 & 0 & \dots \\ 0 & D_1^{(2)} & 0 & \dots \\ \vdots & & \ddots & \\ 0 & 0 & 0 & D_1^{(M)} \end{pmatrix},$$

where $D_1^{(m)}$ $(m = 1, \dots, M)$ are square blocks along the diagonal of D_1, and M is the total number of blocks.

Such correlation structures are often approximately satisfied in genetic studies: for example, subjects are highly genetically correlated within the same family, and are remotely correlated across families. Let s_m denote the number of rows/columns in $D_1^{(m)}$, and $Y^{(m)}$ denote the $s_m \times 1$ subvector of Y corresponding to the m^{th} block. For example, $Y^{(1)}$ is the subvector corresponding to the first s_1 elements of Y. By construction, $Y^{(m)}$'s are independent and normally distributed with zero mean and covariance $\sigma_0^2 I_{s_m} + \sigma_1^2 D_1^{(m)}$, for $m = 1, \dots, M$. For

sparse block-diagonal D_1, we simplify $\tilde{Y}$ and $\tilde{X}$ in (2.3) by discarding elements corresponding to zero entries of D_1:

$$\tilde{Y} = \begin{pmatrix} \tilde{Y}^{(1)} \\ \tilde{Y}^{(2)} \\ \vdots \\ \tilde{Y}^{(M)} \end{pmatrix}, \quad \tilde{X} = \begin{pmatrix} \tilde{X}^{(1)} \\ \tilde{X}^{(2)} \\ \vdots \\ \tilde{X}^{(M)} \end{pmatrix},$$

where we denote $\tilde{Y}^{(m)} = \text{vec}\left(Y^{(m)}Y^{(m)\top}\right)$ and $\tilde{X}^{(m)} = \left(\text{vec}\left(I_n^{(m)}\right), \text{vec}\left(D_1^{(m)}\right)\right)$, for $m = 1, \ldots, M$. The independence of $Y^{(m)}$'s implies independence of $\tilde{Y}^{(m)}$'s. As the number of blocks $M \to \infty$, the HE estimates $\hat{\boldsymbol{\sigma}}_M^2 = \left(\hat{\sigma}_{0,M}^2, \hat{\sigma}_{1,M}^2\right)$ converge in probability to the true variance components $\boldsymbol{\sigma}^2 = \left(\sigma_0^2, \sigma_1^2\right)$ at a rate of $\sqrt{M}$ or faster [235]. Moreover, when the maximum block size $s = \max_{1 \le m \le M}(s_m)$ does not go to infinity too fast as $M \to \infty$, the HE estimates are asymptotically normal, with a rate of $\sqrt{M}$ [235]:

$$W_M^{-1/2} H_M (\hat{\boldsymbol{\sigma}}_M^2 - \boldsymbol{\sigma}^2) \xrightarrow{d} N_2(0, I_2), \tag{A.1}$$

where

$$W_M = \sum_{m=1}^{M} \left(\tilde{X}^{(m)}\right)^\top R^{(m)} \tilde{X}^{(m)}, \quad R^{(m)} = \text{Cor}\left(\tilde{Y}^{(m)}\right), \quad H_M = \sum_{m=1}^{M} \left(\tilde{X}^{(m)}\right)^\top \tilde{X}^{(m)}.$$

In genetic studies, it is typical that subjects belong to small unrelated groups so that s is bounded, and the number of groups M increases with increasing sample sizes. These settings satisfy the conditions for the HE estimator's consistency and asymptotic normality.

We are now ready to establish the asymptotic properties of the REHE estimator. As implied by (2.4), the REHE estimates $\tilde{\boldsymbol{\sigma}}^2$ are different from the HE estimates $\hat{\boldsymbol{\sigma}}^2$ only when HE yields negative estimates for some variance components. Let $\mathbb{P}\left(\hat{\sigma}_0^2 \ge 0,\ \hat{\sigma}_1^2 \ge 0\right)$ denote the probability of the HE estimates being non-negative, which equals $\mathbb{P}\left(\tilde{\boldsymbol{\sigma}}^2 = \hat{\boldsymbol{\sigma}}^2\right)$. Based on the consistency of the HE estimator, we have:

$$\mathbb{P}\left(\tilde{\boldsymbol{\sigma}}^2 = \hat{\boldsymbol{\sigma}}^2\right) = \mathbb{P}\left(\hat{\sigma}_0^2 \ge 0,\ \hat{\sigma}_1^2 \ge 0\right) \xrightarrow{n \to \infty} 1,$$

indicating asymptotic equivalence of the HE and REHE estimators. This implies the consistency of the REHE estimator, which only requires Assumption 6 on D_1. Note that despite their asymptotic equivalence, our simulation results in Section 2.4 clearly show the advantages of REHE over HE in finite samples.

Next, we show that under the stronger Assumption 7, $W_M^{-1/2} H_M(\tilde{\boldsymbol{\sigma}}^2 - \hat{\boldsymbol{\sigma}}^2)$ is $o_p(1)$:

$$\mathbb{P}\left(W_M^{-1/2} H_M(\tilde{\boldsymbol{\sigma}}^2 - \hat{\boldsymbol{\sigma}}^2) = 0\right) \geq \mathbb{P}\left(\tilde{\boldsymbol{\sigma}}^2 - \hat{\boldsymbol{\sigma}}^2 = 0\right) \xrightarrow{M \to \infty} 1.$$

We then have

$$\begin{aligned} W_M^{-1/2} H_M(\tilde{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2) &= W_M^{-1/2} H_M(\tilde{\boldsymbol{\sigma}}^2 - \hat{\boldsymbol{\sigma}}^2) + W_M^{-1/2} H_M(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2) \\ &= o_p(1) + W_M^{-1/2} H_M(\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2). \end{aligned}$$

Therefore, we have established that under Assumption 7, the REHE estimator $\tilde{\boldsymbol{\sigma}}^2$ satisfies:

$$W_M^{-1/2} H_M(\tilde{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2) \xrightarrow{d} N_2(0, I_2). \tag{A.2}$$

Here we note that the REHE estimator is slightly biased in finite samples due to the correction of negative estimates. However, as we have observed in the simulation studies, the variance and mean squared error of the REHE estimator were smaller than those of the HE estimator, indicating the REHE a better estimator.

A.2 Extensions of REHE

A.2.1 Analytical Solutions to REHE

With only two variance components in the linear mixed model ($K = 1$), there is a closed form solution to (2.4) in the main paper for REHE variance component estimates. Equation

(2.4) in the main paper can be reformulated as:

$$(x, y) = \arg\min_{x,y} \; ax^2 + by^2 + cxy + dx + ey + f$$
$$\text{subject to } x, y \geq 0, \; a, b > 0, \; 4ab \neq c^2.$$

The condition $4ab \neq c^2$ corresponds to matrix $\tilde{X}^\top \tilde{X}$ being invertible, where $\tilde{X}$ is introduced in the main paper Section 2.1 as the design matrix in the HE regression. The closed form solution is:

$$(x, y) = \begin{cases} (\frac{ce - 2bd}{4ab - c^2}, \frac{cd - 2ae}{4ab - c^2}), & \text{if } \frac{ce - 2bd}{4ab - c^2} \geq 0, \; \frac{cd - 2ae}{4ab - c^2} \geq 0; \\ (0, -\frac{e}{2b}), & \text{if } \frac{ce - 2bd}{4ab - c^2} < 0, \; \frac{e}{2b} \leq 0; \\ (-\frac{d}{2a}, 0), & \text{if } \frac{cd - 2ae}{4ab - c^2} < 0, \; \frac{d}{2a} \leq 0; \\ (0, 0), & \text{if } \frac{d}{2a} > 0, \; \frac{e}{2b} > 0. \end{cases} \tag{A.3}$$

A.2.2 REHE Confidence Interval Construction with Sparse Correlation Matrix

When the sample size n is large and the correlation matrix D_1 is dense, constructing REHE confidence intervals is computationally demanding. The main burden comes from matrix decomposition when sampling $\tilde{Y}^*$ from $N_n(0, \Sigma)$ in step (a) of Algorithm 2 in the main paper. Subsample $\tilde{Y}^*$ is commonly obtained by computing $\Sigma^{1/2}e$, where $\Sigma = \Sigma^{1/2}(\Sigma^{1/2})^\top$ and e sampled from $N_n(0, I_n)$. Matrix $\Sigma^{1/2}$ is usually obtained by Cholesky decomposition or singular value decomposition, which requires $O(n^3)$ operations. We propose to accelerate this procedure through approximating the dense correlation matrix D_1 with a sparse block-diagonal matrix D_1^* [75]. In genetics studies, it is reasonable to assume high relatedness within small groups of subjects and low relatedness across groups, such as for kinship relatedness and household membership. Such relatedness structure can be well-approximated by sparse block-diagonal correlation matrices, through thresholding the off-diagonal entries of the original dense correlation matrices and rearranging the rows/columns [75]. The resulting sparse block-diagonal correlation matrices can substantially speed up the matrix decomposition step, and thus the inference procedure. Bootstrap confidence intervals obtained with sparse D_1^* are

expected to have good coverage. We used the function `makeSparseMatrix` in the R package `GENESIS` (*v2.14.3*, 40) to implement the correlation matrix sparsification procedure. The sparsification threshold was set at $2^{-5.5}$, which corresponds to the fifth degree relatedness for the kinship matrix [75].

A.2.3 REHE and reREHE for Network-based Pathway Enrichment Analysis

Network-based pathway enrichment analysis assumes the following linear mixed model for the association between gene expressions and the phenotype [147]:

$$Y_i^{(t)} = \Lambda_t \mu_t + \Lambda_t \gamma_i + \epsilon_i, \quad i = 1, \ldots, n_t, \quad t = 1, 2, \ldots T, \tag{A.4}$$

where length p vector $Y_i^{(t)}$ is the observed phenotype for the i^{th} subject in group t, the group-specific network influence matrices Λ_t's are treated as known from external sources, length p vectors μ_t's denote the group-specific mean signal for the p gene entries, random effects γ_i's follow distribution $N(0, \sigma_1^2 I_p)$, error terms $\epsilon_i \sim N(0, \sigma_0^2 I_p)$, and n_t denotes the sample size for group t, $t = 1, \ldots, T$. We can easily format this linear mixed model into the form of model (2.1) in the main paper. The resulting random effect correlation matrix D_1 has a block-diagonal structure:

$$D_1 = \begin{pmatrix} D_1^{(1)} & 0 & 0 & 0 & 0 & \ldots \\ 0 & D_2^{(1)} & 0 & 0 & 0 & \ldots \\ \vdots & & \ddots & & & \\ 0 & 0 & \ldots & D_{n_1}^{(1)} & 0 & \ldots \\ \vdots & & & & \ddots & \\ 0 & 0 & 0 & 0 & 0 & D_{n_T}^{(T)} \end{pmatrix},$$

where $D_i^{(t)} = \Lambda_t \Lambda_t^\top$ $(i = 1, \ldots, n_t,\ t = 1, \ldots, T)$ are $p \times p$ square blocks corresponding to the i^{th} subject in group t. Based on the special correlation structure of D_1, we can write

Algorithm 4 reREHE Approach for NetGSA.

for $b = 1$ to B **do**

(a) Sample $[nr_s]$ out of n subjects with replacement with sampling rate r_s, where $[x]$ denotes rounding x to the nearest integer. The corresponding observation vectors are $Y_{i'},\ i' = 1, \ldots, [nr_s]$;

(b) For each observation vector $Y_{i'}$ in the subsample, randomly sample $[pr_p]$ out of p gene entries with replacement with sampling rate r_p. The resulted observation vector $Y^*_{i'}$ is of dimension $[pr_p]$ by 1, and the outcome vector $Y^{*(b)}$ for the whole subsample is of length $[nr_s][pr_p]$.

(c) Subset the correlation matrix D_1 accordingly as $D_1^{*(b)}$: first only keep blocks $D_1^{(i')}$ corresponding to the subsampled observations; then for each $D_1^{(i')}$, only keep the submatrix with rows and columns corresponding to the subsampled gene entries.

(d) Compute variance component estimations $(\tilde{\sigma}^{2(b)}_{0,re}, \tilde{\sigma}^{2(b)}_{1,re})$ with $Y^{*(b)}$ and $D_1^{*(b)}$ based on the REHE approach.

end for

Compute estimates of variance components as $\tilde{\sigma}^2_{k,re} = \frac{1}{B}\sum_{b=1}^{B} \tilde{\sigma}^{2(b)}_{k,re}, \quad k = 0,\ 1.$

out the expression for computing the REHE variance component estimates as:

$$\left(\tilde{\sigma}_0^2,\ \tilde{\sigma}_1^2\right) = \arg \min_{(\sigma_0^2 \geq 0,\ \sigma_1^2 \geq 0)} \left(\tilde{Y} - \tilde{X}\boldsymbol{\sigma^2}\right)^\top \left(\tilde{Y} - \tilde{X}\boldsymbol{\sigma^2}\right),$$

where

$$\tilde{Y} = \begin{pmatrix} \mathrm{vec}\left(Y_1^{(1)} Y_1^{(1)\top}\right) \\ \mathrm{vec}\left(Y_2^{(1)} Y_2^{(1)\top}\right) \\ \cdots \\ \mathrm{vec}\left(Y_1^{(2)} Y_1^{(2)\top}\right) \\ \cdots \\ \mathrm{vec}\left(Y_{n_T}^{(T)} Y_{n_T}^{(T)\top}\right) \end{pmatrix}, \quad \tilde{X} = \begin{pmatrix} \mathrm{vec}\left(D_1^{(1)}\right) & \mathrm{vec}\left(\mathbf{I}_p\right) \\ \mathrm{vec}\left(D_2^{(1)}\right) & \mathrm{vec}\left(\mathbf{I}_p\right) \\ \cdots & \\ \mathrm{vec}\left(D_1^{(2)}\right) & \mathrm{vec}\left(\mathbf{I}_p\right) \\ \cdots & \\ \mathrm{vec}\left(D_{n_T}^{T}\right) & \mathrm{vec}\left(\mathbf{I}_p\right) \end{pmatrix}, \quad \boldsymbol{\sigma^2} = \begin{pmatrix} \sigma_0^2 \\ \sigma_1^2 \end{pmatrix}.$$

When applying the reREHE approach to NetGSA, in addition to subsampling the n observations, we can sample the p gene entries. The procedure is described in Algorithm 4.

A.3 Additional Simulation Results

A.3.1 Simulation Results on Matrix Sparsification

Correlation matrix sparsification accelerates confidence interval construction with REHE (Supplementary Note A.2.2), and can also benefit computation with REML [75]. We conducted a simulation study to apply sparsification to both REHE and REML for inference of variance components. We used the same HCHS/SOL design setting as had been used by the simulation study in the main paper Section 4. We compared their performances in terms of computation time, root mean squared error (RMSE) of variance component estimates, and confidence interval coverage and width. *sREML* refers to REML with the sparsified kinship correlation matrix. *sREHE Wald* and *sREHE Quantile* refers to, respectively, REHE Wald- and quantile-type confidence intervals based on sparsified kinship matrix. The REHE point estimation is still based on the original dense kinship matrix, because it is fast and more accurate than that obtained with sparsification.

Both REHE and REML inferences based on the sparse kinship matrix are substantially faster than those based on the original kinship matrix (Figure A.1). We noticed that the sparsification procedure itself is computationally demanding. This may make sparsification less appealing when there are multiple dense correlation matrices to process, or when the sample size is not large enough. However, as the sample size grows, inference with REML becomes so slow that it can still be worthwhile to use matrix sparsification. At sample size n =12,000, REHE inference with one sparsified correlation matrix (one random effect) reduces the REML computation with a dense correlation matrix by 50%.

The RMSE of point estimates by sREML is always larger than REML, and sometimes much larger than the RMSE of REHE estimates (Figure A.2). The coverage of confidence intervals by methods based on sparsification is not as stable: sREHE shows conservative coverage, and sREML has coverage lower than 0.75 at sample size n =3,000 (Figure A.3). This illustrates that sparsification speeds up the computation but at the cost of inference accuracy. For the same method, confidence interval width is generally wider with sparsification than without (Figure A.4). We found sREML and REHE yield comparable confidence interval width.

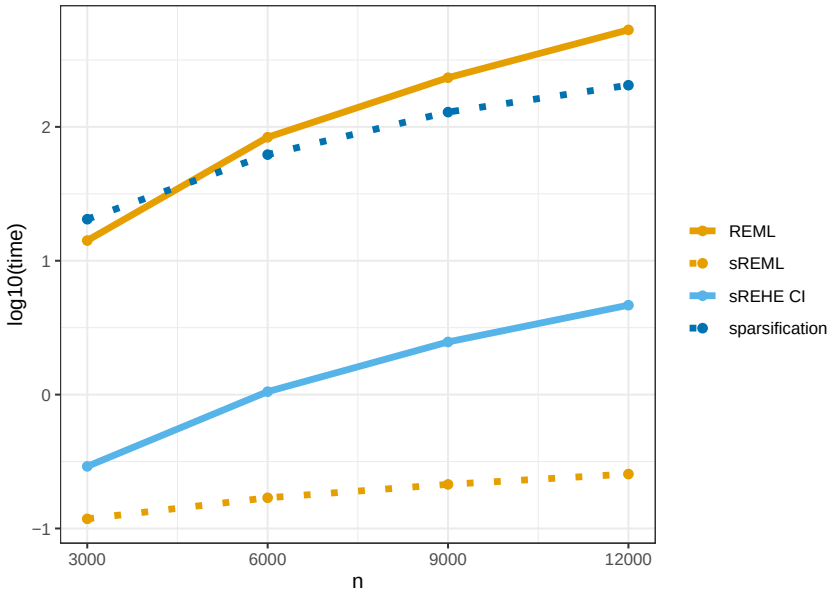

Figure A.1: Results of the simulation study on matrix sparsification: Computation time (in seconds, $\log_{10}$ scale) for REML and REHE based on matrix sparsification, benchmarked with computation time for REML. We separately illustrate time for sparsifying the kinship matrix (sparsification), time for fitting the model with REML after obtaining the sparsified kinship matrix (sREML), and time for REHE confidence interval construction based on sparsified kinship matrix (sREHE CI).

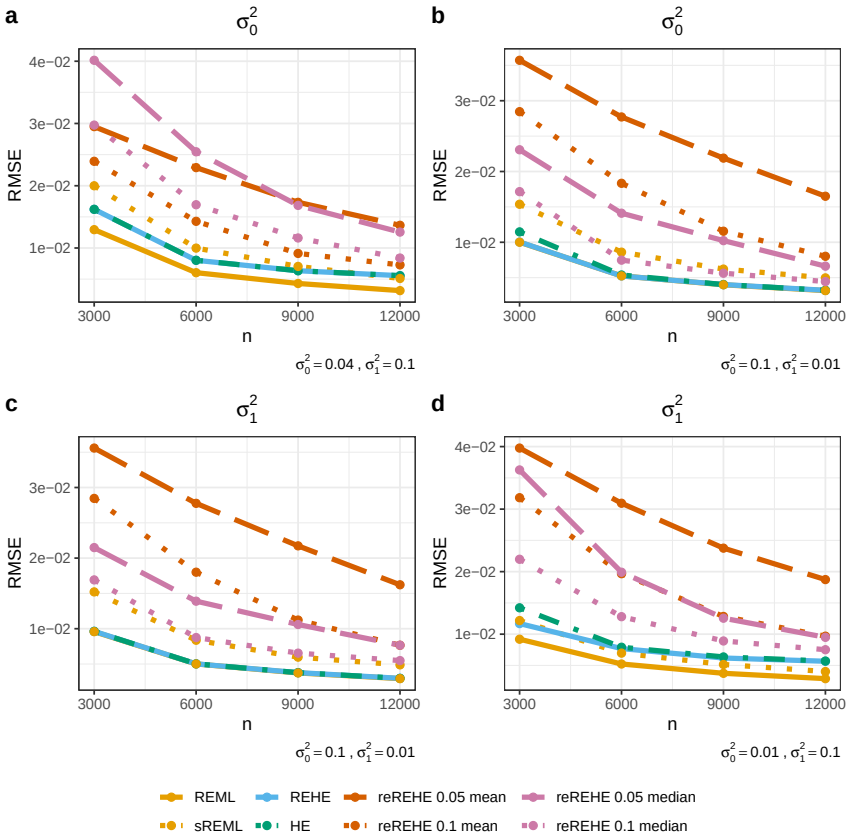

Figure A.2: Results of the simulation study on matrix sparsification and on different reREHE summary functions: RMSE of point estimates by REML, REML with matrix sparsification (sREML), REHE, HE, and different implementations of reREHE: reREHE based on subsampling rate 0.05 and mean summary function (reREHE 0.05 mean), reREHE based on subsampling rate 0.1 and mean summary function (reREHE 0.1 mean), reREHE based on subsampling rate 0.05 and median summary function (reREHE 0.05 median), reREHE based on subsampling rate 0.1 and median summary function (reREHE 0.1 median). **a** - RMSE for σ_0^2 estimates, with true values $\sigma_0^2 = 0.04$, $\sigma_1^2 = 0.1$. **b** - RMSE for σ_0^2 estimates, with true values $\sigma_0^2 = 0.1$, $\sigma_1^2 = 0.01$. **c** - RMSE for σ_1^2 estimates, with true values $\sigma_0^2 = 0.1$, $\sigma_1^2 = 0.01$. **d** - RMSE for σ_1^2 estimates, with true values $\sigma_0^2 = 0.01$, $\sigma_1^2 = 0.1$.

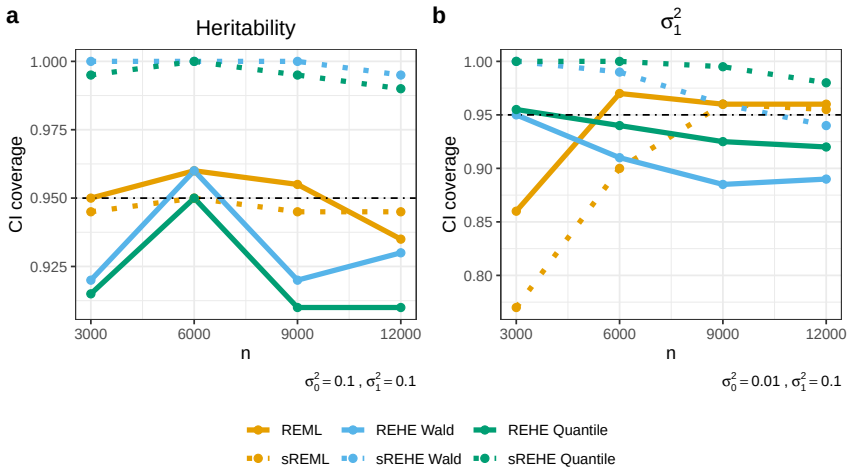

Figure A.3: Results of the simulation study on matrix sparsification: confidence interval coverage with 200 replicates (Monte Carlo error approximately 0.03). We present results for confidence intervals constructed based on REML, REML with matrix sparsification (sREML), Wald-type intervals with REHE (REHE Wald), Wald-type intervals with REHE and matrix sparsification (sREHE Wald), quantile-type intervals with REHE (REHE Quantile), and quantile-type intervals with REHE and matrix sparsification (sREHE Quantile). **a** - Confidence interval coverage for heritability, with true values $\sigma_0^2 = 0.1$, $\sigma_1^2 = 0.1$. **b** - Confidence interval coverage for σ_1^2, with true values $\sigma_0^2 = 0.01$, $\sigma_1^2 = 0.1$.

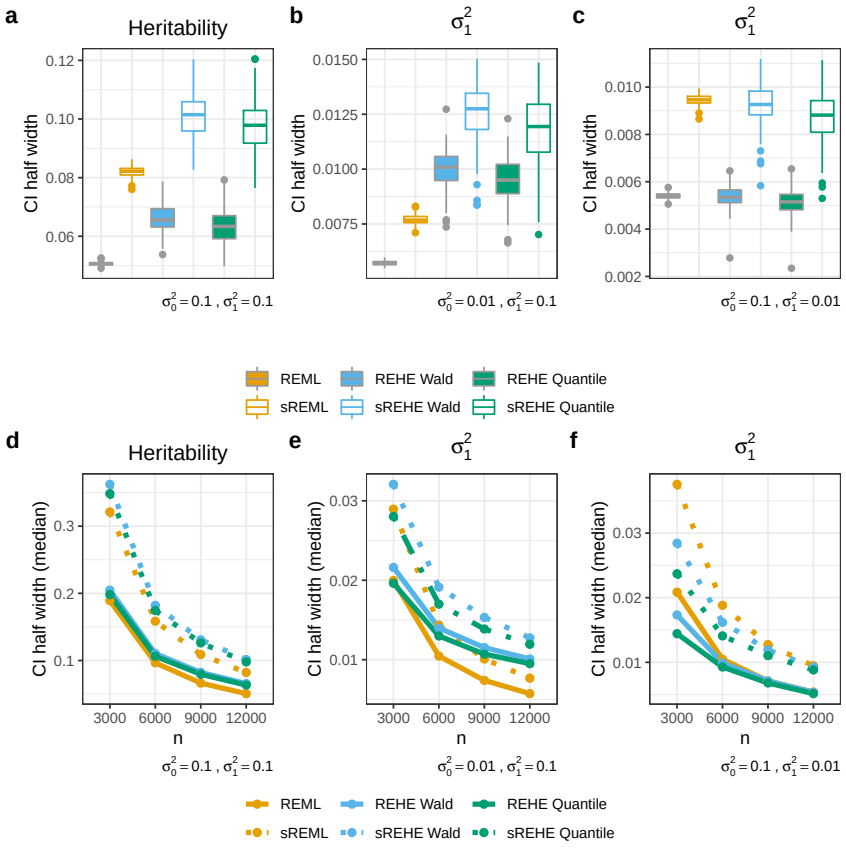

Figure A.4: Results of the simulation study on matrix sparsification: boxplots of confidence interval half width, and line charts of median confidence interval half width against sample sizes. The boxplots follow standard Tukey representations, where center line represents median, box limits represent lower and upper quartiles, whiskers represent 1.5x interquartile range, and points represent outliers. We present results for confidence intervals constructed based on REML, REML with matrix sparsification (sREML), Wald-type intervals with REHE (REHE Wald), Wald-type intervals with REHE and matrix sparsification (sREHE Wald), quantile-type intervals with REHE (REHE Quantile), and quantile-type intervals with REHE and matrix sparsification (sREHE Quantile). The targeted parameter of the confidence interval is indicated by the title of each sub-plot. True values of variance components under each scenario ware presented on the bottom-right corner of each sub-plot.

A.3.2 Simulation Results on Different reREHE Summary Function

The reREHE approach can use different summary functions, such as mean and median functions, to summarize results of repeated subsamples. We implemented median based reREHE for comparison of estimation accuracy against mean based reREHE. We denote mean based reREHE results by *reREHE 0.05 mean* and *reREHE 0.1 mean*, and denote median based reREHE results by *reREHE 0.05 median* and *reREHE 0.1 median*, for subsampling rates 0.05 and 0.1 respectively. We used the same simulation setting as in Supplementary Note A.3.1.

As illustrated by Figure A.2, when the true variance components are very different (taking on values of 0.1 and 0.01), median based reREHE has smaller RMSE than mean based reREHE; otherwise mean based reREHE has smaller RMSE. Higher subsampling rate greatly improves estimation accuracy for both reREHE approaches. In general, reREHE approaches result in less accurate but acceptable estimates compared to REML and REHE results.

Mean based reREHE provides positive estimates with high probability, whereas median based reREHE is more likely to yield a zero estimate in practise. The trade-off among accuracy, interpretability and computational efficiency brought by reREHE will be left at user's choice.

A.3.3 Simulation Results on Heritability

We conducted additional simulation studies based on HCHS/SOL design, using three different artificial kinship matrices to compare REHE and reREHE against REML and HE. In the first two settings, we generated block-diagonal kinship matrix D_1 with 3×3 blocks D_b, such that D_1 had the form

$$D_1 = \begin{pmatrix} D_b & 0 & 0 & \dots \\ 0 & D_b & 0 & \dots \\ 0 & 0 & D_b & \\ \vdots & & & \ddots \end{pmatrix}.$$

Setting 1 represents a population with high correlation within groups, whereas setting 2 represents a population where subjects are remotely related even within groups. For setting 3, we generated a dense kinship matrix which is approximately block-diagonal, and then generated random values from Unif$[-0.001, 0.001]$ to fill the off-diagonal zero entries in D_1. Setting 3 mimics the dense kinship matrix in real data applications, which may render estimation and inference more difficult. We set the values of D_b in each setting as:

Setting 1:
$$D_b = \begin{pmatrix} 1 & 0.8 & 0.2 \\ 0.8 & 1 & 0.4 \\ 0.2 & 0.4 & 1 \end{pmatrix};$$

Setting 2:
$$D_b = \begin{pmatrix} 1 & 0.05 & 0.05 \\ 0.05 & 1 & 0.1 \\ 0.05 & 0.1 & 1 \end{pmatrix};$$

Setting 3:
$$D_b = \begin{pmatrix} 1 & 0.1 & 0.05 \\ 0.1 & 1 & 0.3 \\ 0.05 & 0.3 & 1 \end{pmatrix}.$$

Based on each of the three kinship matrices, we ran 200 replicates for each combination of $(\sigma_0^2,\ \sigma_1^2) \in \{(0.1,\ 0.1),\ (0.04,\ 0.1),\ (0.1, 0.04),\ (0.01,\ 0.1),\ (0.1,\ 0.01)\}$ and $n \in \{3,000,\ 6,000,\ 9,000,\ 12,000\}$. We sparsified the kinship matrices when constructing confidence intervals for REHE and REML.

Computation time of each method is very similar to those presented in the main paper and in Supplementary Note A.3.1 (Figure A.1). Although the kinship matrix is already sparse and block-diagonal under setting 1 and 2, the sparsification algorithm is not able to process the matrices quickly. Sparsification procedures in these settings take similar amount of computation time as for sparsifying dense kinship matrices of the same size. This indicates room of improvement for the sparsification algorithm. We recommend the users to skip the sparsification procedure if the available correlation matrices are approximately sparse and block-diagonal such that computing their inverses are computationally efficient.

Figure A.5 illustrates the RMSE. reREHE has similar performance as those described in Supplementary Note A.3.1. The RMSE of sREML is always indistinguishable from that of REML. REHE estimates are close to REML estimates, and their RMSE often has negligible difference (e.g. Figure A.5-b, A.5-e, A.5-h). When the true variance component values are different, improvement of REHE over HE is more pronounced than those illustrated in

the main paper and Supplementary Note A.3.1 (Figure A.5-e, A.5-f). Such improvement in RMSE occurs when REHE corrects negative values in HE estimates. We counted the proportion of simulation replicates where HE gave negative estimates (before zero truncation) for each setting, and observed a proportion as high as 45% (n =3,000, D_1 under setting 2, $(\sigma_0^2,\ \sigma_1^2) \in \{(0.01, 0.1),\ (0.1, 0.01)\}$). Even at the sample size n =12,000, HE still produces negative estimates in 30% of the replicates. Even with less extreme variance component values $(\sigma_0^2,\ \sigma_1^2) \in \{(0.04, 0.1),\ (0.1, 0.04)\}$, we still observed 15% of the HE estimates being negative at the sample size n =3,000. In all of these cases, we obtained substantially lower RMSE by REHE than by HE (Figure A.5-e, A.5-f).

Confidence intervals have close to nominal coverage for all methods under $(\sigma_0^2, \sigma_1^2) = (0.1, 0.1)$, but show more volatility when the true variance component values are unequal (Figure A.6). Taking into consideration the Monte Carlo error of 0.03 based on 200 replicates, REHE quantile-type confidence intervals always have acceptable coverage around 0.95. The coverage of REHE Wald-type confidence intervals is generally close to 0.95, but slightly off under some scenarios (Figure A.6-c, A.6-f, A.6-i). This suggests that REHE quantile-type interval is more robust than Wald-type interval. REHE confidence intervals constructed from sparsified matrices have similar patterns of coverage. We observed in the main paper and Supplementary Note A.3.1 that the coverage of REML and sREML is poor in some scenarios. Although the coverage improves quickly with increasing sample size in some cases, the under-coverage issue can persist even with large sample sizes (Figure A.6-c, A.6-e, A.6-h). With $(\sigma_0^2, \sigma_1^2) = (0.01, 0.1)$ and kinship matrix structure under setting 2 or 3, the coverage of REML/sREML is lower than 0.7 even at the sample size $n = 12,000$ (Figure A.6-e, A.6-h). In addition, REML may fail to produce a confidence interval due to computational issues, which happens when any variance component estimate reaches zero during the iterative updates. Under setting 2, with $\sigma_0^2 = 0.1$,$\sigma_1^2 = 0.01$, n=3,000, this issue occurs in more than 30% of the simulation replicates. Even at a sample size of n =12,000 we still observed as high as 22% of replicates having problem with REML inference. On the other hand, REHE confidence intervals does not suffer from this issue and can always be obtained.

Figure A.7 illustrates the distribution of confidence interval half width. Overall REHE

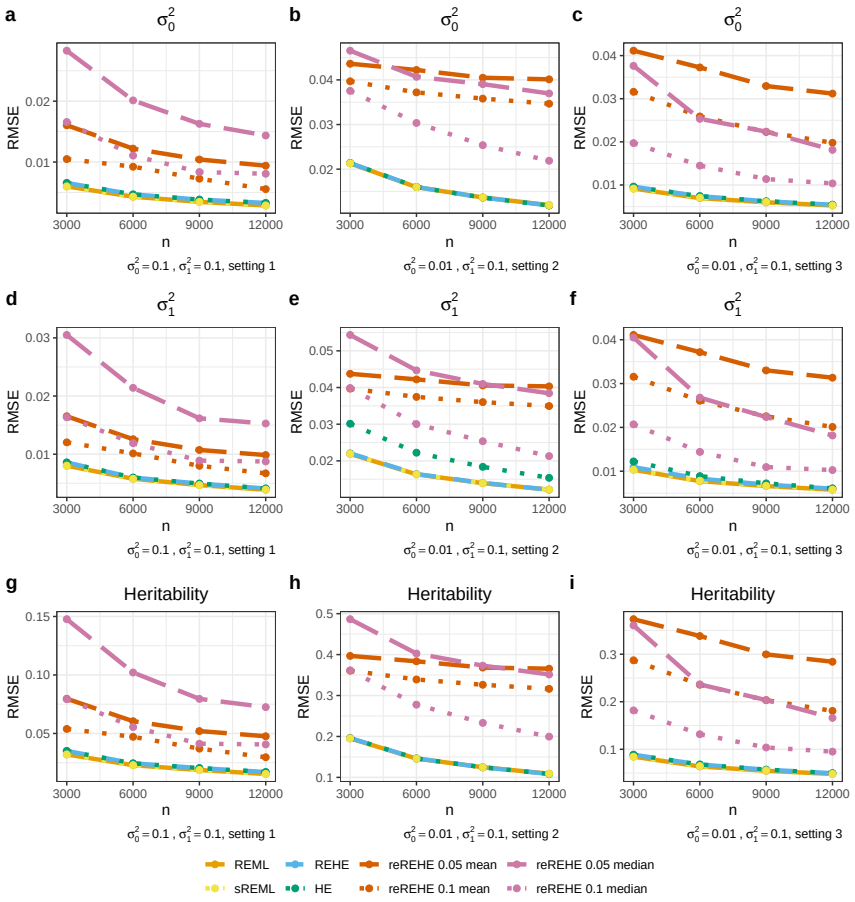

Figure A.5: Results of simulation study on heritability analysis. RMSE of estimates for σ_0^2, σ_1^2 and heritability, based on REML, REML with matrix sparsification (sREML), REHE, HE, and reREHE with different subsampling rates (0.05/0.1) and different summary functions (mean/median). Parameter true values for σ_0^2 and σ_1^2, and kinship matrix structure setting under each scenario are indicated on the bottom-left corner of each sub-plot. The title of each sub-plot indicates the parameter of interest.

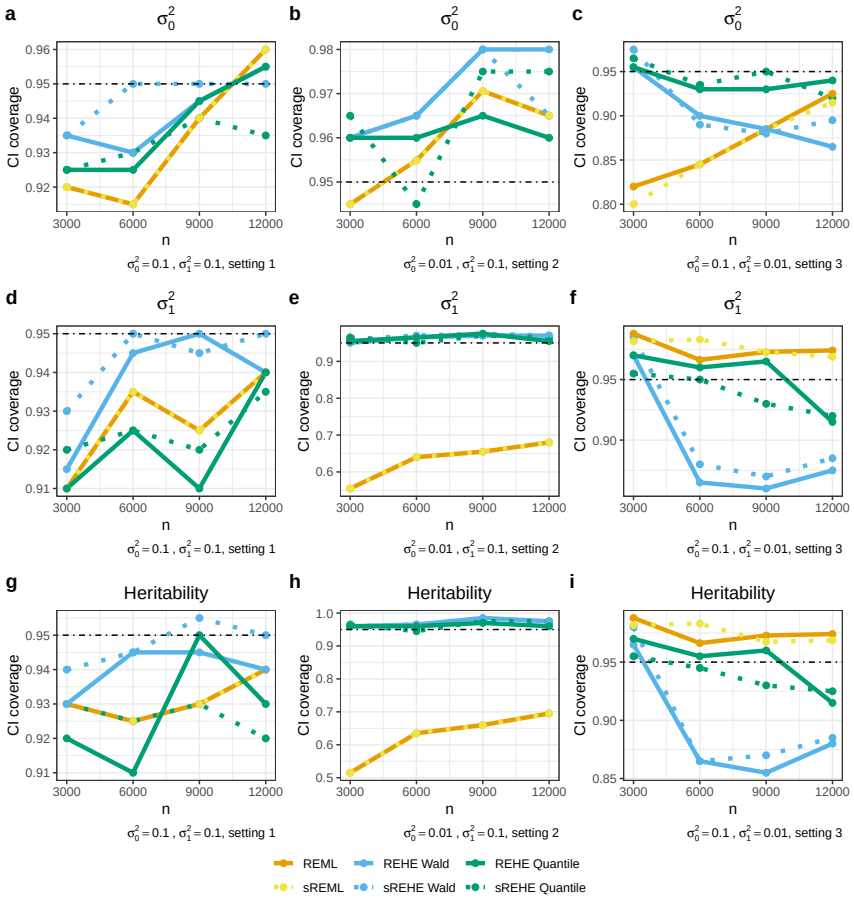

Figure A.6: Results of simulation study on heritability analysis. Confidence interval coverage for σ_0^2, σ_1^2 and heritability with 200 replicates and expected Monte Carlo error of 0.03. We present results for confidence intervals constructed based on REML, REML with matrix sparsification (sREML), Wald-type intervals with REHE (REHE Wald), Wald-type intervals with REHE and matrix sparsification (sREHE Wald), quantile-type intervals with REHE (REHE Quantile), and quantile-type intervals with REHE and matrix sparsification (sREHE Quantile). Parameter true values for σ_0^2 and σ_1^2, and kinship matrix structure setting under each scenario are indicated on the bottom-left corner of each sub-plot. The title of each sub-plot indicates the parameter of interest.

is comparable to REML. REML generally has narrower confidence intervals when true values of variance components are not very different (Figure A.7-a, A.7-d, A.7-g), when they do differ REHE intervals are narrower (Figure A.7-c, A.7-f, A.7-i). As we have seen scenarios with poor coverage for REML and sREML confidence intervals (Figure A.6-e, A.6-h), corresponding interval width has suspiciously skewed distributions (Figure A.7-e, A.7-h).

A.3.4 Simulation Results on Network-Based Pathway Enrichment Analysis

We conducted additional simulation studies with network-based pathway enrichment analysis [147] to benchmark the performance of REHE and reREHE against REML. We designed the simulation study based on the breast cancer data and the dysregulation framework used in [148]. The breast cancer data set has 403 samples in the ER positive group, 117 samples in the ER negative group, with a total of 2,598 gene entries spanning 100 gene pathways. To minimize the difference in gene-gene correlations between the two groups, we randomly split the total number of 520 samples into one subset with 403 samples and another with 117 samples. The first subset of samples was used to estimate an influence matrix Λ_1 for the ER positive group, and the second an influence matrix Λ_2 for the ER negative group [147]. Next, we generated observations according to the linear mixed model (A.4), where $T = 2$, $n_1 = 403$, $n_2 = 117$, and $p =$2,598. We set the mean signal vector to be zero for the ER positive group: $\mu_{1,j} = 0$ for $j = 1, \ldots, p$. For the ER negative group, we set $\mu_{2,j} = \Delta\mu \in \{0.1, 0.2, 0.3\}$ when $j \in S$ for an index set S, otherwise $\mu_{2,j} = 0$. This index set S includes genes selected in the betweeness-type dysregulation design [148]. The true variance component values were set to be $(\sigma_0^2,\ \sigma_1^2) = (0.1, 0.1),\ (0.5, 0.1),\ (0.1, 0.5),\ (0.1, 1),\ (1, 0.1)$. Based on the semi-synthetic data, we performed network-based pathway enrichment analysis to compare the power detecting differential pathway against the true power for each pathway, where true powers are computed according to [147]. We also evaluated the variance component estimates against the true values. Under each setting we ran 200 replicates. All methods were implemented using R package `netgsa` (*v3.1.0*, 147).

Computation with REML is time-demanding, which took around 1.5 hours for the analysis,

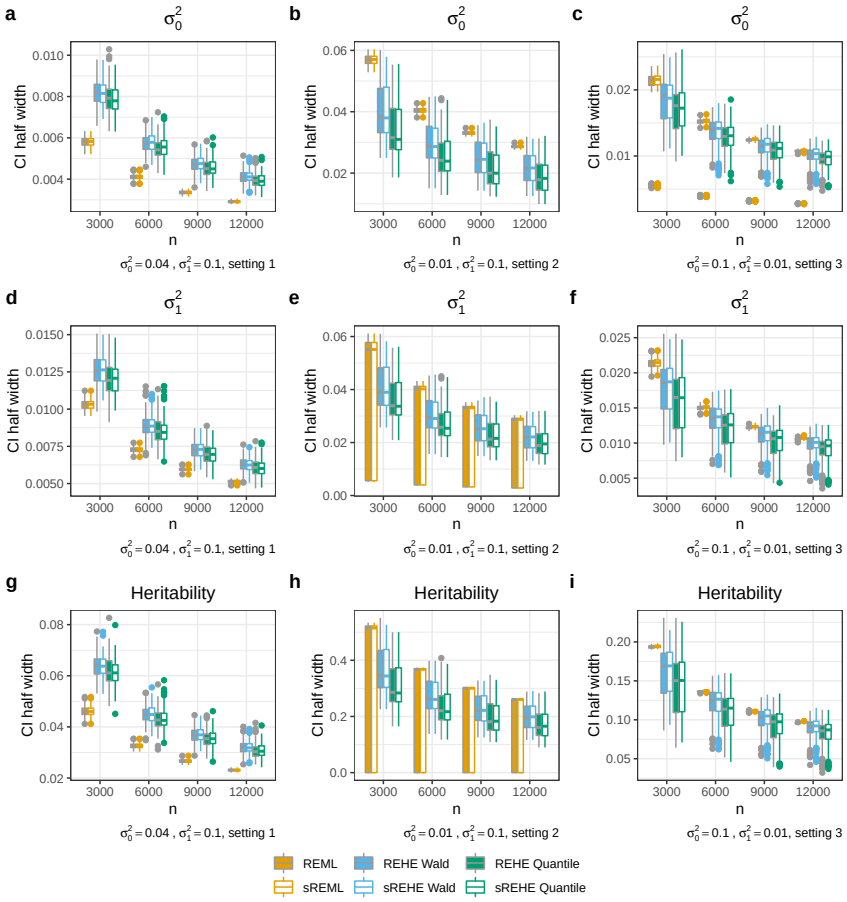

Figure A.7: Results of simulation study on heritability analysis. Boxplots of confidence interval half width for σ_0^2, σ_1^2 and heritability. The boxplots follow standard Tukey representations, where center line represents median, box limits represent lower and upper quartiles, whiskers represent 1.5x interquartile range, and points represent outliers. We present results for confidence intervals constructed based on REML, REML with matrix sparsification (sREML), Wald-type intervals with REHE (REHE Wald), Wald-type intervals with REHE and matrix sparsification (sREHE Wald), quantile-type intervals with REHE (REHE Quantile), and quantile-type intervals with REHE and matrix sparsification (sREHE Quantile). Parameter true values for σ_0^2 and σ_1^2, and kinship matrix structure setting under each scenario are indicated on the bottom-left corner of each sub-plot. The title of each sub-plot indicates the parameter of interest.

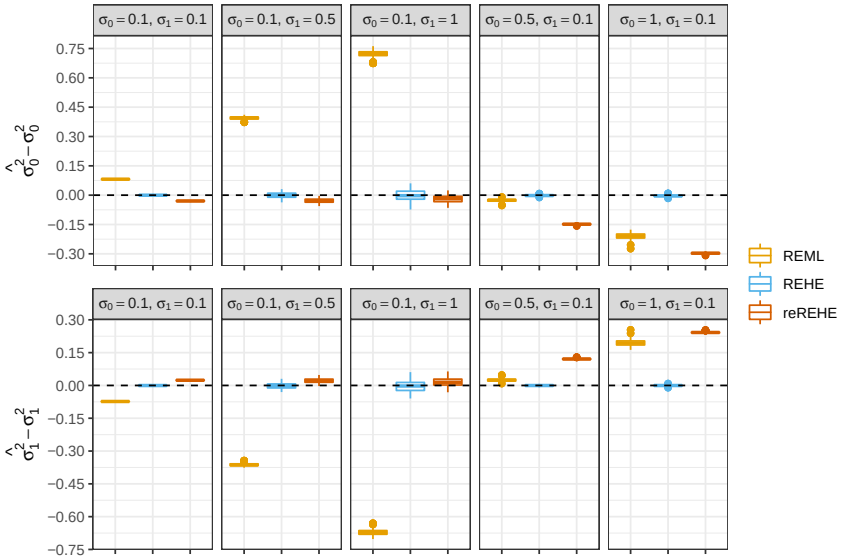

Figure A.8: Results of simulation study on gene-network enrichment analysis. We present results for REML, REHE and reREHE, with boxplots of difference between the variance component estimates and the true values. The boxplots follow standard Tukey representations, where center line represents median, box limits represent lower and upper quartiles, whiskers represent 1.5x interquartile range, and points represent outliers. We present the results based on $\Delta\mu = 0.1$, and sub-plot titles indicate the parameter values of each setting. Results are based on 200 replicates.

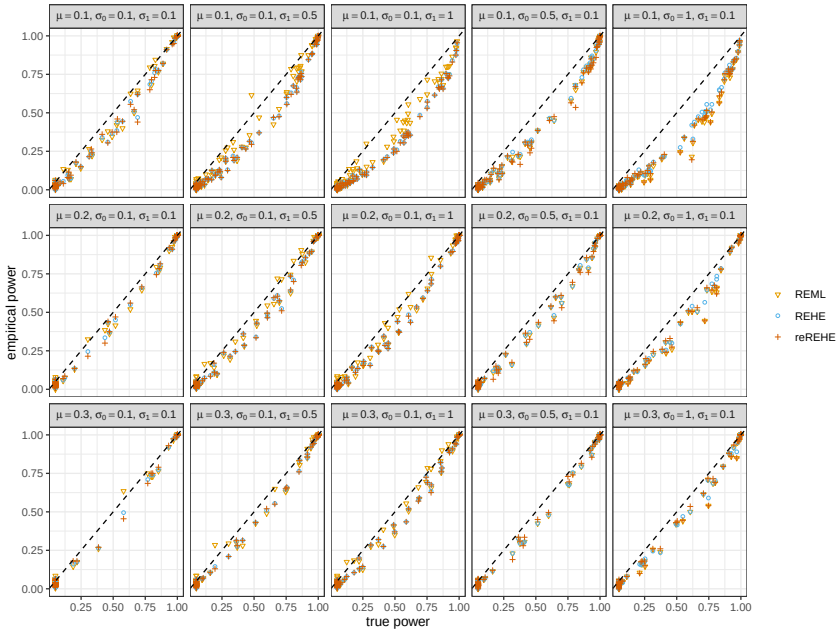

Figure A.9: Simulation study results on gene-network enrichment analysis. We present empirical power against true power of detecting group difference for each pathway, for REML, REHE and reREHE approaches. Results are based on 200 replicates. The sub-plot titles indicate the parameter values of each setting.

compared to only 2 minutes by REHE and reREHE based analysis. In addition to computational gain, REHE provides more accurate variance component estimates than REML (Figure A.8). When the noise variance component $\sigma_0^2 = 0.1$ and the random effect variance component $\sigma_1^2 = 1$, REML estimates are far from truth, but REHE and reREHE estimates are accurate. This suggests that in practice one needs to be cautious about REML estimates if the majority of the outcome variation is attributed to the network random effects. reREHE estimates are comparable to REHE in general, but less accurate when the noise variance component is larger than the random effect variance (Figure A.8).

Although REML provides less accurate variance component estimates than REHE, its empirical power for each pathway is closer to the true power in most settings, except when $(\sigma_0^2,\ \sigma_1^2) = (1, 0.1)$. In this case, the REHE approach has powers closer to the truth (Figure A.9). reREHE empirical powers have similar patterns as those of REHE, but are slightly worse (Figure A.9). In general the empirical powers of all three methods are close to the truth.

We noticed that even when REML fails to provide accurate variance component estimates, its empirical powers are still very close to the true powers. This indicates potential unidentifiability of the model parameters over the likelihood surface. This makes REML difficult to estimate the parameters, but it may still retain sensitivity of tests to group difference based on REML estimates. However, both REHE and reREHE can always yield good empirical powers and at the same time provide good estimates of variance components.

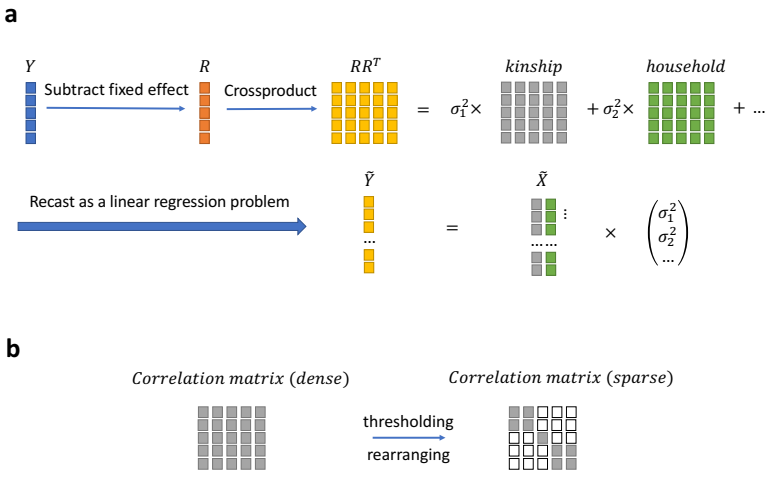

Figure A.10: Illustration of HE/REHE workflow and correlation matrix sparsification. **a** - HE/REHE workflow. Y is the outcome vector and R is the residual vector after subtracting fixed effects from Y. Estimating equations are formed by decomposing the total variation $RR^\top$ into different sources (e.g. kinship, household relatedness) with corresponding variance components σ_k^2's. After vectorizing the total variation $RR^\top$ as $\tilde{Y}$ and the relatedness matrices as $\tilde{X}$, variance components are estimated using ordinary least squares/non-negative least squares. **b** - Illustration of correlation matrix sparsification. Grey area represents non-zero entries, and white area represents zero entries.

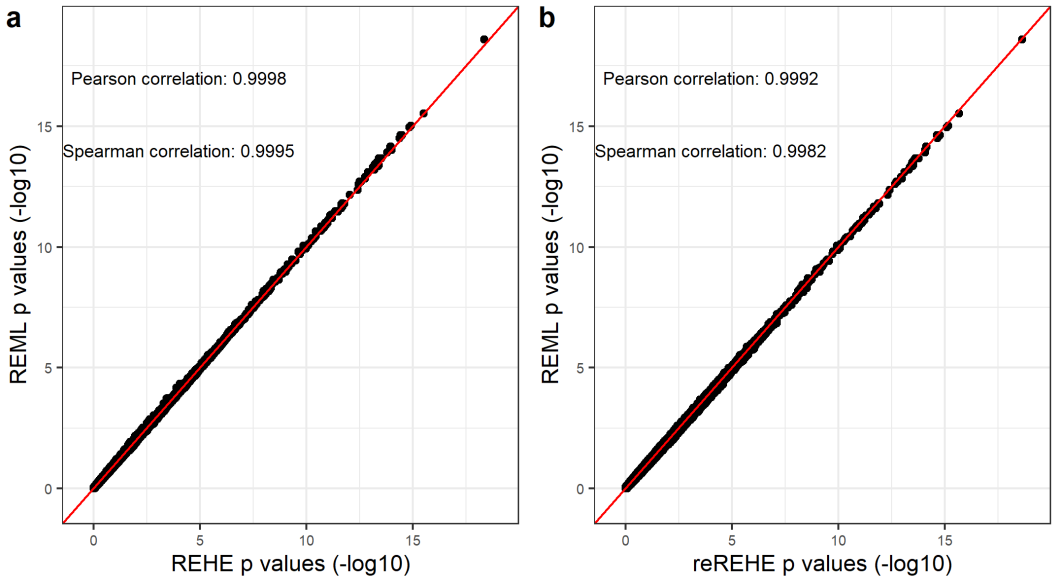

Figure A.11: Results of genome-wide association testing analysis with a HCHS/SOL data set. Association score test p-values ($-\log_{10}$ scale) for all 4,100,028 SNPs based on: **a** - REML against REHE estimated null models; **b** - REML against reREHE estimated null models.

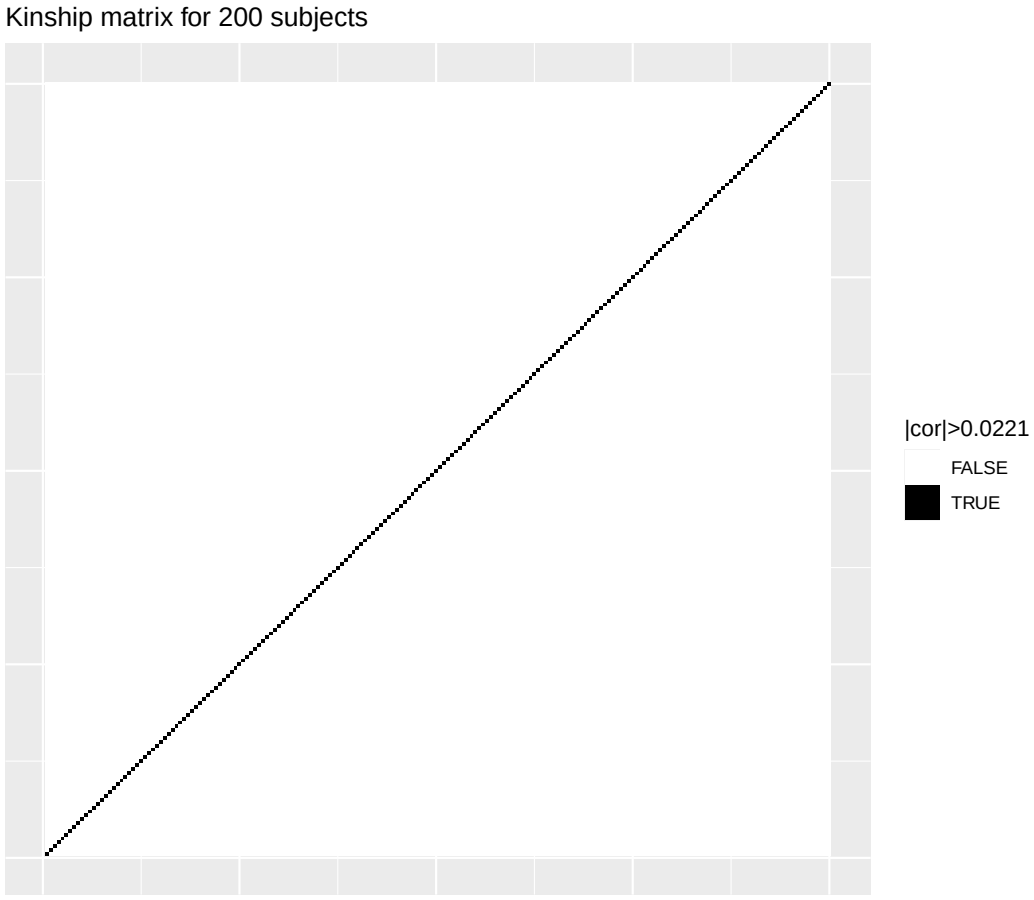

Figure A.12: An illustration of the kinship submatrix corresponding to the first 200 subjects in the HCHS/SOL data set. The entries with absolute values larger than $2^{-5.5} = 0.0221$ are marked as black.

Table A.1: Bias and empirical standard errors of σ_0^2 (error variance) points estimates for all methods, based on 200 replicates for each simulation setting. Setting 1, 2, 3 are described in Appendix A.3.3. Setting 4 corresponds to the main text simulation setting.

n	Setting	σ_0^2	σ_1^2	REHE	HE	REML	sREML	reREHE 0.05	reREHE 0.1	reREHE 0.05 median	reREHE 0.1 median
3000	1	0.10	0.10	-7.940e-04 (6.520e-03)	-7.940e-04 (6.520e-03)	-8.057e-04 (5.943e-03)	-8.057e-04 (5.943e-03)	1.015e-02 (1.243e-02)	2.208e-03 (1.243e-02)	1.492e-02 (1.243e-02)	5.031e-03 (1.243e-02)
		0.04	0.10	-4.415e-04 (4.069e-03)	-4.415e-04 (4.069e-03)	-4.053e-04 (3.076e-03)	-4.053e-04 (3.076e-03)	2.360e-02 (8.322e-03)	8.959e-03 (8.322e-03)	1.663e-02 (8.322e-03)	5.563e-03 (8.322e-03)
		0.10	0.04	-6.702e-04 (5.209e-03)	-6.702e-04 (5.209e-03)	-7.006e-04 (5.129e-03)	-7.006e-04 (5.129e-03)	-9.849e-03 (8.679e-03)	-5.650e-03 (8.679e-03)	2.356e-03 (8.679e-03)	1.212e-03 (8.679e-03)
		0.01	0.10	-2.652e-04 (2.985e-03)	-2.652e-04 (2.985e-03)	-1.955e-04 (1.607e-03)	-1.955e-04 (1.607e-03)	3.052e-02 (6.084e-03)	1.467e-02 (6.084e-03)	1.661e-02 (6.084e-03)	5.942e-03 (6.084e-03)
		0.10	0.01	-6.168e-04 (4.609e-03)	-6.082e-04 (4.624e-03)	-6.370e-04 (4.624e-03)	-6.370e-04 (4.624e-03)	-2.049e-02 (6.610e-03)	-1.182e-02 (6.610e-03)	-8.541e-03 (6.610e-03)	-4.579e-03 (6.610e-03)
	2	0.10	0.10	-5.750e-03 (5.027e-02)	-5.170e-03 (5.160e-02)	-5.897e-03 (5.045e-02)	-5.897e-03 (5.045e-02)	2.140e-03 (1.408e-02)	4.179e-04 (1.408e-02)	-4.104e-03 (1.408e-02)	-5.277e-03 (1.408e-02)
		0.04	0.10	-1.481e-03 (3.353e-02)	-1.481e-03 (3.353e-02)	-1.457e-03 (3.336e-02)	-1.457e-03 (3.336e-02)	2.931e-02 (9.890e-03)	2.593e-02 (9.890e-03)	1.687e-02 (9.890e-03)	9.437e-03 (9.890e-03)
		0.10	0.04	-6.573e-03 (3.184e-02)	-3.973e-03 (3.645e-02)	-6.611e-03 (3.178e-02)	-6.611e-03 (3.178e-02)	-2.637e-02 (9.828e-03)	-2.538e-02 (9.828e-03)	-2.324e-02 (9.828e-03)	-1.734e-02 (9.828e-03)
		0.01	0.10	5.438e-03 (2.068e-02)	5.438e-03 (2.068e-02)	5.324e-03 (2.062e-02)	5.324e-03 (2.062e-02)	4.293e-02 (7.754e-03)	3.877e-02 (7.754e-03)	2.773e-02 (7.754e-03)	1.833e-02 (7.754e-03)
		0.10	0.01	-9.711e-03 (1.950e-02)	-3.144e-03 (2.843e-02)	-9.725e-03 (1.947e-02)	-9.725e-03 (1.947e-02)	-4.065e-02 (7.692e-03)	-3.836e-02 (7.692e-03)	-3.380e-02 (7.692e-03)	-2.374e-02 (7.692e-03)
	3	0.10	0.10	-2.498e-03 (2.008e-02)	-2.498e-03 (2.008e-02)	-2.525e-03 (1.999e-02)	-2.512e-03 (2.000e-02)	5.634e-03 (1.400e-02)	2.425e-03 (1.400e-02)	1.097e-02 (1.400e-02)	4.982e-03 (1.400e-02)
		0.04	0.10	-1.678e-03 (1.397e-02)	-1.678e-03 (1.397e-02)	-1.666e-03 (1.372e-02)	-1.665e-03 (1.374e-02)	2.886e-02 (9.796e-03)	2.122e-02 (9.796e-03)	1.971e-02 (9.796e-03)	6.772e-03 (9.796e-03)
		0.10	0.04	-1.824e-03 (1.420e-02)	-1.820e-03 (1.421e-02)	-1.860e-03 (1.424e-02)	-1.847e-03 (1.423e-02)	-2.101e-02 (9.794e-03)	-1.781e-02 (9.794e-03)	-7.523e-03 (9.794e-03)	-2.525e-03 (9.794e-03)
		0.01	0.10	-2.049e-04 (9.602e-03)	-2.049e-04 (9.602e-03)	-1.697e-04 (9.166e-03)	-1.626e-04 (9.147e-03)	4.043e-02 (7.639e-03)	3.075e-02 (7.639e-03)	2.500e-02 (7.639e-03)	9.446e-03 (7.639e-03)
		0.10	0.01	-2.635e-03 (9.464e-03)	-1.481e-03 (1.132e-02)	-2.668e-03 (9.511e-03)	-2.657e-03 (9.475e-03)	-3.436e-02 (7.669e-03)	-2.818e-02 (7.669e-03)	-1.855e-02 (7.669e-03)	-1.043e-02 (7.669e-03)
	4	0.10	0.10	-2.017e-04 (2.216e-02)	-2.017e-04 (2.216e-02)	-3.118e-04 (1.981e-02)	-3.777e-04 (3.199e-02)	3.882e-03 (1.717e-02)	3.296e-03 (1.717e-02)	7.154e-03 (1.717e-02)	3.839e-03 (1.717e-02)
		0.04	0.10	-3.992e-05 (1.621e-02)	-3.992e-05 (1.621e-02)	-1.245e-04 (1.295e-02)	-1.606e-04 (2.000e-02)	2.704e-02 (1.179e-02)	2.162e-02 (1.179e-02)	1.377e-02 (1.179e-02)	6.724e-03 (1.179e-02)
		0.10	0.04	-2.425e-04 (1.496e-02)	-2.425e-04 (1.496e-02)	-2.750e-04 (1.442e-02)	-6.685e-04 (2.318e-02)	-2.164e-02 (1.205e-02)	-1.696e-02 (1.205e-02)	-5.654e-03 (1.205e-02)	-3.007e-03 (1.205e-02)
		0.01	0.10	1.900e-03 (1.055e-02)	1.900e-03 (1.055e-02)	6.046e-04 (8.216e-03)	1.813e-03 (1.120e-02)	3.875e-02 (9.062e-03)	3.101e-02 (9.062e-03)	1.882e-02 (9.062e-03)	1.023e-02 (9.062e-03)
		0.10	0.01	-1.311e-03 (9.949e-03)	-2.629e-04 (1.147e-02)	-1.319e-03 (9.919e-03)	-3.875e-03 (1.488e-02)	-3.450e-02 (9.266e-03)	-2.734e-02 (9.266e-03)	-1.586e-02 (9.266e-03)	-1.047e-02 (9.266e-03)
6000	1	0.10	0.10	-6.119e-04 (4.646e-03)	-6.119e-04 (4.646e-03)	-5.418e-04 (4.271e-03)	-5.418e-04 (4.271e-03)	5.067e-03 (1.110e-02)	1.173e-03 (1.110e-02)	9.424e-03 (1.110e-02)	1.934e-03 (1.110e-02)
		0.04	0.10	-3.466e-04 (2.857e-03)	-3.466e-04 (2.857e-03)	-2.526e-04 (2.213e-03)	-2.526e-04 (2.213e-03)	1.528e-02 (7.266e-03)	4.746e-03 (7.266e-03)	9.771e-03 (7.266e-03)	2.417e-03 (7.266e-03)
		0.10	0.04	-5.101e-04 (3.747e-03)	-5.101e-04 (3.747e-03)	-4.975e-04 (3.689e-03)	-4.975e-04 (3.689e-03)	-8.294e-03 (7.568e-03)	-2.569e-03 (7.568e-03)	3.002e-03 (7.568e-03)	9.125e-05 (7.568e-03)
		0.01	0.10	-2.139e-04 (2.061e-03)	-2.139e-04 (2.061e-03)	-1.071e-04 (1.160e-03)	-1.071e-04 (1.160e-03)	2.152e-02 (5.082e-03)	9.862e-03 (5.082e-03)	1.014e-02 (5.082e-03)	2.513e-03 (5.082e-03)
		0.10	0.01	-4.592e-04 (3.343e-03)	-4.592e-04 (3.343e-03)	-4.658e-04 (3.344e-03)	-4.658e-04 (3.344e-03)	-1.630e-02 (5.449e-03)	-7.609e-03 (5.449e-03)	-4.451e-03 (5.449e-03)	-2.486e-03 (5.449e-03)
	2	0.10	0.10	-4.069e-03 (3.886e-02)	-4.042e-03 (3.894e-02)	-4.124e-03 (3.884e-02)	-4.124e-03 (3.884e-02)	1.739e-03 (1.392e-02)	-6.164e-04 (1.392e-02)	-1.459e-03 (1.392e-02)	3.787e-04 (1.392e-02)
		0.04	0.10	-1.900e-03 (2.591e-02)	-1.900e-03 (2.591e-02)	-1.946e-03 (2.586e-02)	-1.946e-03 (2.586e-02)	2.826e-02 (9.685e-03)	2.383e-02 (9.685e-03)	1.286e-02 (9.685e-03)	8.152e-03 (9.685e-03)
		0.10	0.04	-3.851e-03 (2.523e-02)	-2.913e-03 (2.716e-02)	-3.904e-03 (2.529e-02)	-3.904e-03 (2.529e-02)	-2.571e-02 (9.819e-03)	-2.464e-02 (9.819e-03)	-1.539e-02 (9.819e-03)	-9.358e-03 (9.819e-03)
		0.01	0.10	3.016e-03 (1.572e-02)	3.016e-03 (1.572e-02)	2.921e-03 (1.571e-02)	2.921e-03 (1.571e-02)	4.154e-02 (7.571e-03)	3.614e-02 (7.571e-03)	2.102e-02 (7.571e-03)	1.358e-02 (7.571e-03)
		0.10	0.01	-6.134e-03 (1.580e-02)	-2.319e-03 (2.122e-02)	-6.155e-03 (1.587e-02)	-6.155e-03 (1.587e-02)	-3.944e-02 (7.691e-03)	-3.665e-02 (7.691e-03)	-2.384e-02 (7.691e-03)	-1.641e-02 (7.691e-03)
	3	0.10	0.10	-1.611e-03 (1.481e-02)	-1.611e-03 (1.481e-02)	-1.641e-03 (1.464e-02)	-1.619e-03 (1.464e-02)	4.685e-03 (1.370e-02)	9.799e-04 (1.370e-02)	1.048e-02 (1.370e-02)	4.470e-03 (1.370e-02)
		0.04	0.10	-1.053e-03 (1.028e-02)	-1.053e-03 (1.028e-02)	-1.067e-03 (9.973e-03)	-1.044e-03 (9.975e-03)	2.589e-02 (9.622e-03)	1.660e-02 (9.622e-03)	1.294e-02 (9.622e-03)	4.742e-03 (9.622e-03)
		0.10	0.04	-1.202e-03 (1.051e-02)	-1.202e-03 (1.051e-02)	-1.224e-03 (1.050e-02)	-1.216e-03 (1.050e-02)	-1.943e-02 (9.382e-03)	-1.523e-02 (9.382e-03)	-1.121e-03 (9.382e-03)	6.551e-04 (9.382e-03)
		0.01	0.10	-3.717e-04 (7.406e-03)	-3.717e-04 (7.406e-03)	-3.790e-04 (6.956e-03)	-4.075e-04 (7.051e-03)	3.650e-02 (7.456e-03)	2.497e-02 (7.456e-03)	1.547e-02 (7.456e-03)	6.156e-03 (7.456e-03)
		0.10	0.01	-1.491e-03 (7.459e-03)	-9.980e-04 (8.387e-03)	-1.505e-03 (7.491e-03)	-1.516e-03 (7.434e-03)	-3.152e-02 (7.189e-03)	-2.391e-02 (7.189e-03)	-1.071e-02 (7.189e-03)	-5.589e-03 (7.189e-03)
	4	0.10	0.10	7.423e-04 (1.062e-02)	7.423e-04 (1.062e-02)	5.510e-04 (9.290e-03)	-1.009e-03 (1.594e-02)	2.753e-03 (1.364e-02)	6.542e-04 (1.364e-02)	7.379e-03 (1.364e-02)	4.698e-03 (1.364e-02)
		0.04	0.10	6.193e-04 (8.002e-03)	6.193e-04 (8.002e-03)	3.662e-04 (6.025e-03)	-6.089e-04 (9.968e-03)	2.078e-02 (9.666e-03)	1.157e-02 (9.666e-03)	8.163e-03 (9.666e-03)	4.023e-03 (9.666e-03)
		0.10	0.04	4.200e-04 (7.034e-03)	4.200e-04 (7.034e-03)	3.713e-04 (6.773e-03)	-7.222e-04 (1.195e-02)	-1.687e-02 (9.077e-03)	-1.044e-02 (9.077e-03)	1.233e-04 (9.077e-03)	2.475e-03 (9.077e-03)
		0.01	0.10	6.903e-04 (6.532e-03)	6.903e-04 (6.532e-03)	2.543e-04 (4.209e-03)	-1.790e-04 (6.454e-03)	3.005e-02 (7.326e-03)	1.847e-02 (7.326e-03)	9.436e-03 (7.326e-03)	4.386e-03 (7.326e-03)
		0.10	0.01	1.945e-04 (5.235e-03)	2.588e-04 (5.372e-03)	1.930e-04 (5.244e-03)	-1.169e-03 (8.524e-03)	-2.690e-02 (6.627e-03)	-1.730e-02 (6.627e-03)	-8.580e-03 (6.627e-03)	-3.163e-03 (6.627e-03)
9000	1	0.10	0.10	-4.857e-04 (3.799e-03)	-4.857e-04 (3.799e-03)	-4.439e-04 (3.432e-03)	-4.439e-04 (3.432e-03)	3.493e-03 (9.817e-03)	-7.852e-05 (9.817e-03)	5.977e-03 (9.817e-03)	1.520e-03 (9.817e-03)
		0.04	0.10	-2.793e-04 (2.390e-03)	-2.793e-04 (2.390e-03)	-2.150e-04 (1.793e-03)	-2.150e-04 (1.793e-03)	1.143e-02 (6.374e-03)	2.316e-03 (6.374e-03)	6.143e-03 (6.374e-03)	1.602e-03 (6.374e-03)
		0.10	0.04	-4.006e-04 (3.014e-03)	-4.006e-04 (3.014e-03)	-3.966e-04 (2.948e-03)	-3.966e-04 (2.948e-03)	-6.412e-03 (6.785e-03)	-1.704e-03 (6.785e-03)	1.815e-03 (6.785e-03)	4.182e-04 (6.785e-03)
		0.01	0.10	-1.761e-04 (1.764e-03)	-1.761e-04 (1.764e-03)	-9.754e-05 (9.503e-04)	-9.754e-05 (9.503e-04)	1.735e-02 (4.348e-03)	7.050e-03 (4.348e-03)	6.093e-03 (4.348e-03)	1.660e-03 (4.348e-03)
		0.10	0.01	-3.581e-04 (2.662e-03)	-3.581e-04 (2.662e-03)	-3.639e-04 (2.660e-03)	-3.639e-04 (2.660e-03)	-1.330e-02 (4.774e-03)	-5.738e-03 (4.774e-03)	-3.412e-03 (4.774e-03)	-1.233e-03 (4.774e-03)
	2	0.10	0.10	-2.816e-03 (3.230e-02)	-2.816e-03 (3.230e-02)	-2.787e-03 (3.232e-02)	-2.787e-03 (3.232e-02)	1.807e-04 (1.442e-02)	9.431e-04 (1.442e-02)	-1.195e-03 (1.442e-02)	-1.182e-03 (1.442e-02)
		0.04	0.10	-1.580e-03 (2.202e-02)	-1.580e-03 (2.202e-02)	-1.548e-03 (2.205e-02)	-1.548e-03 (2.205e-02)	2.651e-02 (1.011e-02)	2.352e-02 (1.011e-02)	1.232e-02 (1.011e-02)	4.707e-03 (1.011e-02)
		0.10	0.04	-2.193e-03 (2.200e-02)	-1.982e-03 (2.245e-02)	-2.202e-03 (2.200e-02)	-2.202e-03 (2.200e-02)	-2.625e-02 (1.006e-02)	-2.224e-02 (1.006e-02)	-1.402e-02 (1.006e-02)	-6.208e-03 (1.006e-02)
		0.01	0.10	2.151e-03 (1.347e-02)	2.151e-03 (1.347e-02)	2.183e-03 (1.349e-02)	2.183e-03 (1.349e-02)	3.970e-02 (7.925e-03)	3.482e-02 (7.925e-03)	1.986e-02 (7.925e-03)	9.841e-03 (7.925e-03)
		0.10	0.01	-4.451e-03 (1.358e-02)	-1.564e-03 (1.753e-02)	-4.458e-03 (1.359e-02)	-4.458e-03 (1.359e-02)	-3.945e-02 (7.923e-03)	-3.385e-02 (7.923e-03)	-2.179e-02 (7.923e-03)	-1.252e-02 (7.923e-03)
	3	0.10	0.10	-6.362e-04 (1.199e-02)	-6.362e-04 (1.199e-02)	-5.250e-04 (1.187e-02)	-7.616e-04 (1.211e-02)	1.926e-03 (1.397e-02)	2.096e-03 (1.397e-02)	8.207e-03 (1.397e-02)	2.701e-03 (1.397e-02)
		0.04	0.10	-4.218e-04 (8.336e-03)	-4.218e-04 (8.336e-03)	-2.993e-04 (8.109e-03)	-4.720e-04 (8.281e-03)	2.193e-02 (9.676e-03)	1.439e-02 (9.676e-03)	1.003e-02 (9.676e-03)	2.365e-03 (9.676e-03)
		0.10	0.04	-4.689e-04 (8.495e-03)	-4.689e-04 (8.495e-03)	-4.332e-04 (8.485e-03)	-5.921e-04 (8.654e-03)	-1.927e-02 (9.651e-03)	-1.176e-02 (9.651e-03)	-3.134e-04 (9.651e-03)	9.745e-04 (9.651e-03)
		0.01	0.10	-1.474e-04 (6.226e-03)	-1.474e-04 (6.226e-03)	-6.644e-05 (5.980e-03)	-1.859e-04 (6.078e-03)	3.213e-02 (7.298e-03)	2.138e-02 (7.298e-03)	1.154e-02 (7.298e-03)	3.500e-03 (7.298e-03)
		0.10	0.01	-6.160e-04 (6.339e-03)	-3.852e-04 (6.773e-03)	-6.146e-04 (6.346e-03)	-7.508e-04 (6.443e-03)	-3.002e-02 (7.377e-03)	-1.969e-02 (7.377e-03)	-8.279e-03 (7.377e-03)	-3.807e-03 (7.377e-03)
	4	0.10	0.10	2.247e-04 (8.260e-03)	2.247e-04 (8.260e-03)	2.704e-04 (6.804e-03)	-9.437e-04 (1.112e-02)	1.854e-03 (1.153e-02)	3.752e-04 (1.153e-02)	3.715e-03 (1.153e-02)	2.585e-03 (1.153e-02)
		0.04	0.10	1.825e-04 (6.345e-03)	1.825e-04 (6.345e-03)	1.628e-04 (4.299e-03)	-5.337e-04 (7.024e-03)	1.545e-02 (7.876e-03)	6.063e-03 (7.876e-03)	3.515e-03 (7.876e-03)	2.057e-03 (7.876e-03)
		0.10	0.04	1.321e-04 (5.357e-03)	1.321e-04 (5.357e-03)	1.814e-04 (5.028e-03)	-7.143e-04 (8.226e-03)	-1.283e-02 (7.809e-03)	-5.284e-03 (7.809e-03)	5.802e-04 (7.809e-03)	1.616e-03 (7.809e-03)
		0.01	0.10	2.679e-04 (5.207e-03)	2.679e-04 (5.207e-03)	9.021e-05 (2.889e-03)	-2.816e-04 (4.713e-03)	2.298e-02 (5.761e-03)	1.165e-02 (5.761e-03)	4.628e-03 (5.761e-03)	2.348e-03 (5.761e-03)
		0.10	0.01	8.575e-05 (4.022e-03)	8.575e-05 (4.022e-03)	1.053e-04 (3.992e-03)	-7.044e-04 (6.174e-03)	-2.115e-02 (5.641e-03)	-1.070e-02 (5.641e-03)	-5.456e-03 (5.641e-03)	-1.563e-03 (5.641e-03)
12000	1	0.10	0.10	-4.040e-04 (3.256e-03)	-4.040e-04 (3.256e-03)	-3.286e-04 (2.850e-03)	-3.286e-04 (2.850e-03)	3.046e-03 (8.896e-03)	9.321e-05 (8.896e-03)	4.693e-03 (8.896e-03)	9.573e-04 (8.896e-03)
		0.04	0.10	-2.330e-04 (2.057e-03)	-2.330e-04 (2.057e-03)	-1.479e-04 (1.479e-03)	-1.479e-04 (1.479e-03)	9.476e-03 (5.712e-03)	1.594e-03 (5.712e-03)	4.836e-03 (5.712e-03)	1.103e-03 (5.712e-03)
		0.10	0.04	-3.326e-04 (2.573e-03)	-3.326e-04 (2.573e-03)	-3.086e-04 (2.471e-03)	-3.086e-04 (2.471e-03)	-4.893e-03 (6.188e-03)	-8.652e-04 (6.188e-03)	1.567e-03 (6.188e-03)	1.125e-04 (6.188e-03)
		0.01	0.10	-1.475e-04 (1.523e-03)	-1.475e-04 (1.523e-03)	-5.809e-05 (7.829e-04)	-5.809e-05 (7.829e-04)	1.497e-02 (3.836e-03)	5.791e-03 (3.836e-03)	4.847e-03 (3.836e-03)	1.271e-03 (3.836e-03)
		0.10	0.01	-2.968e-04 (2.266e-03)	-2.968e-04 (2.266e-03)	-2.941e-04 (2.251e-03)	-2.941e-04 (2.251e-03)	-1.118e-02 (4.278e-03)	-4.403e-03 (4.278e-03)	-2.522e-03 (4.278e-03)	-1.037e-03 (4.278e-03)
	2	0.10	0.10	-2.479e-03 (2.708e-02)	-2.479e-03 (2.708e-02)	-2.478e-03 (2.718e-02)	-2.478e-03 (2.718e-02)	1.444e-03 (1.416e-02)	1.819e-03 (1.416e-02)	-8.404e-03 (1.416e-02)	-3.532e-03 (1.416e-02)
		0.04	0.10	-1.613e-03 (1.887e-02)	-1.613e-03 (1.887e-02)	-1.578e-03 (1.893e-02)	-1.578e-03 (1.893e-02)	2.676e-02 (9.881e-03)	2.300e-02 (9.881e-03)	6.972e-03 (9.881e-03)	1.759e-03 (9.881e-03)
		0.10	0.04	-1.776e-03 (1.879e-02)	-1.756e-03 (1.884e-02)	-1.796e-03 (1.882e-02)	-1.796e-03 (1.882e-02)	-2.469e-02 (9.904e-03)	-2.059e-02 (9.904e-03)	-1.838e-02 (9.904e-03)	-6.732e-03 (9.904e-03)
		0.01	0.10	1.221e-03 (1.181e-02)	1.221e-03 (1.181e-02)	1.250e-03 (1.188e-02)	1.250e-03 (1.188e-02)	3.938e-02 (7.714e-03)	3.366e-02 (7.714e-03)	1.646e-02 (7.714e-03)	7.564e-03 (7.714e-03)
		0.10	0.01	-3.500e-03 (1.170e-02)	-1.394e-03 (1.472e-02)	-3.513e-03 (1.173e-02)	-3.513e-03 (1.173e-02)	-3.776e-02 (7.753e-03)	-3.193e-02 (7.753e-03)	-2.420e-02 (7.753e-03)	-1.171e-02 (7.753e-03)
	3	0.10	0.10	-7.615e-04 (1.017e-02)	-7.615e-04 (1.017e-02)	-7.234e-04 (1.017e-02)	-7.224e-04 (1.019e-02)	2.797e-03 (1.271e-02)	1.896e-03 (1.271e-02)	5.798e-03 (1.271e-02)	2.360e-03 (1.271e-02)
		0.04	0.10	-4.943e-04 (7.046e-03)	-4.943e-04 (7.046e-03)	-4.452e-04 (6.957e-03)	-4.296e-04 (6.978e-03)	2.113e-02 (8.768e-03)	1.236e-02 (8.768e-03)	6.704e-03 (8.768e-03)	2.427e-03 (8.768e-03)
		0.10	0.04	-5.718e-04 (7.230e-03)	-5.718e-04 (7.230e-03)	-5.655e-04 (7.264e-03)	-5.791e-04 (7.270e-03)	-1.727e-02 (8.752e-03)	-9.889e-03 (8.752e-03)	-6.482e-04 (8.752e-03)	3.879e-04 (8.752e-03)
		0.01	0.10	-3.136e-04 (5.416e-03)	-3.136e-04 (5.416e-03)	-2.636e-04 (5.250e-03)	-2.287e-04 (5.252e-03)	3.050e-02 (6.562e-03)	1.887e-02 (6.562e-03)	8.877e-03 (6.562e-03)	3.321e-03 (6.562e-03)
		0.10	0.01	-5.613e-04 (5.594e-03)	-4.770e-04 (5.784e-03)	-5.677e-04 (5.608e-03)	-5.845e-04 (5.625e-03)	-2.757e-02 (6.671e-03)	-1.730e-02 (6.671e-03)	-7.425e-03 (6.671e-03)	-3.533e-03 (6.671e-03)
	4	0.10	0.10	5.959e-04 (7.006e-03)	5.959e-04 (7.006e-03)	3.456e-04 (5.153e-03)	-1.348e-04 (8.244e-03)	1.746e-03 (1.083e-02)	7.589e-04 (1.083e-02)	2.912e-03 (1.083e-02)	2.064e-03 (1.083e-02)
		0.04	0.10	5.061e-04 (5.531e-03)	5.061e-04 (5.531e-03)	1.872e-04 (3.151e-03)	-2.360e-05 (5.108e-03)	1.142e-02 (7.475e-03)	3.582e-03 (7.475e-03)	2.968e-03 (7.475e-03)	2.190e-03 (7.475e-03)
		0.10	0.04	3.282e-04 (4.397e-03)	3.282e-04 (4.397e-03)	2.673e-04 (3.905e-03)	-1.416e-04 (6.224e-03)	-8.897e-03 (7.147e-03)	-2.311e-03 (7.147e-03)	1.286e-03 (7.147e-03)	7.879e-04 (7.147e-03)
		0.01	0.10	4.809e-04 (4.806e-03)	4.809e-04 (4.806e-03)	8.968e-05 (2.011e-03)	3.167e-05 (3.399e-03)	1.780e-02 (5.394e-03)	8.334e-03 (5.394e-03)	3.309e-03 (5.394e-03)	2.316e-03 (5.394e-03)
		0.10	0.01	1.943e-04 (3.205e-03)	1.943e-04 (3.205e-03)	1.977e-04 (3.163e-03)	-1.549e-04 (4.965e-03)	-1.576e-02 (4.922e-03)	-7.170e-03 (4.922e-03)	-2.904e-03 (4.922e-03)	-8.778e-04 (4.922e-03)

Table A.2: Bias and empirical standard errors of σ_1^2 (random effect variance) points estimates for all methods, based on 200 replicates for each simulation setting. Setting 1, 2, 3 are described in Appendix A.3.3. Setting 4 corresponds to the main text simulation setting.

n	Setting	σ_0^2	σ_1^2	REHE	HE	REML	sREML	reREHE.0.05	reREHE.0.1	reREHE.0.05.median	reREHE.0.1.median
3000	1	0.10	0.10	6.187e-04 (8.558e-03)	6.187e-04 (8.558e-03)	6.338e-04 (7.997e-03)	6.338e-04 (7.997e-03)	-9.637e-03 (1.342e-02)	-1.718e-03 (1.342e-02)	-1.837e-02 (1.342e-02)	-6.588e-03 (1.342e-02)
		0.04	0.10	3.844e-04 (6.612e-03)	3.844e-04 (6.612e-03)	3.389e-04 (5.697e-03)	3.389e-04 (5.697e-03)	-2.293e-02 (9.496e-03)	-8.275e-03 (9.496e-03)	-1.872e-02 (9.496e-03)	-6.403e-03 (9.496e-03)
		0.10	0.04	4.818e-04 (5.473e-03)	4.818e-04 (5.473e-03)	5.164e-04 (5.399e-03)	5.164e-04 (5.399e-03)	9.948e-03 (8.901e-03)	5.797e-03 (8.901e-03)	-7.757e-03 (8.901e-03)	-2.790e-03 (8.901e-03)
		0.01	0.10	2.672e-04 (5.692e-03)	2.672e-04 (5.692e-03)	1.728e-04 (4.507e-03)	1.728e-04 (4.507e-03)	-2.971e-02 (7.261e-03)	-1.373e-02 (7.261e-03)	-1.998e-02 (7.261e-03)	-7.959e-03 (7.261e-03)
		0.10	0.01	4.220e-04 (3.999e-03)	4.220e-04 (3.999e-03)	4.435e-04 (4.016e-03)	4.435e-04 (4.016e-03)	2.042e-02 (6.398e-03)	1.185e-02 (6.398e-03)	7.375e-05 (6.398e-03)	-3.040e-04 (6.398e-03)
	2	0.10	0.10	5.387e-03 (5.091e-02)	5.842e-03 (5.184e-02)	5.528e-03 (5.106e-02)	5.528e-03 (5.106e-02)	-2.334e-03 (1.434e-02)	-2.531e-04 (1.434e-02)	-1.838e-02 (1.434e-02)	-7.007e-03 (1.434e-02)
		0.04	0.10	1.256e-03 (3.418e-02)	3.659e-03 (3.769e-02)	1.184e-03 (3.395e-02)	1.184e-03 (3.395e-02)	-2.943e-02 (1.011e-02)	-2.580e-02 (1.011e-02)	-3.292e-02 (1.011e-02)	-1.863e-02 (1.011e-02)
		0.10	0.04	6.308e-03 (3.211e-02)	6.308e-03 (3.211e-02)	6.346e-03 (3.205e-02)	6.346e-03 (3.205e-02)	2.622e-02 (9.993e-03)	2.548e-02 (9.993e-03)	7.414e-03 (9.993e-03)	8.529e-03 (9.993e-03)
		0.01	0.10	-5.544e-03 (2.137e-02)	2.858e-03 (2.999e-02)	-5.595e-03 (2.118e-02)	-5.595e-03 (2.118e-02)	-4.301e-02 (7.942e-03)	-3.864e-02 (7.942e-03)	-4.050e-02 (7.942e-03)	-2.581e-02 (7.942e-03)
		0.10	0.01	9.497e-03 (1.951e-02)	9.497e-03 (1.951e-02)	9.512e-03 (1.948e-02)	9.512e-03 (1.948e-02)	4.051e-02 (7.785e-03)	3.842e-02 (7.785e-03)	2.061e-02 (7.785e-03)	1.653e-02 (7.785e-03)
	3	0.10	0.10	2.185e-03 (2.114e-02)	2.185e-03 (2.114e-02)	2.214e-03 (2.110e-02)	2.201e-03 (2.113e-02)	-5.566e-03 (1.436e-02)	-2.080e-03 (1.436e-02)	-2.089e-02 (1.436e-02)	-8.120e-03 (1.436e-02)
		0.04	0.10	1.483e-03 (1.514e-02)	1.483e-03 (1.514e-02)	1.471e-03 (1.494e-02)	1.471e-03 (1.496e-02)	-2.874e-02 (1.022e-02)	-2.089e-02 (1.022e-02)	-2.710e-02 (1.022e-02)	-1.006e-02 (1.022e-02)
		0.10	0.04	1.580e-03 (1.450e-02)	1.580e-03 (1.450e-02)	1.618e-03 (1.457e-02)	1.605e-03 (1.457e-02)	2.099e-02 (9.918e-03)	1.797e-02 (9.918e-03)	-1.841e-03 (9.918e-03)	-1.507e-03 (9.918e-03)
		0.01	0.10	1.372e-04 (1.093e-02)	1.132e-03 (1.217e-02)	-4.462e-05 (1.033e-02)	-5.103e-05 (1.031e-02)	-4.028e-02 (8.129e-03)	-3.041e-02 (8.129e-03)	-3.164e-02 (8.129e-03)	-1.423e-02 (8.129e-03)
		0.10	0.01	2.426e-03 (9.325e-03)	2.426e-03 (9.325e-03)	2.460e-03 (9.387e-03)	2.449e-03 (9.374e-03)	3.430e-02 (7.696e-03)	2.825e-02 (7.696e-03)	9.178e-03 (7.696e-03)	4.751e-03 (7.696e-03)
	4	0.10	0.10	-1.870e-04 (2.271e-02)	-1.870e-04 (2.271e-02)	-7.783e-05 (2.033e-02)	-7.006e-06 (3.272e-02)	-3.862e-03 (1.745e-02)	-3.059e-03 (1.745e-02)	-1.792e-02 (1.745e-02)	-8.457e-03 (1.745e-02)
		0.04	0.10	-2.305e-04 (1.700e-02)	-2.305e-04 (1.700e-02)	-1.493e-04 (1.376e-02)	-1.084e-04 (2.077e-02)	-2.695e-02 (1.218e-02)	-2.136e-02 (1.218e-02)	-2.180e-02 (1.218e-02)	-1.054e-02 (1.218e-02)
		0.10	0.04	-3.126e-05 (1.490e-02)	-3.126e-05 (1.490e-02)	1.379e-06 (1.433e-02)	3.969e-04 (2.344e-02)	2.158e-02 (1.205e-02)	1.705e-02 (1.205e-02)	-3.399e-03 (1.205e-02)	-1.966e-03 (1.205e-02)
		0.01	0.10	-1.981e-03 (1.153e-02)	-2.523e-04 (1.422e-02)	-8.698e-04 (9.157e-03)	-2.073e-03 (1.200e-02)	-3.862e-02 (9.492e-03)	-3.073e-02 (9.492e-03)	-2.673e-02 (9.492e-03)	-1.498e-02 (9.492e-03)
		0.10	0.01	1.093e-03 (9.567e-03)	1.093e-03 (9.567e-03)	1.102e-03 (9.518e-03)	3.655e-03 (1.477e-02)	3.440e-02 (9.115e-03)	2.737e-02 (9.115e-03)	6.188e-03 (9.115e-03)	4.148e-03 (9.115e-03)
6000	1	0.10	0.10	3.279e-04 (5.975e-03)	3.279e-04 (5.975e-03)	2.506e-04 (5.771e-03)	2.506e-04 (5.771e-03)	-4.732e-03 (1.165e-02)	-1.429e-03 (1.165e-02)	-1.134e-02 (1.165e-02)	-3.114e-03 (1.165e-02)
		0.04	0.10	1.871e-04 (4.603e-03)	1.871e-04 (4.603e-03)	7.227e-05 (4.162e-03)	7.227e-05 (4.162e-03)	-1.480e-02 (8.060e-03)	-4.730e-03 (8.060e-03)	-1.086e-02 (8.060e-03)	-3.002e-03 (8.060e-03)
		0.10	0.04	2.719e-04 (3.839e-03)	2.719e-04 (3.839e-03)	2.610e-04 (3.853e-03)	2.610e-04 (3.853e-03)	8.371e-03 (7.592e-03)	2.315e-03 (7.592e-03)	-4.650e-03 (7.592e-03)	-1.114e-03 (7.592e-03)
		0.01	0.10	1.168e-04 (3.957e-03)	1.168e-04 (3.957e-03)	-2.494e-05 (3.328e-03)	-2.494e-05 (3.328e-03)	-2.091e-02 (6.010e-03)	-9.534e-03 (6.010e-03)	-1.218e-02 (6.010e-03)	-3.919e-03 (6.010e-03)
		0.10	0.01	2.439e-04 (2.838e-03)	2.439e-04 (2.838e-03)	2.524e-04 (2.859e-03)	2.524e-04 (2.859e-03)	1.629e-02 (5.237e-03)	7.381e-03 (5.237e-03)	-1.933e-04 (5.237e-03)	9.409e-05 (5.237e-03)
	2	0.10	0.10	3.673e-03 (3.911e-02)	3.730e-03 (3.926e-02)	3.729e-03 (3.906e-02)	3.729e-03 (3.906e-02)	-1.720e-03 (1.417e-02)	2.984e-04 (1.417e-02)	-1.020e-02 (1.417e-02)	-5.964e-03 (1.417e-02)
		0.04	0.10	1.638e-03 (2.625e-02)	2.555e-03 (2.787e-02)	1.667e-03 (2.617e-02)	1.667e-03 (2.617e-02)	-2.824e-02 (9.856e-03)	-2.403e-02 (9.856e-03)	-2.151e-02 (9.856e-03)	-1.235e-02 (9.856e-03)
		0.10	0.04	3.567e-03 (2.529e-02)	3.567e-03 (2.529e-02)	3.623e-03 (2.533e-02)	3.623e-03 (2.533e-02)	2.571e-02 (9.999e-03)	2.440e-02 (9.999e-03)	6.917e-03 (9.999e-03)	4.293e-03 (9.999e-03)
		0.01	0.10	-3.174e-03 (1.611e-02)	1.981e-03 (2.215e-02)	-3.178e-03 (1.601e-02)	-3.178e-03 (1.601e-02)	-4.150e-02 (7.725e-03)	-3.629e-02 (7.725e-03)	-2.910e-02 (7.725e-03)	-1.754e-02 (7.725e-03)
		0.10	0.01	5.907e-03 (1.573e-02)	5.907e-03 (1.573e-02)	5.930e-03 (1.580e-02)	5.930e-03 (1.580e-02)	3.943e-02 (7.832e-03)	3.645e-02 (7.832e-03)	1.617e-02 (7.832e-03)	1.140e-02 (7.832e-03)
	3	0.10	0.10	1.246e-03 (1.539e-02)	1.246e-03 (1.539e-02)	1.278e-03 (1.516e-02)	1.256e-03 (1.517e-02)	-4.521e-03 (1.411e-02)	-1.226e-03 (1.411e-02)	-1.387e-02 (1.411e-02)	-5.287e-03 (1.411e-02)
		0.04	0.10	8.123e-04 (1.102e-02)	8.123e-04 (1.102e-02)	8.281e-04 (1.065e-02)	8.042e-04 (1.066e-02)	-2.572e-02 (1.001e-02)	-1.670e-02 (1.001e-02)	-1.652e-02 (1.001e-02)	-5.654e-03 (1.001e-02)
		0.10	0.04	9.316e-04 (1.058e-02)	9.316e-04 (1.058e-02)	9.554e-04 (1.054e-02)	9.472e-04 (1.055e-02)	1.949e-02 (9.610e-03)	1.501e-02 (9.610e-03)	-3.296e-03 (9.610e-03)	-2.155e-03 (9.610e-03)
		0.01	0.10	2.194e-04 (8.297e-03)	5.956e-04 (8.864e-03)	1.728e-04 (7.733e-03)	2.042e-04 (7.819e-03)	-3.632e-02 (7.865e-03)	-2.499e-02 (7.865e-03)	-1.992e-02 (7.865e-03)	-8.863e-03 (7.865e-03)
		0.10	0.01	1.267e-03 (7.264e-03)	1.267e-03 (7.264e-03)	1.283e-03 (7.285e-03)	1.294e-03 (7.242e-03)	3.154e-02 (7.337e-03)	2.370e-02 (7.337e-03)	4.715e-03 (7.337e-03)	1.732e-03 (7.337e-03)
	4	0.10	0.10	-1.218e-03 (1.156e-02)	-1.218e-03 (1.156e-02)	-1.017e-03 (9.976e-03)	5.410e-04 (1.613e-02)	-2.630e-03 (1.471e-02)	-1.014e-03 (1.471e-02)	-1.072e-02 (1.471e-02)	-6.116e-03 (1.471e-02)
		0.04	0.10	-9.707e-04 (9.064e-03)	-9.707e-04 (9.064e-03)	-7.015e-04 (6.913e-03)	2.657e-04 (1.035e-02)	-2.060e-02 (1.058e-02)	-1.173e-02 (1.058e-02)	-1.087e-02 (1.058e-02)	-5.123e-03 (1.058e-02)
		0.10	0.04	-7.347e-04 (7.217e-03)	-7.347e-04 (7.217e-03)	-6.831e-04 (6.841e-03)	4.105e-04 (1.186e-02)	1.688e-02 (9.651e-03)	1.014e-02 (9.651e-03)	-4.346e-03 (9.651e-03)	-3.894e-03 (9.651e-03)
		0.01	0.10	-9.623e-04 (7.634e-03)	-8.470e-04 (7.866e-03)	-5.196e-04 (5.205e-03)	-1.134e-04 (6.971e-03)	-2.984e-02 (8.192e-03)	-1.849e-02 (8.192e-03)	-1.384e-02 (8.192e-03)	-7.422e-03 (8.192e-03)
		0.10	0.01	-4.286e-04 (5.003e-03)	-4.286e-04 (5.003e-03)	-4.257e-04 (4.991e-03)	9.354e-04 (8.316e-03)	2.688e-02 (6.936e-03)	1.705e-02 (6.936e-03)	2.009e-03 (6.936e-03)	-1.029e-03 (6.936e-03)
9000	1	0.10	0.10	3.432e-04 (4.936e-03)	3.432e-04 (4.936e-03)	2.990e-04 (4.681e-03)	2.990e-04 (4.681e-03)	-3.487e-03 (1.014e-02)	-2.141e-04 (1.014e-02)	-6.497e-03 (1.014e-02)	-1.756e-03 (1.014e-02)
		0.04	0.10	2.154e-04 (3.796e-03)	2.154e-04 (3.796e-03)	1.379e-04 (3.345e-03)	1.379e-04 (3.345e-03)	-1.122e-02 (6.808e-03)	-2.370e-03 (6.808e-03)	-6.520e-03 (6.808e-03)	-1.974e-03 (6.808e-03)
		0.10	0.04	2.652e-04 (3.171e-03)	2.652e-04 (3.171e-03)	2.636e-04 (3.150e-03)	2.636e-04 (3.150e-03)	6.294e-03 (6.801e-03)	1.437e-03 (6.801e-03)	-2.833e-03 (6.801e-03)	-6.928e-04 (6.801e-03)
		0.01	0.10	1.515e-04 (3.257e-03)	1.515e-04 (3.257e-03)	4.815e-05 (2.647e-03)	4.815e-05 (2.647e-03)	-1.696e-02 (4.818e-03)	-6.808e-03 (4.818e-03)	-7.700e-03 (4.818e-03)	-2.578e-03 (4.818e-03)
		0.10	0.01	2.261e-04 (2.340e-03)	2.261e-04 (2.340e-03)	2.338e-04 (2.351e-03)	2.338e-04 (2.351e-03)	1.315e-02 (4.605e-03)	5.496e-03 (4.605e-03)	-2.600e-04 (4.605e-03)	-8.072e-05 (4.605e-03)
	2	0.10	0.10	2.567e-03 (3.255e-02)	2.567e-03 (3.255e-02)	2.542e-03 (3.255e-02)	2.542e-03 (3.255e-02)	-3.839e-04 (1.433e-02)	-1.322e-03 (1.433e-02)	-8.204e-03 (1.433e-02)	-2.368e-03 (1.433e-02)
		0.04	0.10	1.414e-03 (2.232e-02)	1.791e-03 (2.306e-02)	1.376e-03 (2.232e-02)	1.376e-03 (2.232e-02)	-2.664e-02 (1.006e-02)	-2.377e-02 (1.006e-02)	-1.870e-02 (1.006e-02)	-7.857e-03 (1.006e-02)
		0.10	0.04	2.015e-03 (2.207e-02)	2.015e-03 (2.207e-02)	2.026e-03 (2.206e-02)	2.026e-03 (2.206e-02)	2.610e-02 (9.973e-03)	2.196e-02 (9.973e-03)	7.084e-03 (9.973e-03)	3.423e-03 (9.973e-03)
		0.01	0.10	-2.245e-03 (1.380e-02)	1.403e-03 (1.832e-02)	-2.348e-03 (1.373e-02)	-2.348e-03 (1.373e-02)	-3.980e-02 (7.888e-03)	-3.501e-02 (7.888e-03)	-2.563e-02 (7.888e-03)	-1.339e-02 (7.888e-03)
		0.10	0.01	4.308e-03 (1.353e-02)	4.308e-03 (1.353e-02)	4.317e-03 (1.354e-02)	4.317e-03 (1.354e-02)	3.932e-02 (7.826e-03)	3.362e-02 (7.826e-03)	1.585e-02 (7.826e-03)	8.715e-03 (7.826e-03)
	3	0.10	0.10	3.903e-04 (1.256e-02)	3.903e-04 (1.256e-02)	2.802e-04 (1.238e-02)	5.173e-04 (1.255e-02)	-2.052e-03 (1.415e-02)	-2.456e-03 (1.415e-02)	-1.141e-02 (1.415e-02)	-2.722e-03 (1.415e-02)
		0.04	0.10	2.544e-04 (8.987e-03)	2.544e-04 (8.987e-03)	1.308e-04 (8.703e-03)	3.044e-04 (8.821e-03)	-2.198e-02 (9.900e-03)	-1.459e-02 (9.900e-03)	-1.195e-02 (9.900e-03)	-2.669e-03 (9.900e-03)
		0.10	0.04	2.920e-04 (8.634e-03)	2.920e-04 (8.634e-03)	2.581e-04 (8.601e-03)	4.171e-04 (8.724e-03)	1.915e-02 (9.646e-03)	1.148e-02 (9.646e-03)	-3.209e-03 (9.646e-03)	-1.480e-03 (9.646e-03)
		0.01	0.10	3.017e-05 (6.941e-03)	1.864e-04 (7.223e-03)	-7.333e-05 (6.615e-03)	4.603e-05 (6.669e-03)	-3.213e-02 (7.543e-03)	-2.148e-02 (7.543e-03)	-1.481e-02 (7.543e-03)	-5.668e-03 (7.543e-03)
		0.10	0.01	4.736e-04 (6.245e-03)	4.736e-04 (6.245e-03)	4.739e-04 (6.247e-03)	6.100e-04 (6.322e-03)	2.990e-02 (7.270e-03)	1.945e-02 (7.270e-03)	3.587e-03 (7.270e-03)	8.458e-04 (7.270e-03)
	4	0.10	0.10	-5.309e-04 (8.996e-03)	-5.309e-04 (8.996e-03)	-5.703e-04 (7.365e-03)	6.450e-04 (1.127e-02)	-1.994e-03 (1.207e-02)	-8.203e-04 (1.207e-02)	-4.967e-03 (1.207e-02)	-3.444e-03 (1.207e-02)
		0.04	0.10	-4.103e-04 (7.210e-03)	-4.103e-04 (7.210e-03)	-3.765e-04 (5.050e-03)	3.110e-04 (7.344e-03)	-1.545e-02 (8.472e-03)	-6.292e-03 (8.472e-03)	-4.303e-03 (8.472e-03)	-2.579e-03 (8.472e-03)
		0.10	0.04	-3.330e-04 (5.460e-03)	-3.330e-04 (5.460e-03)	-3.808e-04 (5.049e-03)	5.170e-04 (8.132e-03)	1.266e-02 (7.964e-03)	4.948e-03 (7.964e-03)	-2.432e-03 (7.964e-03)	-2.255e-03 (7.964e-03)
		0.01	0.10	-4.370e-04 (6.172e-03)	-3.500e-04 (6.352e-03)	-2.538e-04 (3.734e-03)	9.382e-05 (5.156e-03)	-2.289e-02 (6.380e-03)	-1.170e-02 (6.380e-03)	-7.669e-03 (6.380e-03)	-4.077e-03 (6.380e-03)
		0.10	0.01	-2.340e-04 (3.770e-03)	-2.340e-04 (3.770e-03)	-2.521e-04 (3.726e-03)	5.575e-04 (5.949e-03)	2.100e-02 (5.632e-03)	1.043e-02 (5.632e-03)	9.468e-04 (5.632e-03)	-9.221e-04 (5.632e-03)
12000	1	0.10	0.10	1.898e-04 (4.098e-03)	1.898e-04 (4.098e-03)	1.045e-04 (3.892e-03)	1.045e-04 (3.892e-03)	-3.274e-03 (9.294e-03)	-4.730e-04 (9.294e-03)	-5.360e-03 (9.294e-03)	-1.749e-03 (9.294e-03)
		0.04	0.10	1.064e-04 (3.161e-03)	1.064e-04 (3.161e-03)	1.735e-06 (2.819e-03)	1.735e-06 (2.819e-03)	-9.478e-03 (6.178e-03)	-1.763e-03 (6.178e-03)	-5.231e-03 (6.178e-03)	-1.653e-03 (6.178e-03)
		0.10	0.04	1.593e-04 (2.625e-03)	1.593e-04 (2.625e-03)	1.347e-04 (2.598e-03)	1.347e-04 (2.598e-03)	4.657e-03 (6.271e-03)	5.651e-04 (6.271e-03)	-2.048e-03 (6.271e-03)	-8.227e-04 (6.271e-03)
		0.01	0.10	6.474e-05 (2.718e-03)	6.474e-05 (2.718e-03)	-5.393e-05 (2.266e-03)	-5.393e-05 (2.266e-03)	-1.477e-02 (4.373e-03)	-5.684e-03 (4.373e-03)	-6.425e-03 (4.373e-03)	-2.027e-03 (4.373e-03)
		0.10	0.01	1.440e-04 (1.932e-03)	1.440e-04 (1.932e-03)	1.421e-04 (1.936e-03)	1.421e-04 (1.936e-03)	1.096e-02 (4.147e-03)	4.148e-03 (4.147e-03)	5.900e-05 (4.147e-03)	-2.070e-04 (4.147e-03)
	2	0.10	0.10	2.199e-03 (2.729e-02)	2.199e-03 (2.729e-02)	2.199e-03 (2.737e-02)	2.199e-03 (2.737e-02)	-1.759e-03 (1.425e-02)	-2.224e-03 (1.425e-02)	6.422e-04 (1.425e-02)	1.783e-03 (1.425e-02)
		0.04	0.10	1.421e-03 (1.910e-02)	1.522e-03 (1.932e-02)	1.384e-03 (1.915e-02)	1.384e-03 (1.915e-02)	-2.697e-02 (1.001e-02)	-2.327e-02 (1.001e-02)	-1.294e-02 (1.001e-02)	-3.723e-03 (1.001e-02)
		0.10	0.04	1.576e-03 (1.885e-02)	1.576e-03 (1.885e-02)	1.597e-03 (1.888e-02)	1.597e-03 (1.888e-02)	2.447e-02 (9.915e-03)	2.030e-02 (9.915e-03)	1.272e-02 (9.915e-03)	4.804e-03 (9.915e-03)
		0.01	0.10	-1.345e-03 (1.206e-02)	1.184e-03 (1.534e-02)	-1.423e-03 (1.208e-02)	-1.423e-03 (1.208e-02)	-3.954e-02 (7.858e-03)	-3.386e-02 (7.858e-03)	-2.165e-02 (7.858e-03)	-1.027e-02 (7.858e-03)
		0.10	0.01	3.340e-03 (1.168e-02)	3.340e-03 (1.168e-02)	3.354e-03 (1.170e-02)	3.354e-03 (1.170e-02)	3.757e-02 (7.730e-03)	3.169e-02 (7.730e-03)	1.947e-02 (7.730e-03)	8.921e-03 (7.730e-03)
	3	0.10	0.10	4.906e-04 (1.063e-02)	4.906e-04 (1.063e-02)	4.536e-04 (1.057e-02)	4.513e-04 (1.062e-02)	-3.043e-03 (1.304e-02)	-2.291e-03 (1.304e-02)	-6.143e-03 (1.304e-02)	-2.835e-03 (1.304e-02)
		0.04	0.10	3.109e-04 (7.592e-03)	3.109e-04 (7.592e-03)	2.620e-04 (7.446e-03)	2.446e-04 (7.492e-03)	-2.126e-02 (9.127e-03)	-1.258e-02 (9.127e-03)	-7.666e-03 (9.127e-03)	-3.119e-03 (9.127e-03)
		0.10	0.04	3.758e-04 (7.323e-03)	3.758e-04 (7.323e-03)	3.706e-04 (7.335e-03)	3.839e-04 (7.363e-03)	1.707e-02 (8.847e-03)	9.582e-03 (8.847e-03)	-1.215e-03 (8.847e-03)	-1.250e-03 (8.847e-03)
		0.01	0.10	1.772e-04 (6.010e-03)	2.211e-04 (6.095e-03)	1.203e-04 (5.777e-03)	8.237e-05 (5.799e-03)	-3.057e-02 (6.911e-03)	-1.900e-02 (6.911e-03)	-1.144e-02 (6.911e-03)	-5.267e-03 (6.911e-03)
		0.10	0.01	4.028e-04 (5.505e-03)	4.028e-04 (5.505e-03)	4.101e-04 (5.516e-03)	4.268e-04 (5.547e-03)	2.739e-02 (6.648e-03)	1.705e-02 (6.648e-03)	4.134e-03 (6.648e-03)	1.103e-03 (6.648e-03)
	4	0.10	0.10	-9.755e-04 (7.662e-03)	-9.755e-04 (7.662e-03)	-7.029e-04 (5.756e-03)	-2.425e-04 (8.715e-03)	-2.088e-03 (1.089e-02)	-1.293e-03 (1.089e-02)	-4.039e-03 (1.089e-02)	-3.047e-03 (1.089e-02)
		0.04	0.10	-7.982e-04 (6.309e-03)	-7.982e-04 (6.309e-03)	-4.404e-04 (3.931e-03)	-2.664e-04 (5.666e-03)	-1.157e-02 (7.771e-03)	-3.907e-03 (7.771e-03)	-3.871e-03 (7.771e-03)	-2.695e-03 (7.771e-03)
		0.10	0.04	-5.674e-04 (4.477e-03)	-5.674e-04 (4.477e-03)	-5.027e-04 (3.948e-03)	-9.692e-05 (6.328e-03)	8.618e-03 (7.013e-03)	1.944e-03 (7.013e-03)	-1.926e-03 (7.013e-03)	-1.553e-03 (7.013e-03)
		0.01	0.10	-7.248e-04 (5.622e-03)	-7.096e-04 (5.657e-03)	-2.788e-04 (2.875e-03)	-2.774e-04 (4.017e-03)	-1.781e-02 (5.799e-03)	-8.456e-03 (5.799e-03)	-5.772e-03 (5.799e-03)	-3.320e-03 (5.799e-03)
		0.10	0.01	-3.634e-04 (2.952e-03)	-3.634e-04 (2.952e-03)	-3.661e-04 (2.905e-03)	-1.389e-05 (4.879e-03)	1.553e-02 (4.711e-03)	6.893e-03 (4.711e-03)	-2.737e-04 (4.711e-03)	-7.964e-04 (4.711e-03)

Table A.3: Bias and empirical standard errors of heritability points estimates for all methods, based on 200 replicates for each simulation setting. Setting 1, 2, 3 are described in Appendix A.3.3. Setting 4 corresponds to the main text simulation setting.

n	Setting	σ_0^2	σ_1^2	REHE	HE	REML	sREML	reREHE.0.05	reREHE.0.1	reREHE.0.05.median	reREHE.0.1.median
3000	1	0.10	0.10	3.160e-03 (3.494e-02)	3.160e-03 (3.494e-02)	3.251e-03 (3.190e-02)	3.251e-03 (3.190e-02)	-4.972e-02 (6.226e-02)	-1.025e-02 (6.226e-02)	-8.517e-02 (6.226e-02)	-2.905e-02 (6.226e-02)
		0.04	0.10	2.578e-03 (3.211e-02)	2.578e-03 (3.211e-02)	2.397e-03 (2.479e-02)	2.397e-03 (2.479e-02)	-1.669e-01 (6.017e-02)	-6.282e-02 (6.017e-02)	-1.246e-01 (6.017e-02)	-4.157e-02 (6.017e-02)
		0.10	0.04	3.586e-03 (3.637e-02)	3.586e-03 (3.637e-02)	3.824e-03 (3.580e-02)	3.824e-03 (3.580e-02)	7.052e-02 (6.084e-02)	4.076e-02 (6.084e-02)	-4.931e-02 (6.084e-02)	-1.692e-02 (6.084e-02)
		0.01	0.10	1.934e-03 (2.857e-02)	1.934e-03 (2.857e-02)	1.481e-03 (1.603e-02)	1.481e-03 (1.603e-02)	-2.752e-01 (5.576e-02)	-1.319e-01 (5.576e-02)	-1.557e-01 (5.576e-02)	-5.560e-02 (5.576e-02)
		0.10	0.01	3.899e-03 (3.610e-02)	3.899e-03 (3.610e-02)	4.095e-03 (3.626e-02)	4.095e-03 (3.626e-02)	1.856e-01 (5.690e-02)	1.075e-01 (5.690e-02)	4.513e-03 (5.690e-02)	1.896e-04 (5.690e-02)
	2	0.10	0.10	2.715e-02 (2.533e-01)	2.715e-02 (2.533e-01)	2.790e-02 (2.541e-01)	2.790e-02 (2.541e-01)	-1.130e-02 (6.905e-02)	-2.008e-03 (6.905e-02)	-4.402e-02 (6.905e-02)	-3.776e-03 (6.905e-02)
		0.04	0.10	9.237e-03 (2.413e-01)	9.237e-03 (2.413e-01)	9.082e-03 (2.400e-01)	9.082e-03 (2.400e-01)	-2.099e-01 (6.944e-02)	-1.851e-01 (6.944e-02)	-1.708e-01 (6.944e-02)	-8.764e-02 (6.944e-02)
		0.10	0.04	4.514e-02 (2.291e-01)	4.514e-02 (2.291e-01)	4.542e-02 (2.286e-01)	4.542e-02 (2.286e-01)	1.877e-01 (6.877e-02)	1.813e-01 (6.877e-02)	9.266e-02 (6.877e-02)	8.002e-02 (6.877e-02)
		0.01	0.10	-5.045e-02 (1.895e-01)	-5.045e-02 (1.895e-01)	-4.939e-02 (1.889e-01)	-4.939e-02 (1.889e-01)	-3.907e-01 (6.937e-02)	-3.523e-01 (6.937e-02)	-2.922e-01 (6.937e-02)	-1.773e-01 (6.937e-02)
		0.10	0.01	8.643e-02 (1.775e-01)	8.643e-02 (1.775e-01)	8.658e-02 (1.772e-01)	8.658e-02 (1.772e-01)	3.688e-01 (6.834e-02)	3.487e-01 (6.834e-02)	2.206e-01 (6.834e-02)	1.572e-01 (6.834e-02)
	3	0.10	0.10	1.120e-02 (1.024e-01)	1.120e-02 (1.024e-01)	1.132e-02 (1.019e-01)	1.124e-02 (1.020e-01)	-2.815e-02 (6.878e-02)	-1.172e-02 (6.878e-02)	-8.960e-02 (6.878e-02)	-3.612e-02 (6.878e-02)
		0.04	0.10	1.094e-02 (1.018e-01)	1.094e-02 (1.018e-01)	1.083e-02 (9.982e-02)	1.082e-02 (9.993e-02)	-2.059e-01 (6.930e-02)	-1.511e-01 (6.930e-02)	-1.597e-01 (6.930e-02)	-5.366e-02 (6.930e-02)
		0.10	0.04	1.143e-02 (1.024e-01)	1.143e-02 (1.024e-01)	1.168e-02 (1.027e-01)	1.157e-02 (1.027e-01)	1.499e-01 (6.831e-02)	1.275e-01 (6.831e-02)	-4.805e-03 (6.831e-02)	-8.575e-03 (6.831e-02)
		0.01	0.10	1.054e-03 (8.858e-02)	1.054e-03 (8.858e-02)	7.819e-04 (8.448e-02)	7.242e-04 (8.429e-02)	-3.672e-01 (6.934e-02)	-2.790e-01 (6.934e-02)	-2.394e-01 (6.934e-02)	-8.950e-02 (6.934e-02)
		0.10	0.01	2.208e-02 (8.460e-02)	2.208e-02 (8.460e-02)	2.237e-02 (8.510e-02)	2.226e-02 (8.494e-02)	3.119e-01 (6.776e-02)	2.563e-01 (6.776e-02)	8.928e-02 (6.776e-02)	4.486e-02 (6.776e-02)
	4	0.10	0.10	-2.711e-04 (1.110e-01)	-2.711e-04 (1.110e-01)	3.253e-04 (9.927e-02)	3.256e-04 (1.616e-01)	-1.953e-02 (8.432e-02)	-1.603e-02 (8.432e-02)	-7.201e-02 (8.432e-02)	-3.266e-02 (8.432e-02)
		0.04	0.10	-7.612e-04 (1.165e-01)	-7.612e-04 (1.165e-01)	-5.165e-05 (9.322e-02)	-3.799e-05 (1.441e-01)	-1.930e-01 (8.326e-02)	-1.538e-01 (8.326e-02)	-1.184e-01 (8.326e-02)	-5.343e-02 (8.326e-02)
		0.10	0.04	2.046e-04 (1.056e-01)	2.046e-04 (1.056e-01)	4.589e-04 (1.017e-01)	2.901e-03 (1.670e-01)	1.542e-01 (8.398e-02)	1.214e-01 (8.398e-02)	-1.603e-02 (8.398e-02)	-8.607e-03 (8.398e-02)
		0.01	0.10	-1.789e-02 (9.667e-02)	-1.789e-02 (9.667e-02)	-6.074e-03 (7.518e-02)	-1.719e-02 (1.026e-01)	-3.520e-01 (8.183e-02)	-2.811e-01 (8.183e-02)	-1.837e-01 (8.183e-02)	-9.626e-02 (8.183e-02)
		0.10	0.01	1.012e-02 (8.680e-02)	1.012e-02 (8.680e-02)	1.022e-02 (8.647e-02)	3.331e-02 (1.345e-01)	3.131e-01 (8.147e-02)	2.486e-01 (8.147e-02)	6.177e-02 (8.147e-02)	4.041e-02 (8.147e-02)
6000	1	0.10	0.10	2.179e-03 (2.436e-02)	2.179e-03 (2.436e-02)	1.798e-03 (2.282e-02)	1.798e-03 (2.282e-02)	-2.463e-02 (5.526e-02)	-6.722e-03 (5.526e-02)	-5.268e-02 (5.526e-02)	-1.287e-02 (5.526e-02)
		0.04	0.10	1.942e-03 (2.223e-02)	1.942e-03 (2.223e-02)	1.243e-03 (1.785e-02)	1.243e-03 (1.785e-02)	-1.080e-01 (5.228e-02)	-3.410e-02 (5.228e-02)	-7.274e-02 (5.228e-02)	-1.874e-02 (5.228e-02)
		0.10	0.04	2.319e-03 (2.551e-02)	2.319e-03 (2.551e-02)	2.223e-03 (2.545e-02)	2.223e-03 (2.545e-02)	5.950e-02 (5.270e-02)	1.695e-02 (5.270e-02)	-3.120e-02 (5.270e-02)	-6.153e-03 (5.270e-02)
		0.01	0.10	1.652e-03 (1.963e-02)	1.652e-03 (1.963e-02)	7.137e-04 (1.160e-02)	7.137e-04 (1.160e-02)	-1.943e-01 (4.682e-02)	-8.931e-02 (4.682e-02)	-9.386e-02 (4.682e-02)	-2.378e-02 (4.682e-02)
		0.10	0.01	2.351e-03 (2.561e-02)	2.351e-03 (2.561e-02)	2.422e-03 (2.579e-02)	2.422e-03 (2.579e-02)	1.480e-01 (4.683e-02)	6.740e-02 (4.683e-02)	-2.930e-04 (4.683e-02)	2.118e-03 (4.683e-02)
	2	0.10	0.10	1.917e-02 (1.951e-01)	1.917e-02 (1.951e-01)	1.947e-02 (1.949e-01)	1.947e-02 (1.949e-01)	-8.746e-03 (6.880e-02)	2.152e-03 (6.880e-02)	-2.462e-02 (6.880e-02)	-1.565e-02 (6.880e-02)
		0.04	0.10	1.272e-02 (1.858e-01)	1.272e-02 (1.858e-01)	1.306e-02 (1.854e-01)	1.306e-02 (1.854e-01)	-2.019e-01 (6.833e-02)	-1.711e-01 (6.833e-02)	-1.172e-01 (6.833e-02)	-6.390e-02 (6.833e-02)
		0.10	0.04	2.596e-02 (1.807e-01)	2.596e-02 (1.807e-01)	2.637e-02 (1.810e-01)	2.637e-02 (1.810e-01)	1.835e-01 (6.944e-02)	1.750e-01 (6.944e-02)	6.518e-02 (6.944e-02)	3.808e-02 (6.944e-02)
		0.01	0.10	-2.792e-02 (1.436e-01)	-2.792e-02 (1.436e-01)	-2.706e-02 (1.434e-01)	-2.706e-02 (1.434e-01)	-3.775e-01 (6.804e-02)	-3.294e-01 (6.804e-02)	-2.096e-01 (6.804e-02)	-1.245e-01 (6.804e-02)
		0.10	0.01	5.394e-02 (1.434e-01)	5.394e-02 (1.434e-01)	5.415e-02 (1.440e-01)	5.415e-02 (1.440e-01)	3.584e-01 (6.930e-02)	3.321e-01 (6.930e-02)	1.573e-01 (6.930e-02)	1.054e-01 (6.930e-02)
	3	0.10	0.10	6.941e-03 (7.498e-02)	6.941e-03 (7.498e-02)	7.124e-03 (7.397e-02)	7.005e-03 (7.392e-02)	-2.316e-02 (6.814e-02)	-5.678e-03 (6.814e-02)	-6.382e-02 (6.814e-02)	-2.468e-02 (6.814e-02)
		0.04	0.10	6.744e-03 (7.458e-02)	6.744e-03 (7.458e-02)	6.889e-03 (7.223e-02)	6.715e-03 (7.218e-02)	-1.845e-01 (6.853e-02)	-1.191e-01 (6.853e-02)	-1.001e-01 (6.853e-02)	-3.492e-02 (6.853e-02)
		0.10	0.04	7.093e-03 (7.500e-02)	7.093e-03 (7.500e-02)	7.275e-03 (7.485e-02)	7.211e-03 (7.481e-02)	1.389e-01 (6.662e-02)	1.078e-01 (6.662e-02)	-2.137e-02 (6.662e-02)	-1.398e-02 (6.662e-02)
		0.01	0.10	2.872e-03 (6.799e-02)	2.872e-03 (6.799e-02)	2.998e-03 (6.386e-02)	3.260e-03 (6.467e-02)	-3.314e-01 (6.781e-02)	-2.273e-01 (6.781e-02)	-1.456e-01 (6.781e-02)	-5.709e-02 (6.781e-02)
		0.10	0.01	1.167e-02 (6.616e-02)	1.167e-02 (6.616e-02)	1.182e-02 (6.637e-02)	1.191e-02 (6.589e-02)	2.865e-01 (6.495e-02)	2.161e-01 (6.495e-02)	4.454e-02 (6.495e-02)	1.657e-02 (6.495e-02)
	4	0.10	0.10	-5.174e-03 (5.455e-02)	-5.174e-03 (5.455e-02)	-4.095e-03 (4.704e-02)	3.811e-03 (7.950e-02)	-1.384e-02 (6.987e-02)	-4.374e-03 (6.987e-02)	-4.636e-02 (6.987e-02)	-2.729e-02 (6.987e-02)
		0.04	0.10	-5.517e-03 (5.875e-02)	-5.517e-03 (5.875e-02)	-3.503e-03 (4.392e-02)	3.563e-03 (7.139e-02)	-1.483e-01 (7.084e-02)	-8.335e-02 (7.084e-02)	-6.311e-02 (7.084e-02)	-3.113e-02 (7.084e-02)
		0.10	0.04	-4.779e-03 (5.032e-02)	-4.779e-03 (5.032e-02)	-4.372e-03 (4.781e-02)	3.537e-03 (8.441e-02)	1.202e-01 (6.615e-02)	7.311e-02 (6.615e-02)	-2.695e-02 (6.615e-02)	-2.615e-02 (6.615e-02)
		0.01	0.10	-6.962e-03 (6.027e-02)	-6.962e-03 (6.027e-02)	-2.735e-03 (3.872e-02)	1.300e-03 (5.887e-02)	-2.729e-01 (6.844e-02)	-1.683e-01 (6.844e-02)	-8.916e-02 (6.844e-02)	-4.221e-02 (6.844e-02)
		0.10	0.01	-3.794e-03 (4.529e-02)	-3.794e-03 (4.529e-02)	-3.759e-03 (4.520e-02)	8.686e-03 (7.561e-02)	2.441e-01 (6.099e-02)	1.556e-01 (6.099e-02)	2.116e-02 (6.099e-02)	-8.105e-03 (6.099e-02)
9000	1	0.10	0.10	1.950e-03 (2.033e-02)	1.950e-03 (2.033e-02)	1.733e-03 (1.868e-02)	1.733e-03 (1.868e-02)	-1.753e-02 (4.902e-02)	-4.793e-04 (4.902e-02)	-3.117e-02 (4.902e-02)	-8.314e-03 (4.902e-02)
		0.04	0.10	1.715e-03 (1.875e-02)	1.715e-03 (1.875e-02)	1.247e-03 (1.462e-02)	1.247e-03 (1.462e-02)	-8.118e-02 (4.564e-02)	-1.681e-02 (4.564e-02)	-4.462e-02 (4.564e-02)	-1.234e-02 (4.564e-02)
		0.10	0.04	2.091e-03 (2.114e-02)	2.091e-03 (2.114e-02)	2.068e-03 (2.088e-02)	2.068e-03 (2.088e-02)	4.517e-02 (4.778e-02)	1.073e-02 (4.778e-02)	-1.862e-02 (4.778e-02)	-4.511e-03 (4.778e-02)
		0.01	0.10	1.427e-03 (1.681e-02)	1.427e-03 (1.681e-02)	7.442e-04 (9.525e-03)	7.442e-04 (9.525e-03)	-1.569e-01 (3.948e-02)	-6.386e-02 (3.948e-02)	-5.644e-02 (3.948e-02)	-1.570e-02 (3.948e-02)
		0.10	0.01	2.131e-03 (2.112e-02)	2.131e-03 (2.112e-02)	2.197e-03 (2.121e-02)	2.197e-03 (2.121e-02)	1.198e-01 (4.156e-02)	5.023e-02 (4.156e-02)	-7.944e-04 (4.156e-02)	-2.225e-04 (4.156e-02)
	2	0.10	0.10	1.328e-02 (1.623e-01)	1.328e-02 (1.623e-01)	1.315e-02 (1.623e-01)	1.315e-02 (1.623e-01)	-1.377e-03 (7.142e-02)	-5.681e-03 (7.142e-02)	-2.065e-02 (7.142e-02)	-2.389e-03 (7.142e-02)
		0.04	0.10	1.069e-02 (1.580e-01)	1.069e-02 (1.580e-01)	1.047e-02 (1.582e-01)	1.047e-02 (1.582e-01)	-1.897e-01 (7.162e-02)	-1.688e-01 (7.162e-02)	-1.017e-01 (7.162e-02)	-3.766e-02 (7.162e-02)
		0.10	0.04	1.465e-02 (1.576e-01)	1.465e-02 (1.576e-01)	1.473e-02 (1.576e-01)	1.473e-02 (1.576e-01)	1.870e-01 (7.104e-02)	1.577e-01 (7.104e-02)	6.202e-02 (7.104e-02)	2.678e-02 (7.104e-02)
		0.01	0.10	-1.992e-02 (1.231e-01)	-1.992e-02 (1.231e-01)	-2.019e-02 (1.232e-01)	-2.019e-02 (1.232e-01)	-3.613e-01 (7.148e-02)	-3.173e-01 (7.148e-02)	-1.902e-01 (7.148e-02)	-9.123e-02 (7.148e-02)
		0.10	0.01	3.930e-02 (1.231e-01)	3.930e-02 (1.231e-01)	3.938e-02 (1.233e-01)	3.938e-02 (1.233e-01)	3.580e-01 (7.103e-02)	3.065e-01 (7.103e-02)	1.512e-01 (7.103e-02)	8.005e-02 (7.103e-02)
	3	0.10	0.10	2.397e-03 (6.098e-02)	2.397e-03 (6.098e-02)	1.863e-03 (6.019e-02)	3.068e-03 (6.126e-02)	-1.002e-02 (6.973e-02)	-1.150e-02 (6.973e-02)	-5.147e-02 (6.973e-02)	-1.403e-02 (6.973e-02)
		0.04	0.10	2.445e-03 (6.058e-02)	2.445e-03 (6.058e-02)	1.599e-03 (5.880e-02)	2.855e-03 (5.995e-02)	-1.569e-01 (6.922e-02)	-1.035e-01 (6.922e-02)	-7.491e-02 (6.922e-02)	-1.776e-02 (6.922e-02)
		0.10	0.04	2.336e-03 (6.106e-02)	2.336e-03 (6.106e-02)	2.100e-03 (6.086e-02)	3.256e-03 (6.187e-02)	1.371e-01 (6.837e-02)	8.269e-02 (6.837e-02)	-2.065e-02 (6.837e-02)	-1.017e-02 (6.837e-02)
		0.01	0.10	9.677e-04 (5.716e-02)	9.677e-04 (5.716e-02)	2.694e-04 (5.483e-02)	1.376e-03 (5.571e-02)	-2.922e-01 (6.657e-02)	-1.948e-01 (6.657e-02)	-1.063e-01 (6.657e-02)	-3.252e-02 (6.657e-02)
		0.10	0.01	4.377e-03 (5.672e-02)	4.377e-03 (5.672e-02)	4.380e-03 (5.674e-02)	5.633e-03 (5.750e-02)	2.723e-01 (6.590e-02)	1.774e-01 (6.590e-02)	3.402e-02 (6.590e-02)	8.245e-03 (6.590e-02)
	4	0.10	0.10	-2.051e-03 (4.256e-02)	-2.051e-03 (4.256e-02)	-2.205e-03 (3.461e-02)	3.938e-03 (5.548e-02)	-9.784e-03 (5.840e-02)	-3.148e-03 (5.840e-02)	-2.175e-02 (5.840e-02)	-1.525e-02 (5.840e-02)
		0.04	0.10	-2.013e-03 (4.674e-02)	-2.013e-03 (4.674e-02)	-1.724e-03 (3.156e-02)	3.306e-03 (5.035e-02)	-1.106e-01 (5.729e-02)	-4.406e-02 (5.729e-02)	-2.639e-02 (5.729e-02)	-1.596e-02 (5.729e-02)
		0.10	0.04	-2.055e-03 (3.824e-02)	-2.055e-03 (3.824e-02)	-2.376e-03 (3.538e-02)	4.088e-03 (5.791e-02)	9.079e-02 (5.587e-02)	3.604e-02 (5.587e-02)	-1.490e-02 (5.587e-02)	-1.491e-02 (5.587e-02)
		0.01	0.10	-2.914e-03 (4.809e-02)	-2.914e-03 (4.809e-02)	-1.077e-03 (2.670e-02)	2.347e-03 (4.301e-02)	-2.089e-01 (5.336e-02)	-1.062e-01 (5.336e-02)	-4.370e-02 (5.336e-02)	-2.249e-02 (5.336e-02)
		0.10	0.01	-2.030e-03 (3.429e-02)	-2.030e-03 (3.429e-02)	-2.191e-03 (3.388e-02)	5.208e-03 (5.417e-02)	1.912e-01 (5.065e-02)	9.526e-02 (5.065e-02)	1.063e-02 (5.065e-02)	-7.367e-03 (5.065e-02)
12000	1	0.10	0.10	1.409e-03 (1.698e-02)	1.409e-03 (1.698e-02)	9.964e-04 (1.535e-02)	9.964e-04 (1.535e-02)	-1.591e-02 (4.488e-02)	-1.592e-03 (4.488e-02)	-2.549e-02 (4.488e-02)	-6.911e-03 (4.488e-02)
		0.04	0.10	1.314e-03 (1.583e-02)	1.314e-03 (1.583e-02)	6.658e-04 (1.203e-02)	6.658e-04 (1.203e-02)	-6.779e-02 (4.124e-02)	-1.191e-02 (4.124e-02)	-3.552e-02 (4.124e-02)	-9.131e-03 (4.124e-02)
		0.10	0.04	1.443e-03 (1.753e-02)	1.443e-03 (1.753e-02)	1.259e-03 (1.715e-02)	1.259e-03 (1.715e-02)	3.373e-02 (4.411e-02)	4.516e-03 (4.411e-02)	-1.387e-02 (4.411e-02)	-4.547e-03 (4.411e-02)
		0.01	0.10	1.178e-03 (1.441e-02)	1.178e-03 (1.441e-02)	3.627e-04 (7.870e-03)	3.627e-04 (7.870e-03)	-1.358e-01 (3.516e-02)	-5.264e-02 (3.516e-02)	-4.520e-02 (3.516e-02)	-1.208e-02 (3.516e-02)
		0.10	0.01	1.416e-03 (1.743e-02)	1.416e-03 (1.743e-02)	1.395e-03 (1.745e-02)	1.395e-03 (1.745e-02)	1.000e-01 (3.755e-02)	3.792e-02 (3.755e-02)	1.607e-03 (3.755e-02)	-1.179e-03 (3.755e-02)
	2	0.10	0.10	1.158e-02 (1.359e-01)	1.158e-02 (1.359e-01)	1.158e-02 (1.364e-01)	1.158e-02 (1.364e-01)	-8.054e-03 (7.086e-02)	-1.035e-02 (7.086e-02)	2.389e-02 (7.086e-02)	1.432e-02 (7.086e-02)
		0.04	0.10	1.096e-02 (1.353e-01)	1.096e-02 (1.353e-01)	1.072e-02 (1.357e-01)	1.072e-02 (1.357e-01)	-1.919e-01 (7.087e-02)	-1.654e-01 (7.087e-02)	-6.109e-02 (7.087e-02)	-1.271e-02 (7.087e-02)
		0.10	0.04	1.159e-02 (1.345e-01)	1.159e-02 (1.345e-01)	1.173e-02 (1.348e-01)	1.173e-02 (1.348e-01)	1.755e-01 (7.059e-02)	1.457e-01 (7.059e-02)	1.041e-01 (7.059e-02)	3.584e-02 (7.059e-02)
		0.01	0.10	-1.139e-02 (1.077e-01)	-1.139e-02 (1.077e-01)	-1.164e-02 (1.084e-01)	-1.164e-02 (1.084e-01)	-3.587e-01 (7.056e-02)	-3.069e-01 (7.056e-02)	-1.572e-01 (7.056e-02)	-6.909e-02 (7.056e-02)
		0.10	0.01	3.050e-02 (1.062e-01)	3.050e-02 (1.062e-01)	3.062e-02 (1.064e-01)	3.062e-02 (1.064e-01)	3.423e-01 (7.016e-02)	2.887e-01 (7.016e-02)	1.850e-01 (7.016e-02)	8.178e-02 (7.016e-02)
	3	0.10	0.10	3.016e-03 (5.159e-02)	3.016e-03 (5.159e-02)	2.844e-03 (5.143e-02)	2.827e-03 (5.161e-02)	-1.473e-02 (6.407e-02)	-1.072e-02 (6.407e-02)	-3.005e-02 (6.407e-02)	-1.333e-02 (6.407e-02)
		0.04	0.10	3.001e-03 (5.113e-02)	3.001e-03 (5.113e-02)	2.674e-03 (5.034e-02)	2.549e-03 (5.056e-02)	-1.515e-01 (6.344e-02)	-8.912e-02 (6.344e-02)	-4.872e-02 (6.344e-02)	-1.899e-02 (6.344e-02)
		0.10	0.04	3.014e-03 (5.178e-02)	3.014e-03 (5.178e-02)	2.982e-03 (5.189e-02)	3.070e-03 (5.207e-02)	1.224e-01 (6.262e-02)	6.903e-02 (6.262e-02)	-7.561e-03 (6.262e-02)	-7.796e-03 (6.262e-02)
		0.01	0.10	2.547e-03 (4.965e-02)	2.547e-03 (4.965e-02)	2.125e-03 (4.805e-02)	1.796e-03 (4.811e-02)	-2.776e-01 (6.052e-02)	-1.721e-01 (6.052e-02)	-8.195e-02 (6.052e-02)	-3.124e-02 (6.052e-02)
		0.10	0.01	3.762e-03 (4.999e-02)	3.762e-03 (4.999e-02)	3.828e-03 (5.010e-02)	3.975e-03 (5.038e-02)	2.495e-01 (6.023e-02)	1.554e-01 (6.023e-02)	3.820e-02 (6.023e-02)	1.050e-02 (6.023e-02)
	4	0.10	0.10	-4.057e-03 (3.607e-02)	-4.057e-03 (3.607e-02)	-2.708e-03 (2.639e-02)	-3.695e-04 (4.183e-02)	-9.619e-03 (5.381e-02)	-5.266e-03 (5.381e-02)	-1.769e-02 (5.381e-02)	-1.308e-02 (5.381e-02)
		0.04	0.10	-4.408e-03 (4.073e-02)	-4.408e-03 (4.073e-02)	-1.953e-03 (2.335e-02)	-5.318e-04 (3.713e-02)	-8.206e-02 (5.372e-02)	-2.649e-02 (5.372e-02)	-2.318e-02 (5.372e-02)	-1.704e-02 (5.372e-02)
		0.10	0.04	-3.632e-03 (3.122e-02)	-3.632e-03 (3.122e-02)	-3.167e-03 (2.741e-02)	-2.810e-04 (4.453e-02)	6.226e-02 (4.992e-02)	1.462e-02 (4.992e-02)	-1.301e-02 (4.992e-02)	-9.809e-03 (4.992e-02)
		0.01	0.10	-4.839e-03 (4.433e-02)	-4.839e-03 (4.433e-02)	-1.060e-03 (1.868e-02)	-5.893e-04 (3.122e-02)	-1.619e-01 (4.953e-02)	-7.613e-02 (4.953e-02)	-3.156e-02 (4.953e-02)	-2.183e-02 (4.953e-02)
		0.10	0.01	-3.184e-03 (2.677e-02)	-3.184e-03 (2.677e-02)	-3.208e-03 (2.634e-02)	-3.072e-05 (4.422e-02)	1.416e-01 (4.282e-02)	6.304e-02 (4.282e-02)	-1.460e-03 (4.282e-02)	-6.661e-03 (4.282e-02)

Appendix B

APPENDIX FOR CHAPTER 3

In the main discussion, we have established an estimation and inference framework for doubly high-dimensional linear mixed models (LMMs) in the context of neighborhood-based graphical modeling of heterogeneous data. In that specific setting, the fixed and the random effects had identical design matrices. However, doubly high-dimensional LMMs may also arise from settings where the two matrices are not identical, but yet share a set of variables or have highly correlated columns, therefore still violating assumptions on the conditional independence of the fixed and random design matrices. A typical example is in longitudinal data analysis with random slope models — if we want to allow random slopes for a large number of explanatory variables, we end up with a doubly high-dimensional LMM. With advancing technology allowing us to collect more variables than ever, such examples will only become more ubiquitous. Thus, we extend our framework to a generic doubly high-dimensional LMM, formulated as:

$$y^i = X^i\beta + Z^i\gamma_i + \epsilon_i, \quad i = 1, \ldots, n. \tag{B.1}$$

Here, each $y^i \in \mathbb{R}^m$ is the observation vector, $X^i \in \mathbb{R}^{m\times p}$ and $Z^i \in \mathbb{R}^{m\times q}$ are the design matrices, where $X^i \mid \Sigma_X^i \sim MN_{m\times p}(0, I_m, \Sigma_X^i)$ and $Z^i \mid \Sigma_Z^i \sim MN_{m\times q}(0, I_m, \Sigma_Z^i)$. The subject-level covariance matrices Σ_X^i and Σ_Z^i are centered at the population-level covariance matrices Σ_X, Σ_Z, respectively. We assume the q random effect covariates are a subset of the p fixed effect covariates. Moreover, we assume that conditional on X^i, the random effect coefficients $\gamma_i \in \mathbb{R}^q$ and the noise term $\epsilon_i \in \mathbb{R}^m$ are independent with variance Ψ and $\mathbb{R}^i$, and satisfy $\gamma_i \in \mathrm{SGV}(c_1\|\Psi\|_2)$, $\epsilon_i \in \mathrm{SGV}(c_2\|R^i\|_2)$, respectively. We allow for either $q > c_0 m$ or $m > c_0 q$, for some constant $c_0 > 1$. The fixed effect coefficients β has support S with cardinality s.

We present the estimators, the inference framework and their theoretical properties in the Appendix B.1–B.4. All results are directly applicable to the doubly high-dimensional LMM in (3.2) of the main text in the context of graphical model selection, where we simply need to set $X^i = Z^i$.

We define some relevant quantities to facilitate the discussion. We use $\mathbf{1}\{\cdot\}$ to represent the indicator function. The proxy covariance matrix is denoted by $\Sigma_a^i = aZ^i(Z^i)^\top + I_m$ and $\Sigma_a = \text{diag}\left(\{\Sigma_a^i\}_{i=1}^n\right)$. The vectors/matrices y, γ, ϵ and X are formed by vertically stacking the vectors/matrices y^i's, γ_i's, ϵ_i's and X^i's respectively. A random variable X is sub-exponential with parameters (a, b) if $\forall\ |t| \leq 1/b$, $\mathbb{E}(\exp(tX)) \leq \exp(at^2/2)$. We define $\text{SE}(a, b)$ as the class of all sub-exponential random variables with mean zero and parameters (a, b). Denote e_j as a length q unite vector with the jth entry taking value 1.

B.1 Fixed Effect Estimator $\hat{\beta}$

We define the estimator $\hat{\beta}$ for the fixed effect coefficients β in model (B.1) as follows:

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\text{argmin}} \left(2\,\text{tr}(\Sigma_a^{-1})\right)^{-1} \|\Sigma_a^{-1/2}(y - X\beta)\|_2^2 + \lambda_a \|\beta\|_1.$$

We state a generalized version of Theorem 3.3.1 of the main text and its related assumptions for the generic doubly high-dimensional LMM in (B.1):

Assumption 8. *1. Let $q > c_0 m$ or $m > c_0 q$ for some constant $c_0 > 1$ and let $m \vee q > c_1$ for some suitably large constant $c_1 > 0$. Moreover, let $\log(q)\,(q/m)^{\mathbf{1}\{q > c_0 m\}}/n = o(1)$.*

2. $\forall\ i$, $\sigma(\Sigma_X) \asymp \sigma(R^i) \asymp \|\Psi\|_2 \asymp 1$, $\|\Sigma_X^i - \Sigma_X\|_2 \leq \sigma_{\min}(\Sigma_X) - c_2$, for some constant $c_2 > 0$.

Assumption 8.2 implies $\sigma(\Sigma_X^i) \asymp 1$ for all $i = 1, \ldots, n$ (Weyl's theorem, [15]).

Assumption 9. *1. $\Psi = \text{diag}(\psi)$ for a vector $\psi \in \mathbb{R}^q$. The support of ψ is S_ψ with cardinality $s_\psi < c_2 m \wedge n$ for some constant $c_2 > 0$, and $\min(\psi_{S_\psi}) \asymp \max(\psi_{S_\psi}) \asymp 1$.*

2. $\sigma_{\min}(\Psi) \asymp 1$.

Theorem B.1.1 (Fixed effect estimator consistency)**.** *Under Assumption 8.1 and Assumption 8.2, with probability at least* $1 - 4\exp\{-cn\} - 12\exp\{-c\log(n)\} - 2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - \exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\}$*, we have the following results:*

1. *When* $q > c_0 m$*: Taking* $\lambda_a = c_1\sqrt{q\log(p)/(nm)}$ *for a suitably large* $c_1 > 0$*, we have that:*

$$\|\hat{\beta} - \beta^*\|_2 = O_p\left(\sqrt{\frac{sq\log(p)}{mn}}\right),$$
$$\|\hat{\beta} - \beta^*\|_1 = O_p\left(s\sqrt{\frac{q\log(p)}{mn}}\right),$$
$$\left\|\Sigma_a^{-1/2}X(\hat{\beta} - \beta^*)\right\|_2^2 = O_p\left(s\log(p)\right).$$

2. *When* $q > c_0 m$ *and Assumption 9.1 also holds: Taking* $\lambda_a = c_2\sqrt{\log(p)\log^2(n)/n}$ *for suitably large* $c_2 > 0$*, we have that:*

$$\|\hat{\beta} - \beta^*\|_2 = O_p\left(\sqrt{\frac{s\log^2(n)\log(p)}{n}}\right),$$
$$\|\hat{\beta} - \beta^*\|_1 = O_p\left(s\sqrt{\frac{\log^2(n)\log(p)}{n}}\right),$$
$$\|\Sigma_a^{-1/2}X(\hat{\beta} - \beta^*)\|_2^2 = O_p\left(\frac{sm\log^2(n)\log(p)}{q}\right).$$

3. *When* $m > c_0 q$ *and* $p = q$*: Taking* $\lambda_a = c_3\sqrt{\log(p)/(nm^2)}$ *for suitably large* $c_3 > 0$*, we have that:*

$$\|\hat{\beta} - \beta^*\|_2 = O_p\left(\sqrt{\frac{s\log(p)}{n}}\right),$$
$$\|\hat{\beta} - \beta^*\|_1 = O_p\left(s\sqrt{\frac{\log(p)}{n}}\right),$$
$$\|\Sigma_a^{-1/2}X(\hat{\beta} - \beta^*)\|_2^2 = O_p\left(s\log(p)\right).$$

4. *When $m > c_0 q$ and $p > q$: Taking $\lambda_a = c_4\sqrt{\log(p)/(nm)}$ for suitably large $c_4 > 0$, we have that:*

$$\|\hat{\beta} - \beta^*\|_2 = O_p\left(\sqrt{\frac{sm\log(p)}{n}}\right),$$
$$\|\hat{\beta} - \beta^*\|_1 = O_p\left(s\sqrt{\frac{m\log(p)}{n}}\right),$$
$$\|\Sigma_a^{-1/2} X(\hat{\beta} - \beta^*)\|_2^2 = O_p\left(sm\log(p)\right).$$

B.1.1 Related lemmas for Theorem B.1.1

Lemma B.1.2 (Core Lemma)**.** *Assume $q > c_0 m$ or $m > c_0 q$ for some constant $c_0 > 1$. Z is a $m \times q$ matrix with entries independently following the $N(0,1)$ distribution, and Z^i, $i = 1, \ldots, n$ are identical copies of Z. Then the following properties hold for the non-zero singular values $\sigma(Z)$ of Z and $\sigma(Z^i)$ of Z^i's:*

1. $|\sqrt{q} - \sqrt{m}| \leq \mathbb{E}(\sigma(Z)) \leq \sqrt{m} + \sqrt{q}$, $\mathbb{E}(\sigma(Z)) \asymp \sqrt{m} \vee \sqrt{q}$.
2. $\sigma(Z) - \mathbb{E}(\sigma(Z)) \in \mathrm{SG}(1)$.
3. $\mathbb{E}(\sigma^2(Z)) \in [\mathbb{E}(\sigma(Z))^2, \mathbb{E}(\sigma(Z))^2 + 1]$, $\mathbb{E}(\sigma^2(Z)) \asymp m \vee q$.
4. $\sigma^2(Z) - \mathbb{E}(\sigma^2(Z)) \in \mathrm{SE}(32, 4)$. $\sum_{i=1}^n \sigma^2(Z^i) \asymp n(m \vee q)$ *with probability at least* $1 - 2\exp\{-c_1 n(m \vee q)\}$, *for some* $c_1 > 0$.

Further assume $m \vee q > c_2 > 0$ for some suitably large constant c_2. Denote $\Sigma_a^i = aZ^i(Z^i)^\top + I_m$, where a is a positive constant. Then we have the following properties hold for any constant $c > 0$, for positive constants $c_3, c_4, \ldots$:

5. $\frac{1}{\mathbb{E}(\sigma^2(Z))+c} \leq \mathbb{E}\left(\frac{1}{\sigma^2(Z)+c}\right) \leq \frac{4}{\mathbb{E}(\sigma^2(Z))+c}$.
6. $\frac{1}{\sigma^2(Z)+c} - \mathbb{E}\left(\frac{1}{\sigma^2(Z)+c}\right) \in \mathrm{SE}\left(c_3\,\mathbb{E}\left(\frac{1}{\sigma^2(Z)+c}\right)^2, c_3\,\mathbb{E}\left(\frac{1}{\sigma^2(Z)+c}\right)\right)$.
7. $\sum_{i=1}^n \frac{1}{\sigma^2(Z^i)+c} \asymp \frac{n}{m\vee q}$ *with probability at least* $1 - 2\exp\{-c_4 n\}$. $\max_{1\leq i\leq n} \frac{1}{\sigma^2(Z^i)+c} \leq \frac{1}{c} \wedge \frac{c_5\log(n)}{m\vee q}$ *with probability at least* $1 - 2\exp\{-c_6\log(n)\}$, $\min_{1\leq i\leq n} \frac{1}{\sigma^2(Z^i)+c} \geq \frac{c_7}{\log(n)+m\vee q}$ *with probability at least* $1 - 2\exp\{-c_6\log(n)\}$.

8. $\sum_{i=1}^{n} \operatorname{tr}\left((\Sigma_a^i)^{-1}\right) \asymp mnq^{-\mathbf{1}\{q>c_0m\}}$ *with probability at least* $1-4\exp\{-c_8 n\}$.
   When $q > c_0 m$: $\frac{c_9 m}{\log(n)+q} \le \min_i \operatorname{tr}\left((\Sigma_a^i)^{-1}\right) \le \max_i \operatorname{tr}\left((\Sigma_a^i)^{-1}\right) \le c_{10} m\left(1 \wedge \frac{\log(n)}{q}\right)$ *with probability at least* $1-4\exp\{-c_{11}\log(n)\}$.
   When $m > c_0 q$: $c_{12}\left(m + \frac{q}{m+\log(n)}\right) \le \min_i \operatorname{tr}\left((\Sigma_a^i)^{-1}\right) \le \max_i \operatorname{tr}\left((\Sigma_a^i)^{-1}\right) \le c_{13}\left(m + q\left(1 \wedge \frac{\log(n)}{m}\right)\right)$ *with probability at least* $1-4\exp\{-c_{14}\log(n)\}$.

9. $\mathbb{E}\left(Z^\top (aZZ^\top + I_m)^{-1} Z\right) = kI_q$, $k \asymp (m/q)^{\mathbf{1}\{q>c_0m\}}$. $\sigma\left(\sum_{i=1}^{n} (Z^i)^\top (\Sigma_a^i)^{-1} Z^i\right) \asymp n\,(m/q)^{\mathbf{1}\{q>c_0m\}}$ *with probability at least* $1-\exp\left\{\log(q) - c_{16} n\,(m/q)^{\mathbf{1}\{q>c_0m\}}\right\}$.

10. *Assume* $\log(q)\,(q/m)^{\mathbf{1}\{q>c_0m\}}/n = o(1)$. *Then with probability at least* $1 - 2\exp\{-c_{18}n\} - 2\exp\{-c_{18}\log(n)\} - \exp\left\{-c_{18} n\,(m/q)^{\mathbf{1}\{q>c_0m\}}\right\}$:

$$\sigma\left(\sum_{i=1}^{n} (Z^i)^\top (\Sigma_a^i)^{-2} Z^i\right) \le \begin{cases} c_{17}\frac{n}{q}\left(1 \wedge \frac{m\log(n)}{q}\right) & , q > c_0 m, \\ c_{17}\frac{n}{m} & , m > c_0 q. \end{cases}$$

Lemma B.1.3. *1. Under Assumption 8.1 and Assumption 8.2, with probability at least* $1 - 4\exp\{-cn\} - 2\exp\{-c\log(n)\} - 2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - \exp\left\{-cn\,(m/q)^{\mathbf{1}\{q>c_0m\}}\right\}$, *we have:*

$$\begin{cases} \sigma(X^\top \Sigma_a^{-1} X) \asymp \frac{mn}{q} & , q > c_0 m, \\ \sigma(X^\top \Sigma_a^{-1} X) \asymp n & , m > c_0 q \text{ and } p = q, \\ c_1 n \le \sigma(X^\top \Sigma_a^{-1} X) \le c_2 mn & , m > c_0 q \text{ and } q < p. \end{cases}$$

2. *Under Assumption 8.1 and Assumption 8.2,* $\max_i \sigma((X^i)^\top (\Sigma_a^i)^{-1} X^i) \le c_1$ *when* $p = q$; *when* $p > q$, *with probability at least* $1 - 8\exp\{-c_2\log(n)\}$ *we have:*

$$\max_i \sigma((X^i)^\top (\Sigma_a^i)^{-1} X^i) \le \begin{cases} c_3\left(1 \vee \frac{m\log^2(n)}{q}\right) & , q > c_0 m, \\ c_3 m \log^2(n) & , m > c_0 q. \end{cases}$$

Lemma B.1.4. *1. Under Assumption 8.1 and Assumption 8.2, we have*

$$\max_{1\le j\le p}\sum_{i=1}^{n}\left\|(Z^i)^\top\left(\Sigma_a^i\right)^{-1}X_j^i\right\|_2^2\|\Psi\|_2+\left\|\left(\Sigma_a^i\right)^{-1}X_j^i\right\|_2^2\|R^i\|_2\le\begin{cases}c_1\frac{nm}{q} & ,q>c_0m,\\ c_1n & ,m>c_0q \text{ and } q=p,\\ c_1mn & ,m>c_0q \text{ and } q<p,\end{cases}$$

with probability at least $1-4\exp\{-cn\}-12\exp\{-c\log(n)\}-2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}-\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\}$.

If we additionally assume Assumption 9.1, we have

$$\max_{1\le j\le p}\sum_{i=1}^{n}\left\|(Z^i_{S_\psi})^\top\left(\Sigma_a^i\right)^{-1}X_j^i\right\|_2^2\|\Psi\|_2+\left\|\left(\Sigma_a^i\right)^{-1}X_j^i\right\|_2^2\|R^i\|_2\le\begin{cases}\frac{c_1mn}{q}\left(1\wedge\frac{m\log^2(n)}{q}\right) & ,q>c_0m,\\ c_1n & ,m>c_0q \text{ and}\\ c_1mn & ,m>c_0q \text{ and}\end{cases}$$

with probability at least $1-4\exp\{-cn\}-6\exp\{-c\log(n)\}-2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}-\exp\left\{-cn\,(m/q)^{\mathbf{1}\{q>c_0m\}}\right\}$.

2. Define

$$z_0^*=\max_{1\le j\le p}\left|\frac{1}{\mathrm{tr}(\Sigma_a^{-1})}X_j^\top\Sigma_a^{-1}(y-X\beta^*)\right|.$$

Under Assumption 8.1 and Assumption 8.2, we have

$$z_0^*\le\begin{cases}c_1\sqrt{\frac{q\log(p)}{nm}} & ,q>c_0m,\\ c_1\sqrt{\frac{\log(p)\log^2(n)}{n}} & ,q>c_0m \text{ and Assumption 9.1alsoholds},\\ c_1\sqrt{\frac{\log(p)}{nm^2}} & ,m>c_0q \text{ and } p=q,\\ c_1\sqrt{\frac{\log(p)}{nm}} & ,m>c_0q \text{ and } p>q.\end{cases}$$

with probability at least $1-4\exp\{-cn\}-12\exp\{-c\log(n)\}-2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}-\exp\left\{-cn\,(m/q)^{\mathbf{1}\{q>c_0m\}}\right\}$.

B.1.2 Proof of Theorem B.1.1

Proof.

Based on the definition of $\hat{\beta}$, we can get:

$$\frac{1}{2\,\mathrm{tr}(\Sigma_a^{-1})}\left\|\Sigma_a^{-1/2}(y-X\hat{\beta})\right\|_2^2+\lambda_a\|\hat{\beta}\|_1\leq\frac{1}{2\,\mathrm{tr}(\Sigma_a^{-1})}\left\|\Sigma_a^{-1/2}(y-X\beta^*)\right\|_2^2+\lambda_a\|\beta^*\|_1.$$

Denoting $\hat{u}=\hat{\beta}-\beta^*$ and rearranging the terms, we get

$$0\leq\frac{1}{2\,\mathrm{tr}(\Sigma_a^{-1})}\left\|\Sigma_a^{-1/2}X\hat{u}\right\|_2^2\leq(z_0^*+\lambda_a)\|\hat{u}_S\|_1+(z_0^*-\lambda_a)\|\hat{u}_{S^c}\|_1, \tag{B.2}$$

where

$$z_0^*:=\max_{1\leq j\leq p}\left|\frac{1}{\mathrm{tr}(\Sigma_a^{-1})}X_j^\top\Sigma_a^{-1}(y-X\beta^*)\right|.$$

Based on Lemma B.1.3.1 and Lemma B.1.2.8, we have that:

$$\frac{1}{2\,\mathrm{tr}(\Sigma_a^{-1})}\left\|\Sigma_a^{-1/2}X\hat{u}\right\|_2^2\geq\frac{\|\hat{u}\|_2^2}{2\,\mathrm{tr}(\Sigma_a^{-1})}\sigma_{\min}(X^\top\Sigma_a^{-1}X)\geq\begin{cases}c_1\|\hat{u}\|_2^2 & ,q>c_0m,\\ c_1\frac{1}{m}\|\hat{u}\|_2^2 & ,m>c_0q.\end{cases}$$

Based on Lemma B.1.4.2:

$$z_0^*\leq\begin{cases}c_1\sqrt{\frac{q\log(p)}{nm}} & ,q>c_0m,\\ c_1\sqrt{\frac{\log(p)\log^2(n)}{n}} & ,q>c_0m \text{ and assume Assumption 9.1},\\ c_1\sqrt{\frac{\log(p)}{nm^2}} & ,m>c_0q \text{ and } p=q,\\ c_1\sqrt{\frac{\log(p)}{nm}} & ,m>c_0q \text{ and } p>q.\end{cases}$$

Thus with probability at least $1-4\exp\{-cn\}-12\exp\{-c\log(n)\}-2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}-\exp\left\{-cn\,(m/q)^{\mathbf{1}\{q>c_0m\}}\right\}$, we can get the following results:

1. When $q>c_0m$: With $\lambda_a=c_2\sqrt{\frac{q\log(p)}{nm}}$ for suitably large c_2, we have $z_0^*\leq\lambda_a/2$ with

high probability. Then based on (B.2) we have

$$c_1\|\hat{u}\|_2^2 \leq 3\lambda_a\|\hat{u}_S\|_1 - \lambda_a\|\hat{u}_{S^c}\|_1 \leq 3\lambda_a\|\hat{u}_S\|_2^2 \leq 3\lambda_a\sqrt{s}\|\hat{u}\|_2^2.$$

Thus we can obtain

$$\begin{aligned}
\|\hat{u}\|_2 &\leq O_p\left(\sqrt{\frac{sq\log(p)}{mn}}\right), \\
\|\hat{u}\|_1 &\leq O_p\left(s\sqrt{\frac{q\log(p)}{mn}}\right), \\
\|\Sigma_a^{-1/2}X\hat{u}\|_2^2 &\leq O_p\left(s\log(p)\right).
\end{aligned}$$

2. When $q > c_0 m$ and additionally assume Assumption 9.1: With $\lambda_a = c_2\sqrt{\frac{\log(p)\log^2(n)}{n}}$ for suitably large c_2, we have $z_0^* \leq \lambda_a/2$ with high probability. Then following similar arguments as before, we can obtain

$$\begin{aligned}
\|\hat{u}\|_2 &\leq O_p\left(\sqrt{\frac{s\log^2(n)\log(p)}{n}}\right), \\
\|\hat{u}\|_1 &\leq O_p\left(s\sqrt{\frac{\log^2(n)\log(p)}{n}}\right), \\
\|\Sigma_a^{-1/2}X\hat{u}\|_2^2 &\leq O_p\left(\frac{sm\log^2(n)\log(p)}{q}\right).
\end{aligned}$$

3. When $m > c_0 q$ and $p = q$: With $\lambda_a = c_2\sqrt{\frac{\log(p)}{nm^2}}$ for suitably large c_2, we have $z_0^* \leq \lambda_a/2$ with high probability. Then following similar arguments as before, we can obtain

$$\begin{aligned}
\|\hat{u}\|_2 &\leq O_p\left(\sqrt{\frac{s\log(p)}{n}}\right), \\
\|\hat{u}\|_1 &\leq O_p\left(s\sqrt{\frac{\log(p)}{n}}\right), \\
\|\Sigma_a^{-1/2}X\hat{u}\|_2^2 &\leq O_p\left(s\log(p)\right).
\end{aligned}$$

4. When $m > c_0 q$ and $p > q$: With $\lambda_a = c_2\sqrt{\frac{\log(p)}{nm}}$ for suitably large c_2, we have $z_0^* \leq \lambda_a/2$ with high probability. Then following similar arguments as before, we can obtain

$$\|\hat{u}\|_2 \leq O_p\left(\sqrt{\frac{sm\log(p)}{n}}\right),$$
$$\|\hat{u}\|_1 \leq O_p\left(s\sqrt{\frac{m\log(p)}{n}}\right),$$
$$\|\Sigma_a^{-1/2}X\hat{u}\|_2^2 \leq O_p\left(sm\log(p)\right).$$

□

B.1.3 Proof of related lemmas for Theorem B.1.1

Proof of Lemma B.1.2.

Lemma B.1.2.1)

By Theorem 5.32 in [223], we have $|\sqrt{q} - \sqrt{m}| \leq \mathbb{E}(\sigma(Z)) \leq \sqrt{q} + \sqrt{m}$. With $q > c_0 m$ or $m > c_0 q$, we have $(\sqrt{q} \vee \sqrt{m})\left(1 - \frac{1}{\sqrt{c_0}}\right) \leq \mathbb{E}(\sigma(Z)) \leq (\sqrt{q} \vee \sqrt{m})\left(1 + \frac{1}{\sqrt{c_0}}\right)$, and thus $\mathbb{E}(\sigma(Z)) \asymp \sqrt{q} \vee \sqrt{m}$.

Lemma B.1.2.2)

By Proposition 5.34 in [223], we have

$$\mathbb{P}\left(|\sigma(Z) - \mathbb{E}(\sigma(Z))| > t\right) \leq 2\exp\{-t^2/2\}.$$

Thus $\sigma(Z) - \mathbb{E}(\sigma(Z)) \in \mathrm{SG}(1)$.

Lemma B.1.2.3)

By Lemma B.1.2.2, we have $\mathrm{Var}(\sigma(Z)) \leq 1$. Then

$$\mathbb{E}(\sigma^2(Z)) = \mathrm{Var}(\sigma(Z)) + \mathbb{E}(\sigma(Z))^2 \leq 1 + \mathbb{E}(\sigma(Z))^2 \asymp m \vee q;$$
$$\mathbb{E}(\sigma^2(Z)) \geq \mathbb{E}(\sigma(Z))^2 \asymp m \vee q.$$

Lemma B.1.2.4)

By Lemma B.1.2.2 and the results in the Appendix B of [97], we have $\sigma^2(Z) - \mathbb{E}(\sigma^2(Z)) \in \mathrm{SE}(32, 4)$. Then for independent copies Z^i's, we have

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} \sigma^2(Z^i) - n\,\mathbb{E}(\sigma^2(Z))\right|\right) \geq 2\exp\left\{-\frac{1}{2}\min\left(\frac{t^2}{32n}, \frac{t}{4}\right)\right\}.$$

Thus by Lemma B.1.2.3, we have $\sum_{i=1}^{n} \sigma^2(Z^i) \asymp n(m \vee q)$ with probability at least $1 - 2\exp\{-c_1 n(m \vee q)\}$.

Lemma B.1.2.5)

By Jensen's inequality,

$$\mathbb{E}\left(\frac{1}{\sigma^2(Z) + c}\right) \geq \frac{1}{\mathbb{E}(\sigma^2(Z)) + c}.$$

For simpler notation, denote $\sigma^2 = \sigma^2(Z)$ and $\mu = \mathbb{E}(\sigma^2(Z))$ in the following proof. For some $0 < k < 1$, with $f(\sigma^2)$ being the probability density function of σ^2, we have

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{\sigma^2 + c}\right) &= \int_0^{k\mu} \frac{1}{\sigma^2 + c} f(\sigma^2)\mathrm{d}\sigma^2 + \int_{k\mu}^{\infty} \frac{1}{\sigma^2 + c} f(\sigma^2)\mathrm{d}\sigma^2 \\
&\leq \int_0^{k\mu} \frac{1}{c} f(\sigma^2)\mathrm{d}\sigma^2 + \int_{k\mu}^{\infty} \frac{1}{k\mu + c} f(\sigma^2)\mathrm{d}\sigma^2 \\
&= \frac{1}{c}\,\mathbb{P}(\sigma^2 < k\mu) + \frac{1}{k\mu + c}\,\mathbb{P}(\sigma^2 \geq k\mu) \\
&\leq \frac{1}{c} 2\exp\left\{-\frac{1}{2}\min\left(\frac{(1-k)^2\mu^2}{32}, \frac{(1-k)\mu}{4}\right)\right\} + \frac{1}{k\mu + c},
\end{aligned}$$

using the fact that $\sigma^2 - \mathbb{E}(\sigma^2) \in \mathrm{SE}(32, 4)$ (Lemma B.1.2.4). Then take $k = 1/2$, we have

$$\begin{aligned}\mathbb{E}\left(\frac{1}{\sigma^2 + c}\right) &\leq \frac{2}{\mu + 2c} + \frac{2}{c}\exp\left\{-\frac{1}{2}\min\left(\frac{\mu^2}{128}, \frac{\mu}{8}\right)\right\} \\ &\leq \frac{4}{\mu + c}\end{aligned}$$

for all $\mu > c_1 > 0$ for some suitably large constant c_1. By Lemma B.1.2.3, it suffices to assume $m \vee q > c_2 > 0$ for some suitably large constant c_2.

Lemma B.1.2.6)

For simpler notation, we denote $\sigma^2 = \sigma^2(Z)$, $\mu = \mathbb{E}(\sigma^2)$ and $E = \mathbb{E}\left(\frac{1}{\sigma^2+c}\right)$ in the following proof. We would like to show that $\forall\ t \in \mathbb{R}$,

$$\mathbb{P}\left(\left|\frac{1}{\sigma^2 + c} - E\right| < t\right) \geq 1 - 2\exp\left\{-\frac{1}{2}\min\left(\frac{t^2}{c_1 E^2}, \frac{t}{c_1 E}\right)\right\}.$$

We consider three cases based on the value of t:

- Case 1: $t \leq E$:

$$\mathbb{P}\left(\left|\frac{1}{\sigma^2 + c} - E\right| < t\right) = \mathbb{P}\left(\frac{1}{E + t} - \mu - c \leq \sigma^2 - \mu \leq \frac{1}{E - t} - \mu - c\right). \tag{B.3}$$

  Denote $t_1 = \mu + c - \frac{1}{E+t}$, $t_2 = \frac{1}{E-t} - \mu - c$. Note that we can show $t_1 \geq 0, t_2 \geq 0$ because of $\frac{1}{\mu+c} \leq E \leq \frac{4}{\mu+c}$ (Lemma B.1.2.5). Then based on the sub-gaussian property of σ^2 (Lemma B.1.2.4) and the equation (B.3), we have

$$\mathbb{P}\left(\left|\frac{1}{\sigma^2 + c} - E\right| < t\right) \geq 1 - \exp\left\{-\frac{1}{2}\min\left(\frac{t_1^2}{32}, \frac{t_1}{4}\right)\right\} - \exp\left\{-\frac{1}{2}\min\left(\frac{t_2^2}{32}, \frac{t_2}{4}\right)\right\}.$$

  We next show that $\forall\ 0 \leq t \leq E$,

$$\exp\left\{-\frac{1}{2}\min\left(\frac{t_1^2}{32}, \frac{t_1}{4}\right)\right\} + \exp\left\{-\frac{1}{2}\min\left(\frac{t_2^2}{32}, \frac{t_2}{4}\right)\right\} \leq 2\exp\left\{-\frac{1}{2}\min\left(\frac{t^2}{32E^2}, \frac{t}{32E}\right)\right\}. \tag{B.4}$$

Note that $\frac{t_1^2}{32} \le \frac{t_1}{4}$ for $0 \le t \le t_1^* := \frac{1}{\mu+c-8} - E$, and $\frac{t_2^2}{32} \le \frac{t_2}{4}$ for $0 \le t \le t_2^* := E - \frac{1}{\mu+c+8}$. Then:

1. On $t \in [0, t_1^* \wedge t_2^*]$:

$$\begin{aligned}
\text{LHS of (B.4)} &= \exp\left\{-\frac{1}{2}\frac{t_1^2}{32}\right\} + \exp\left\{-\frac{1}{2}\frac{t_2^2}{32}\right\} \\
&\le 2\exp\left\{-\frac{1}{4}\left(\frac{t_1}{8}+\frac{t_2}{8}\right)^2\right\} \qquad \text{(By Jensen's inequality)} \\
&\le 2\exp\left\{-\frac{1}{256}\left(\frac{2t}{E^2-t^2}\right)^2\right\} \\
&\le 2\exp\left\{-\frac{t^2}{64E^2}\right\} \qquad \text{(For } m \vee q > c_1 \text{ for suitably large } c_1 > 0, \text{ we have } E^2 < 1\text{)} \\
&\le \text{RHS of (B.4)}
\end{aligned}$$

2. On $t \ge t_1^* \vee t_2^*$:

$$\begin{aligned}
\text{LHS of (B.4)} &= \exp\left\{-\frac{1}{2}\frac{t_1}{4}\right\} + \exp\left\{-\frac{1}{2}\frac{t_2}{4}\right\} \\
&\le 2\exp\left\{-\frac{1}{2}\left(\frac{t_1}{8}+\frac{t_2}{8}\right)\right\} \qquad \text{(By Jensen's inequality)} \\
&= 2\exp\left\{-\frac{1}{16}\left(\frac{2t}{E^2-t^2}\right)\right\} \\
&\le 2\exp\left\{-\frac{t^2}{64E^2}\right\} \qquad \text{(For } m \vee q > c_1 \text{ for suitably large } c_1 > 0, \text{ we have } E < 1\text{)} \\
&\le \text{RHS of (B.4).}
\end{aligned}$$

3. On $t \in [t_1^* \wedge t_2^*, t_1^* \vee t_2^*]$:

   3.1. When $t_1^* \le t_2^*$:

$$\begin{aligned}
\text{LHS of (B.4)} &= \exp\left\{-\frac{1}{2}\frac{t_1}{4}\right\} + \exp\left\{-\frac{1}{2}\frac{t_2^2}{32}\right\} \\
&= \exp\left\{-\frac{1}{8}\left(\mu+c-\frac{1}{E+t}\right)\right\} + \exp\left\{-\frac{1}{64}\left(\frac{1}{E-t}-\mu-c\right)^2\right\},
\end{aligned}$$

where the last expression takes maximum value at $t = t_1^*$. Thus, denoting $\Delta_1 = \frac{1}{E - t_1^*} - \mu - c$, we have $\Delta_1 \geq 0$, and

$$\begin{aligned}
\text{LHS of (B.4)} &\leq \exp\{-1\} + \exp\left\{-\frac{\Delta_1^2}{64}\right\} \\
&\leq 2\exp\left\{-\frac{1}{4}\left(1 + \frac{\Delta_1}{8}\right)^2\right\} \\
&\leq 2\exp\left\{-\frac{1}{64}\left(1 - \frac{1}{E(\mu + c + 8)}\right)^2\right\} \qquad \text{(since } E \geq \frac{1}{\mu + c} \text{ by Lemma B.1.2} \\
&= 2\exp\left\{-\frac{(t_2^*)^2}{64E^2}\right\} \\
&\leq 2\exp\left\{-\frac{t^2}{64E^2}\right\} \\
&\leq \text{RHS of(B.4).}
\end{aligned}$$

3.2. When $t_1^* > t_2^*$: similarly, the LHS of (B.4) reaches maximum value at $t = t_2^*$. Then denoting $\Delta_2 = \mu + c - \frac{1}{E + t_2^*}$, we have

$$\begin{aligned}
\text{LHS of (B.4)} &= \exp\left\{-\frac{1}{2}\frac{t_2}{4}\right\} + \exp\left\{-\frac{1}{2}\frac{t_1^2}{32}\right\} \\
&\leq \exp\{-1\} + \exp\left\{-\frac{\Delta_2^2}{64}\right\} \\
&\leq 2\exp\left\{-\frac{1}{4}\left(1 + \frac{\Delta_2}{8}\right)^2\right\}; \\
\text{RHS of (B.4)} &\geq 2\exp\left\{-\frac{(t_1^*)^2}{64E^2}\right\};
\end{aligned}$$

and

$$\begin{aligned}
\frac{t_1^*}{8E} &= \frac{1}{8}\left(\frac{1}{E(\mu + c - 8)} - 1\right) \\
&\leq \frac{1}{8}\left(\frac{\mu + c}{\mu + c - 8} - 1\right) \qquad \text{(since } E \geq \frac{1}{\mu + c} \text{ by Lemma B.1.2.5)} \\
&\leq \frac{1}{2} + \frac{\Delta_2}{16},
\end{aligned}$$

where the last inequality hold for $m \vee q > c_2$ for suitably large $c_2 > 0$, based on Lemma B.1.2.3.

Thus we have shown (B.4) holds for $t \leq E$.

- Case 2: $E < t < \frac{1}{c}$: Denoting $t_1 = \frac{1}{E+t} - \mu - c$, using Lemma B.1.2.4, we have that:

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{\sigma^2+c} - E\right| \leq t\right) &= \mathbb{P}\left(\sigma^2 - \mu \geq \frac{1}{E+t} - c - \mu\right) \\ &\geq 1 - \exp\left\{-\frac{1}{2}\min\left(\frac{t_1^2}{32}, \frac{t_1}{4}\right)\right\}.\end{aligned}$$

Then we want to show

$$1 - \exp\left\{-\frac{1}{2}\min\left(\frac{t_1^2}{32}, \frac{t_1}{4}\right)\right\} \geq 1 - 2\exp\left\{-\frac{1}{2}\min\left(\frac{c_1^2 t^2}{128E^2}, \frac{ct}{8E}\right)\right\}. \tag{B.5}$$

To show (B.5), note that

$$\frac{t_1^2}{32} \geq \frac{c^2t^2}{128E^2} \Leftrightarrow t_1 \geq \frac{ct}{2E} \Leftrightarrow \mu + c \geq \frac{ct}{2E} + \frac{1}{E+t};$$

and we can show

$$\begin{aligned}\frac{ct}{2E} + \frac{1}{E+t} &\leq \max\left(\frac{c}{2} + \frac{1}{2E}, \frac{1}{2E} + \frac{1}{E + \frac{1}{c}}\right) \\ &\leq \max\left(\frac{c}{2} + \frac{\mu+c}{2}, \frac{\mu+c}{2} + c\right) \\ &\leq \mu + c\end{aligned}$$

using that fact that $E \geq \frac{1}{\mu+c}$ (Lemma B.1.2.5) for $m \vee q > c_2$ for suitably large $c_2 > 0$. Thus, (B.5) holds.

- Case 3: $t \geq \frac{1}{c}$: We have $\mathbb{P}\left(\left|\frac{1}{\sigma^2+c} - E\right| \leq t\right) = 1$ since $\frac{1}{\sigma^2+c} \leq \frac{1}{c}$.

Thus, we have shown $\mathbb{P}\left(\left|\frac{1}{\sigma^2+c} - E\right| \leq t\right) \geq 1 - 2\exp\left\{-c_2\min\left(\frac{t^2}{E^2}, \frac{t}{E}\right)\right\}$, which implies $\frac{1}{\sigma^2+c} - E \in \mathrm{SE}(c_2E^2, c_2E)$.

Lemma B.1.2.7)

By Lemma B.1.2.6, it is straightforward that

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}\frac{1}{\sigma^2(Z^i)+c}-\sum_{i=1}^{n}\mathbb{E}\left(\frac{1}{\sigma^2(Z^i)+c}\right)\right|\le t\right)$$
$$\ge 1-2\exp\left\{-c_1\min\left(\frac{t^2}{\sum_{i=1}^{n}\mathbb{E}\left(\frac{1}{\sigma^2(Z^i)+c}\right)^2},\frac{t}{\max_i\mathbb{E}\left(\frac{1}{\sigma^2(Z^i)+c}\right)}\right)\right\}.$$

Based on Lemma B.1.2.5, we have $\sum_{i=1}^{n}\frac{1}{\sigma^2(Z^i)+c}\asymp\frac{n}{m\vee q}$ with probability at least $1-2\exp\{-c_2n\}$.

By Lemma B.1.2.6, we have

$$\mathbb{P}\left(\forall\ i,\left|\frac{1}{\sigma^2(Z^i)+c}-\mathbb{E}\left(\frac{1}{\sigma^2(Z^i)+c}\right)\right|\ge t\right)$$
$$\le 2\exp\left\{\log(n)-c_1\min\left(\frac{t^2}{\max_i\mathbb{E}\left(\frac{1}{\sigma^2(Z^i)+c}\right)^2},\frac{t}{\max_i\mathbb{E}\left(\frac{1}{\sigma^2(Z^i)+c}\right)}\right)\right\}.$$

Thus based on Lemma B.1.2.5, and the trivial bound $\max_i\frac{1}{\sigma^2(Z^i)+c}\le\frac{1}{c}$, we have $\max_i\frac{1}{\sigma^2(Z^i)+c}\le\frac{1}{c}\wedge\frac{c_2\log(n)}{m\vee q}$ with probability at least $1-2\exp\{-c_3n\}$.

By Lemma B.1.2.4, we have

$$\mathbb{P}\left(\forall\ i,\left|\sigma^2(Z^i)-\mathbb{E}(\sigma^2(Z^i))\right|\ge t\right)\le 2\exp\left\{\log(n)-c_1\min\left(\frac{t^2}{32},\frac{t}{4}\right)\right\}.$$

Thus $\max_i\sigma^2(Z^i)\le c_2(\log(n)+m\vee q)$ and $\min_i\frac{1}{\sigma^2(Z^i)+c}\ge\frac{c_3}{\log(n)+m\vee q}$ with probability at least $1-2\exp\{-c_4n\}$.

Lemma B.1.2.8)

Note that $\mathrm{tr}\left((\Sigma_a^i)^{-1}\right)=\sum_{l=1}^{m}\frac{1}{a\sigma_l^2(Z^i)+1}$, where $\sigma_l^2(Z^i)$'s are the singular values of Z^i.

For $q > c_0 m$: $\sigma_l^2(Z^i)$'s are positive. We then have

$$\frac{m}{a\sigma_{\max}^2(Z^i)+1} \leq \operatorname{tr}\left((\Sigma_a^i)^{-1}\right) \leq \frac{m}{a\sigma_{\min}^2(Z^i)+1};$$

and based on Lemma B.1.2.7:

$$\begin{aligned}
&\mathbb{P}\left(\max_i \operatorname{tr}\left((\Sigma_a^i)^{-1}\right) \leq c_1 m\left(1 \wedge \frac{\log(n)}{q}\right)\right) \geq 1 - 2\exp\{-c_2\log(n)\},\\
&\mathbb{P}\left(\min_i \operatorname{tr}\left((\Sigma_a^i)^{-1}\right) \geq \frac{c_1 m}{\log(n)+q}\right) \geq 1 - 2\exp\{-c_2\log(n)\},\\
&\mathbb{P}\left(\sum_{i=1}^n \operatorname{tr}\left((\Sigma_a^i)^{-1}\right) \asymp \frac{nm}{q}\right) \geq 1 - 4\exp\{-c_3 n\}.
\end{aligned}$$

For $m > c_0 q$: $\sigma_l^2(Z^i)$'s are zero for $l > q$, and $\sigma_l^2(Z^i)$'s are positive for $1 \leq l \leq q$. Then $\operatorname{tr}\left((\Sigma_a^i)^{-1}\right) = m - q + \sum_{l=1}^q \frac{1}{a\sigma_l^2(Z^i)+1}$. We thus have:

$$\begin{aligned}
&\mathbb{P}\left(\max_i \operatorname{tr}\left((\Sigma_a^i)^{-1}\right) \leq m - q + c_1 q\left(1 \wedge \frac{\log(n)}{m}\right) \leq c_1\left(m + q\left(1 \wedge \frac{\log(n)}{m}\right)\right)\right) \geq 1 - 2\exp\{-c_2\log(n)\},\\
&\mathbb{P}\left(\min_i \operatorname{tr}\left((\Sigma_a^i)^{-1}\right) \geq m - q + \frac{c_1 q}{\log(n)+m} \geq c_1\left(m + \frac{q}{\log(n)+m}\right)\right) \geq 1 - 2\exp\{-c_2\log(n)\},\\
&\mathbb{P}\left(\sum_{i=1}^n \operatorname{tr}\left((\Sigma_a^i)^{-1}\right) \asymp n(m-q) + \frac{nq}{m} \asymp nm\right) \geq 1 - 4\exp\{-c_3 n\}.
\end{aligned}$$

Lemma B.1.2.9)

For any $q \times q$ orthogonal matrix P, Z and ZP follow the same distribution. Thus

$$\begin{aligned}
\mathbb{E}\left(Z^\top(aZZ^\top + I_m)^{-1}Z\right) &= \mathbb{E}\left((ZP)^\top(aZPP^\top Z^\top + I_m)^{-1}ZP\right)\\
&= P^\top\,\mathbb{E}\left(Z^\top(aZZ^\top + I_m)^{-1}Z\right)P.
\end{aligned}$$

Then by Proposition 2.14 in [56], we have $\mathbb{E}\left(Z^\top(aZZ^\top + I_m)^{-1}Z\right) = kI_q$ for some $k \in \mathbb{R}$.

We then have

$$\begin{aligned} kq = \mathrm{tr}(kI_q) &= \mathrm{tr}\left(\mathbb{E}\left(Z^\top (aZZ^\top + I_m)^{-1} Z\right)\right) \\ &= \mathbb{E}\left(\mathrm{tr}\left(I_m - (aZZ^\top + I_m)^{-1}\right)\right) \\ &= m - \mathbb{E}\left(\sum_{l=1}^{m} \frac{1}{a\sigma_l^2(Z)+1}\right). \end{aligned}$$

When $q > c_0 m$, $\sigma_l^2(Z)$'s are positive. By Lemma B.1.2.3 and Lemma B.1.2.5 we have

$$m - \frac{c_1 m}{q} \le m - m\,\mathbb{E}\left(\frac{1}{a\sigma_{\min}^2(Z)+1}\right) \le kq \le m - m\,\mathbb{E}\left(\frac{1}{a\sigma_{\max}^2(Z)+1}\right) \le m - \frac{c_2 m}{q}. \tag{B.6}$$

Thus $k \asymp \frac{m}{q}$ when $q > c_0 m$.

When $m > c_0 q$, $\sigma_l^2(Z)$'s are positive and $\sigma_l^2(Z) = 0$ for $l > q$. Then similar to the proof of (B.6), by Lemma B.1.2.3 and Lemma B.1.2.5, we have

$$q - \frac{c_1 q}{m} \le q - q\,\mathbb{E}\left(\frac{1}{a\sigma_{\min}^2(Z)+1}\right) \le kq = q - \mathbb{E}\left(\sum_{l=1}^{q} \frac{1}{a\sigma_l^2(Z)+1}\right) \le q - q\,\mathbb{E}\left(\frac{1}{a\sigma_{\max}^2(Z)+1}\right) \le q - \frac{c_2 q}{m}$$

Thus $k \asymp 1$ when $m > c_0 q$.

Then since $\|(Z^i)^\top (\Sigma_a^i)^{-1} Z^i\|_2 \le \max_l \frac{\sigma_l^2(Z^i)}{a\sigma_l^2(Z^i)+1} \le \frac{1}{a}$, by Matrix Chernoff bound [219] we have:

$$\begin{aligned} &\mathbb{P}\left(\sigma_{\min}\left(\sum_{i=1}^{n} (Z^i)^\top (\Sigma_a^i)^{-1} Z^i\right) \ge \frac{1}{2} n\sigma_{\min}\left(\mathbb{E}\left((Z^i)^\top (\Sigma_a^i)^{-1} Z^i\right)\right)\right) \\ &\quad \ge 1 - \exp\left\{\log(q) - \left(\frac{1}{2} - \log\sqrt{2}\right) an\sigma_{\min}\left(\mathbb{E}\left((Z^i)^\top (\Sigma_a^i)^{-1} Z^i\right)\right)\right\}; \\ &\mathbb{P}\left(\sigma_{\max}\left(\sum_{i=1}^{n} (Z^i)^\top (\Sigma_a^i)^{-1} Z^i\right) \le \frac{3}{2} n\sigma_{\max}\left(\mathbb{E}\left((Z^i)^\top (\Sigma_a^i)^{-1} Z^i\right)\right)\right) \\ &\quad \ge 1 - \exp\left\{\log(q) - \left(\frac{3}{2}\log\frac{3}{2} - \frac{1}{2}\right) an\sigma_{\min}\left(\mathbb{E}\left((Z^i)^\top (\Sigma_a^i)^{-1} Z^i\right)\right)\right\}. \end{aligned}$$

Plugging in $\mathbb{E}\left(Z^\top(aZZ^\top + I_m)^{-1}Z\right) = kI_q$ with $k \asymp \left(\frac{m}{q}\right)^{\mathbf{1}\{q>c_0m\}}$, we obtain the stated results in the lemma.

Lemma B.1.2.10)

By Lemma B.1.2.7 and Lemma B.1.2.9,

$$\begin{aligned}\sigma_{\max}\left(\sum_{i=1}^n (Z^i)^\top(\Sigma_a^i)^{-2}Z^i\right) &\le \sigma_{\max}\left(\sum_{i=1}^n (Z^i)^\top(\Sigma_a^i)^{-1}Z^i\right)\sigma_{\max}\left((\Sigma_a^i)^{-1}\right)\\ &\le c_1 n\left(\frac{m}{q}\right)^{\mathbf{1}\{q>c_0m\}}\left(1\wedge\frac{\log(n)}{m\vee q}\right)\end{aligned}$$

with probability at least $1-\exp\left\{-c_2 n\left(\frac{m}{q}\right)^{\mathbf{1}\{q>c_0m\}}\right\} - 2\exp\{-c_2\log(n)\}$. At the same time, based on Lemma B.1.2.7 we have

$$\begin{aligned}\sigma_{\max}\left(\sum_{i=1}^n (Z^i)^\top(\Sigma_a^i)^{-2}Z^i\right) &\le \sigma_{\max}\left(\sum_{i=1}^n (\Sigma_a^i)^{-1}\right)\max_i \sigma_{\max}\left((Z^i)^\top(\Sigma_a^i)^{-1}Z^i\right)\\ &\le \frac{c_1 n}{m\vee q}\end{aligned}$$

with probability at least $1-2\exp\{-c_2 n\}$. Thus with probability at least $1-2\exp\{-c_2n\} - \exp\left\{-c_2 n\left(\frac{m}{q}\right)^{\mathbf{1}\{q>c_0m\}}\right\} - 2\exp\{-c_2\log(n)\}$ we have

$$\sigma_{\max}\left(\sum_{i=1}^n (Z^i)^\top(\Sigma_a^i)^{-2}Z^i\right) \le \begin{cases} c_1\frac{n}{q}\left(1\wedge\frac{m\log(n)}{q}\right) & , q > c_0 m \\ c_1\frac{n}{m} & , m > c_0 q\end{cases}$$

□

Proof of Lemma B.1.3.

Lemma B.1.3.1)

Define $\check{X}^i = X^i(\Sigma_X^i)^{-1/2}$ and $\check{Z}^i = Z^i(\Sigma_Z^i)^{-1/2}$. Conditional on $\{\Sigma_X^i\}_{i=1}^n$ and $\{\Sigma_Z^i\}_{i=1}^n$, the

entries of $\check{X}^i$'s and $\check{Z}^i$'s independently follow $N(0,1)$ distribution. Then we have

$$\begin{aligned} X^\top \Sigma_a X &= \sum_{i=1}^n (\Sigma_X^i)^{1/2} (\check{X}^i)^\top (a\check{Z}^i \Sigma_Z^i (\check{Z}^i)^\top + I_m)^{-1} \check{X}^i (\Sigma_X^i)^{1/2} \\ &\preceq \max_i \sigma_{\max}(\Sigma_X^i) \sum_{i=1}^n (\check{X}^i)^\top (a\sigma_{\min}(\Sigma_Z^i) \check{Z}^i (\check{Z}^i)^\top + I_m)^{-1} \check{X}^i, \end{aligned}$$

and similarly,

$$X^\top \Sigma_a X \succeq \min_i \sigma_{\min}(\Sigma_X^i) \sum_{i=1}^n (\check{X}^i)^\top (a\sigma_{\max}(\Sigma_Z^i) \check{Z}^i (\check{Z}^i)^\top + I_m)^{-1} \check{X}^i.$$

Under Assumption 8.2, we have $\sigma(\Sigma_X^i) \asymp \sigma(\Sigma_Z^i) \asymp 1$ based on [116], and thus $\sigma(X^\top \Sigma_a X)$ has the same rate as $\sigma\left(\sum_{i=1}^n (\check{X}^i)^\top (c\check{Z}^i (\check{Z}^i)^\top + I_m)^{-1} \check{X}^i\right)$. Therefore, without loss of generality, we work with $\Sigma_X^i = I_p$, $\Sigma_Z^i = I_q$ for the rest of the proof.

When $p = q$: $X^i = Z^i$ for $i = 1, \ldots, n$. Then by Lemma B.1.2.9, we have $\sigma(X^\top \Sigma_a X) \asymp n\,(m/q)^{\mathbf{1}\{q > c_0 m\}}$ with probability at least $1 - \exp\left\{-cn\,(m/q)^{\mathbf{1}\{q > c_0 m\}}\right\}$.

When $p > q$: Recall that $Z^i = X_{1:q}^i$ and denote $Q^i = X_{-\{1:q\}}^i$. Suppose u is a non-random length q vector, and denote $u_Z = u_{1:q}$, $u_Q = u_{-\{1:q\}}$. We have $Q^i u_Q | Z^i \sim N(0, \|u_Q\|_2^2 I_m)$. Then we have

$$\begin{aligned} u^\top X^\top \Sigma_a^{-1} X u &= \sum_{i=1}^n \left(Z^i u_Z + Q^i u_Q\right)^\top \left(aZ^i (Z^i)^\top + I_m\right)^{-1} \left(Z^i u_Z + Q^i u_Q\right) \\ &= \sum_{i=1}^n u_Z^\top (Z^i)^\top (\Sigma_a^i)^{-1} Z^i u_Z \qquad \text{(B.7)} \\ &\quad + \sum_{i=1}^n u_Q^\top (Q^i)^\top (\Sigma_a^i)^{-1} Q^i u_Q \qquad \text{(B.8)} \\ &\quad + 2\sum_{i=1}^n u_Z^\top (Z^i)^\top (\Sigma_a^i)^{-1} Q^i u_Q. \qquad \text{(B.9)} \end{aligned}$$

1. For (B.7): By Lemma B.1.2.9, (B.7) $\asymp \|u_Z\|_2^2 n\,(m/q)^{\mathbf{1}\{q > c_0 m\}}$ with probability at least $1 - \exp\left\{-c_1 n\,(m/q)^{\mathbf{1}\{q > c_0 m\}}\right\}$.

2. For (B.9): Note that entries of Q^i are i.i.d. realizations of a standard normal distri-

bution. Thus given Z^i, $u_Z^\top (Z^i)^\top (\Sigma_a^i)^{-1} Q^i u_Q$ is a Gaussian random variable. Then by the tail bound of Gaussian random variables and Lemma B.1.2.10, we have

$$\mathbb{P}\left(\left|2\sum_{i=1}^n u_Z^\top (Z^i)^\top (\Sigma_a^i)^{-1} Q^i u_Q\right| > 2t \mid \{Z^i\}_{i=1}^n\right) \le 2\exp\left\{-\frac{t^2}{\sum_{i=1}^n u_Z^\top (Z^i)^\top (\Sigma_z^i)^{-2} Z^i u_Z \|u_Q\|_2^2}\right\}$$
$$\le \begin{cases} 2\exp\left\{-\frac{t^2}{\frac{c_1 n}{q}\left(1\wedge\frac{m\log(n)}{q}\right)\|u_Z\|_2^2\|u_Q\|_2^2}\right\} & , q > c_0 m \\ 2\exp\left\{-\frac{t^2}{\frac{c_1 n}{m}\|u_Z\|_2^2\|u_Q\|_2^2}\right\} & , m > c_0 q. \end{cases}$$

Thus with probability at least $1 - 2\exp\{-c_2 n\} - 4\exp\{-c_2\log(n)\} - \exp\left\{-c_2 n\,(m/q)^{\mathbf{1}\{q>c_0 m\}}\right\}$, we have:

$$\text{(B.9)} \le \begin{cases} c_3\sqrt{\frac{n\log(n)\left(1\wedge\frac{m\log(n)}{q}\right)}{q}}\|u_Z\|_2\|u_Q\|_2 & , q > c_0 m \\ c_3\sqrt{\frac{n\log(n)}{m}}\|u_Z\|_2\|u_Q\|_2 & , m > c_0 q. \end{cases}$$

3. For (B.8): With $\mathbb{E}\left(\sum_{i=1}^n u_Q^\top (Z^i)^\top (\Sigma_a^i)^{-1} Q^i u_Q | \{Z^i\}_{i=1}^n\right) = \operatorname{tr}(\Sigma_a^{-1})\|u_Q\|_2^2$, Hanson-Wright inequality gives:

$$\mathbb{P}\left(\left|\sum_{i=1}^n u_Q^\top (Z^i)^\top (\Sigma_a^i)^{-1} Q^i u_Q - \operatorname{tr}(\Sigma_a^{-1})\|u_Q\|_2^2\right| > t \mid \{Z^i\}_{i=1}^n\right)$$
$$\le 2\exp\left\{-c_1\min\left(\frac{t^2}{\|u_Q\|_2^4\|\Sigma_a^{-1}\|_F^2}, \frac{t}{\|u_Q\|_2^2\|\Sigma_a^{-1}\|_2}\right)\right\},$$

where based on Lemma B.1.2.8:

$$\operatorname{tr}\left(\Sigma_a^{-1}\right) \asymp mnq^{-\mathbf{1}\{q>c_0 m\}}$$
$$\|\Sigma_a^{-1}\|_F^2 = \operatorname{tr}(\Sigma_a^{-2}) \le a^{-1}\operatorname{tr}(\Sigma_a^{-1}) \asymp mnq^{-\mathbf{1}\{q>c_0 m\}}$$
$$\|\Sigma_a^{-1}\|_2 \le a^{-1}.$$

Then (B.8)$\asymp \|u_Q\|_2^2 mnq^{-\mathbf{1}\{q>c_0 m\}}$ with probability at least $1 - 2\exp\{-c_1 mnq^{-\mathbf{1}\{q>c_0 m\}}\} - 4\exp\{-c_1 n\}$.

Thus with probability at least $1-4\exp\{-cn\}-2\exp\{-c\log(n)\}-2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}-\exp\left\{-cn\,(m/q)^{\mathbf{1}\{q>c_0m\}}\right\}$:

For $m > c_0 q$: Since $\frac{\log(n)}{nm} = o(1)$, we have

$$\begin{aligned}
&c_1 n\|u\|_2^2 \le c_2\left(n\|u_Z\|_2^2 + mn\|u_Q\|_2^2 - \sqrt{\frac{n\log(n)}{m}}\|u_Z\|_2\|u_Q\|_2\right)\\
&\le u^\top X^\top \Sigma_a^{-1} X u\\
&\le c_2\left(n\|u_Z\|_2^2 + mn\|u_Q\|_2^2 + \sqrt{\frac{n\log(n)}{m}}\|u_Z\|_2\|u_Q\|_2\right) \le c_3 mn\|u\|_2^2,
\end{aligned}$$

which implies $c_1 n \le \sigma(X^\top \Sigma_a^{-1} X) \le c_2 mn$.

For $q > c_0 m$: Since $\frac{\log^2(n)}{mn} = o(1)$, we have

$$\begin{aligned}
&c_1\frac{nm}{q}\|u\|_2^2 \le c_2\left(\frac{nm}{q}\|u_Z\|_2^2 + \frac{nm}{q}\|u_Q\|_2^2 - \sqrt{\frac{n\log(n)\left(1\wedge\frac{m\log(n)}{q}\right)}{q}}\|u_Z\|_2\|u_Q\|_2\right)\\
&\le u^\top X^\top \Sigma_a^{-1} X u\\
&\le c_2\left(\frac{nm}{q}\|u_Z\|_2^2 + \frac{nm}{q}\|u_Q\|_2^2 + \sqrt{\frac{n\log(n)\left(1\wedge\frac{m\log(n)}{q}\right)}{q}}\|u_Z\|_2\|u_Q\|_2\right) \le c_3\frac{nm}{q}\|u\|_2^2,
\end{aligned}$$

which implies $\sigma(X^\top \Sigma_a^{-1} X) \asymp nm/q$.

Lemma B.1.3.2)

As discussed in the proof of Lemma B.1.3.1, without loss of generality, we can let $\Sigma_X^i = I_p$, $\Sigma_Z^i = I_q$.

When $p = q$: $X^i = Z^i$ for $i = 1, \ldots, n$. Then based on the singular value decomposition

$Z^i = U^i D^i V^i$ with $D^i = \operatorname{diag}\left(\{\sigma_l(Z^i)\}_{l=1}^m\right)$, we have:

$$\begin{aligned}
&\max_i \sigma\left((X^i)^\top(\Sigma_a^i)^{-1}X^i\right) = \max_i \sigma\left(\left(aZ^i(Z^i)^\top + 1\right)^{-1} Z^i(Z^i)^\top\right) \\
&= \max_i \sigma\left(U^i(a(D^i)^2 + I_m)^{-1}(U^i)^\top U^i(D^i)^2(U^i)^\top\right) \\
&\leq \max_{i,l} \frac{\sigma_l^2(Z^i)}{a\sigma_l^2(Z^i)+1} \leq \frac{1}{a}. \qquad \text{(B.10)}
\end{aligned}$$

When $p > q$: Following the proof of Lemma B.1.3.1 (the case when $p > q$), denoting $X^i = (Z^i, Q^i)$ and $u = (u_Z^\top, u_Q^\top)^\top$, we have

$$\begin{aligned}
u^\top(X^i)^\top(\Sigma_a^i)^{-1}X^i u &= u_Z^\top(Z^i)^\top(\Sigma_a^i)^{-1}Z^i u_Z \qquad \text{(B.11)} \\
&+ u_Q^\top(Z^i)^\top(\Sigma_a^i)^{-1}Q^i u_Q \qquad \text{(B.12)} \\
&+ 2u_Z^\top(Z^i)^\top(\Sigma_a^i)^{-1}Q^i u_Q. \qquad \text{(B.13)}
\end{aligned}$$

1. For (B.11): $u_Z^\top(Z^i)^\top(\Sigma_a^i)^{-1}Z^i u_Z \leq \|u_Z\|_2^2 \sigma_{\max}\left((Z^i)^\top(\Sigma_a^i)^{-1}Z^i\right) \leq c_1\|u_Z\|_2^2$, where we have used the inequality $\sigma_{\max}\left((Z^i)^\top(\Sigma_a^i)^{-1}Z^i\right) \leq c_1$ based on (B.10).

2. For (B.12): Hanson-Wright inequality gives

$$\begin{aligned}
&\mathbb{P}\left(\forall\, i, \left|u_Q^\top(Z^i)^\top(\Sigma_a^i)^{-1}Q^i u_Q - \operatorname{tr}\left((\Sigma_a^i)^{-1}\right)\|u_Q\|_2^2\right| \mid Z^i\right) \\
&\leq 2\exp\left\{\log(n) - c_1 \min\left(\frac{t^2}{\max_i \|u_Q\|_2^4 \|(\Sigma_a^i)^{-1}\|_F^2}, \frac{t}{\max_i \|u_Q\|_2^2 \|(\Sigma_a^i)^{-1}\|_2}\right)\right\}
\end{aligned}$$

   where by Lemma B.1.2.7 and Lemma B.1.2.8:

$$\|(\Sigma_a^i)^{-1}\|_2 \leq c_1\left(1 \wedge \frac{\log(n)}{m \vee q}\right),$$

$$\|(\Sigma_a^i)^{-1}\|_F^2 \leq \max_i \operatorname{tr}\left((\Sigma_a^i)^{-1}\right)\sigma_{\max}\left((\Sigma_a^i)^{-1}\right) \leq \begin{cases} c_1 m\left(1 \wedge \frac{\log(n)}{q}\right)^2 & , q > c_0 m, \\ c_1\left(m + q\left(1 \wedge \frac{\log(n)}{m}\right)\right)\left(1 \wedge \frac{\log(n)}{m}\right) & , m > c_0 q, \end{cases}$$

$$\operatorname{tr}\left((\Sigma_a^i)^{-1}\right) \leq \begin{cases} c_1 m\left(1 \wedge \frac{\log(n)}{q}\right) & , q > c_0 m, \\ c_1\left(m + q\left(1 \wedge \frac{\log(n)}{m}\right)\right) & , m > c_0 q. \end{cases}$$

Then with probability at least $1 - 6\exp\{-c\log(n)\}$,

$$u_Q^\top (Z^i)^\top (\Sigma_a^i)^{-1} Q^i u_Q \leq \begin{cases} c_1 \|u_Q\|_2^2 m \log(n) \left(1 \wedge \frac{\log(n)}{q}\right) & , q > c_0 m, \\ c_1 \|u_Q\|_2^2 \log(n) \left(m + q\left(1 \wedge \frac{\log(n)}{m}\right)\right) & , m > c_0 q. \end{cases}$$

3. For (B.13): Since entries of Q^i are i.i.d. realizations of a standard normal distribution, we have $u_Z^\top (Z^i)^\top (\Sigma_a^i)^{-1} Q^i u_Q$ following a normal distribution conditional on Z^i. Thus based on the tail bound of a Gaussian random variable, we have that:

$$\begin{aligned} &\mathbb{P}\left(\max_i \left| u_Z^\top (Z^i)^\top (\Sigma_a^i)^{-1} Q^i u_Q \right| > t \mid Z^i \right) \leq 2\exp\left\{ \log(n) - c_1 \frac{t^2}{\|u_Q\|_2^2 \max_i \|(\Sigma_a^i)^{-1}(Z^i)^\top u_Z\|_2^2} \right\} \\ &\leq 2\exp\left\{ \log(n) - c_1 \frac{t^2}{\|u_Q\|_2^2 \|u_Z\|_2^2 \max_i \sigma_{\max}\left((Z^i)^\top (\Sigma_a^i)^{-1} Z^i\right) \sigma_{\max}\left((\Sigma_a^i)^{-1}\right)} \right\} \\ &\leq 2\exp\left\{ \log(n) - c_1 \frac{t^2}{\|u_Q\|_2^2 \|u_Z\|_2^2 \max_i \left(1 \wedge \frac{\log(n)}{m \vee q}\right)} \right\} \qquad \text{(based on Lemma B.1.2.7).} \end{aligned}$$

Then $\max_i \left| u_Z^\top (Z^i)^\top (\Sigma_a^i)^{-1} Q^i u_Q \right| \leq c_1 \|u_Q\|_2 \|u_Z\|_2 \sqrt{\log(n)\left(1 \wedge \frac{\log(n)}{m \vee q}\right)}$ with probability at least $1 - 4\exp\{-c\log(n)\}$.

Based on the bounds of (B.11), (B.12) and (B.13), with probability at least $1 - 8\exp\{-c\log(n)\}$, we have that:

$$\begin{aligned} u^\top (X^i)^\top (\Sigma_a^i)^{-1} X^i u &\leq c_1 \|u_Z\|_2^2 + c_1 \|u_Q\|_2 \|u_Z\|_2 \sqrt{\log(n)\left(1 \wedge \frac{\log(n)}{m \vee q}\right)} \\ &\quad + \begin{cases} c_1 \|u_Q\|_2^2 m \log(n) \left(1 \wedge \frac{\log(n)}{q}\right) & , q > c_0 m, \\ c_1 \|u_Q\|_2^2 \log(n) \left(m + q\left(1 \wedge \frac{\log(n)}{m}\right)\right) & , m > c_0 q. \end{cases} \\ &\leq \begin{cases} c_1 \|u\|_2^2 \left(1 \vee \frac{m \log^2(n)}{q}\right) & , q > c_0 m, \\ c_1 \|u\|_2^2 m \log^2(n) & , m > c_0 q. \end{cases} \end{aligned}$$

□

Proof of Lemma B.1.4.

Lemma B.1.4.1)

We can show that:

$$\begin{aligned}
&\max_{1\le j\le p}\sum_{i=1}^{n}\left\|(Z^i)^\top\left(\Sigma_a^i\right)^{-1}X_j^i\right\|_2^2\|\Psi\|_2+\left\|\left(\Sigma_a^i\right)^{-1}X_j^i\right\|_2^2\|R^i\|_2\\
&=\max_{1\le j\le p}\sum_{i=1}^{n}e_j^\top(X^i)^\top\left(\Sigma_a^i\right)^{-1}Z^i(Z^i)^\top\left(\Sigma_a^i\right)^{-1}X^ie_j\,\|\Psi\|_2+e_j^\top(X^i)^\top\left(\Sigma_a^i\right)^{-1}\left(\Sigma_a^i\right)^{-1}X^ie_j\|R^i\|_2\\
&\le\max_j\max_i\|e_j\|_2^2\sigma_{\max}\left(X^\top\Sigma_a^{-1}X\right)\left(\sigma_{\max}\left(\Sigma_a^{-1}\right)\|R^i\|_2+\sigma_{\max}\left((Z^i)^\top\left(\Sigma_a^i\right)^{-1}Z^i\right)\|\Psi\|_2\right).
\end{aligned}$$

We can bound $\sigma_{\max}(X^\top\Sigma_a^{-1}X)$ based on Lemma B.1.3.1. The remaining terms can be bound under Assumption 8.2 and using the results from Lemma B.1.2.7 and Lemma B.1.3.2:

$$\begin{aligned}
&\max_i\ \sigma_{\max}(\Sigma_a^{-1})\|R^i\|_2+\sigma_{\max}\left((Z^i)^\top\left(\Sigma_a^i\right)^{-1}Z^i\right)\|\Psi\|_2\\
&\le c_1+c_2\left(1\wedge\frac{\log(n)}{m\vee q}\right)\\
&\le c_3,
\end{aligned}$$

with probability at least $1-10\exp\{-c_4\log(n)\}$. Thus we have

$$\max_{1\le j\le p}\sum_{i=1}^{n}\left\|(Z^i)^\top\left(\Sigma_a^i\right)^{-1}X_j^i\right\|_2^2\|\Psi\|_2+\left\|\left(\Sigma_a^i\right)^{-1}X_j^i\right\|_2^2\|R^i\|_2\le\begin{cases}\frac{c_5mn}{q} & ,q>c_0m,\\ c_5n & ,m>c_0q \text{ and } p=q,\\ c_5mn & ,m>c_0q \text{ and } q<p,\end{cases}$$

with probability at least $1-4\exp\{-cn\}-12\exp\{-c\log(n)\}-2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}-\exp\left\{-cn\,(m/q)^{\mathbf{1}\{q>c_0m\}}\right\}$.

When $q > c_0 m$ and we additionally assume Assumption 9.1: We have

$$\begin{aligned}
&\max_{1\le j\le p}\sum_{i=1}^{n}\left\|(Z^i_{S_\psi})^\top\left(\Sigma^i_a\right)^{-1}X^i_j\right\|_2^2\|\Psi\|_2+\left\|\left(\Sigma^i_a\right)^{-1}X^i_j\right\|_2^2\|R^i\|_2\\
&=\max_{1\le j\le p}\sum_{i=1}^{n}e_j^\top(X^i)^\top\left(\Sigma^i_a\right)^{-1}Z^i_{S_\psi}(Z^i_{S_\psi})^\top\left(\Sigma^i_a\right)^{-1}X^ie_j\,\|\Psi\|_2+e_j^\top(X^i)^\top\left(\Sigma^i_a\right)^{-1}\left(\Sigma^i_a\right)^{-1}X^ie_j\|R^i\|_2\\
&\le\max_j\max_i\sum_{i=1}^{n}\|e_j\|_2^2\sigma_{\max}\left(X^\top\Sigma_a^{-1}X\right)\sigma_{\max}\left(\Sigma_a^{-1}\right)\left(\sigma_{\max}\left(Z^i_{S_\psi}\left(Z^i_{S_\psi}\right)^\top\right)\|\Psi\|_2+\|R^i\|_2\right).
\end{aligned}$$

Here, based on Lemma B.1.2.4, $\max_i\sigma_{\max}\left(Z^i_{S_\psi}(Z^i_{S_\psi})^\top\right)\le c_1 m\log(n)$ with probability at least $1-2\exp\{-c\log(n)\}$ assuming $s_\psi\le c_2 m$ for some constant $c_2>0$ (Assumption 9.1). Then plugging in the upper bound for $\sigma_{\max}\left(X^\top\Sigma_a^{-1}X\right)$ from Lemma B.1.3.1, and the upper bound for $\sigma_{\max}\left(\Sigma_a^{-1}\right)$ from Lemma B.1.2.7, we obtain:

$$\max_{1\le j\le p}\sum_{i=1}^{n}\left\|(Z^i_{S_\psi})^\top\left(\Sigma^i_a\right)^{-1}X^i_j\right\|_2^2\|\Psi\|_2+\left\|\left(\Sigma^i_a\right)^{-1}X^i_j\right\|_2^2\|R^i\|_2\le\begin{cases}\frac{c_1m^2n\log^2(n)}{q^2} & ,q>c_0m,\\ c_1n\log^2(n) & ,m>c_0q\text{ and }p=q,\\ c_1mn\log^2(n) & ,m>c_0q\text{ and }q<p.\end{cases}$$

Since we also have

$$\begin{aligned}
&\max_{1\le j\le p}\sum_{i=1}^{n}\left\|(Z^i_{S_\psi})^\top\left(\Sigma^i_a\right)^{-1}X^i_j\right\|_2^2\|\Psi\|_2+\left\|\left(\Sigma^i_a\right)^{-1}X^i_j\right\|_2^2\|R^i\|_2\\
&\le\max_{1\le j\le p}\sum_{i=1}^{n}\left\|(Z^i)^\top\left(\Sigma^i_a\right)^{-1}X^i_j\right\|_2^2\|\Psi\|_2+\left\|\left(\Sigma^i_a\right)^{-1}X^i_j\right\|_2^2\|R^i\|_2,
\end{aligned}$$

we take the smaller upper bounds and obtain:

$$\max_{1\le j\le p}\sum_{i=1}^{n}\left\|(Z^i_{S_\psi})^\top\left(\Sigma^i_a\right)^{-1}X^i_j\right\|_2^2\|\Psi\|_2+\left\|\left(\Sigma^i_a\right)^{-1}X^i_j\right\|_2^2\|R^i\|_2\le\begin{cases}\frac{c_1mn}{q}\left(1\wedge\frac{m\log^2(n)}{q}\right) & ,q>c_0m,\\ c_1n & ,m>c_0q\text{ and }p=q\\ c_1mn & ,m>c_0q\text{ and }q<p\end{cases}$$

with probability at least $1-4\exp\{-cn\}-6\exp\{-c\log(n)\}-2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}-\exp\left\{-cn\,(m/q)^{\mathbf{1}\{q>c_0m\}}\right\}$.

Lemma B.1.4.2)

Conditional on X, the random effect γ_i's and the noise ϵ_i's are independent sub-Gaussian random vectors with parameters $c_1\|\Psi\|_2$ and $c_2\|R^i\|_2$. Thus for fixed vectors u_i, $i = 1, \ldots, n$, we have $\sum_{i=1}^n u_i^\top Z^i\gamma_i + u_i^\top\epsilon_i \mid X \in \mathrm{SG}\left(c_3\sum_{i=1}^n \|(Z^i)^\top u_i\|_2^2\|\Psi\|_2 + \|u_i\|_2^2\|R^i\|_2\right)$. Then conditional on X, letting $u_i = \left(\Sigma_a^i\right)^{-1} X_j^i$, we have $z_0^* = \max_j(\mathrm{tr}(\Sigma_a^{-1}))^{-1}\sum_{i=1}^n u_i^\top Z^i\gamma_i + u_i^\top\epsilon_i$. We thus have the following inequality based on the tail bound for sub-Gaussian random variables:

$$\mathbb{P}\left(z_0^* > t \mid X\right) \leq 2\exp\left\{\log(p) - c\frac{t^2\,\mathrm{tr}^2(\Sigma_a^{-1})}{\max_j\sum_{i=1}^n \|(Z^i)^\top(\Sigma_a^i)^{-1}X_j^i\|_2^2\|\Psi\|_2 + \|(X_j^i)^\top(\Sigma_a^i)^{-1}\|_2^2\|R^i\|_2}\right\}.$$

Let $t = c_1(\mathrm{tr}(\Sigma_a^{-1}))^{-1}\sqrt{\log(p)\max_j\sum_{i=1}^n \|(Z^i)^\top(\Sigma_a^i)^{-1}X_j^i\|_2^2\|\Psi\|_2 + \|(X_j^i)^\top(\Sigma_a^i)^{-1}\|_2^2\|R^i\|_2}$. Then using Lemma B.1.4.1 and the bound for $\mathrm{tr}(\Sigma_a^{-1})$ from Lemma B.1.2.8, we obtain the following bounds for z_0^* with probability at least $1 - 4\exp\{-cn\} - 12\exp\{-c\log(n)\} - 2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - \exp\left\{-cn\,(m/q)^{\mathbf{1}\{q>c_0m\}}\right\}$:

$$z_0^* \leq \begin{cases} c_1\sqrt{\frac{q\log(p)}{nm}} & , q > c_0m, \\ c_1\sqrt{\frac{\log(p)}{nm^2}} & , m > c_0q \text{ and } p = q, \\ c_1\sqrt{\frac{\log(p)}{nm}} & , m > c_0q \text{ and } p > q. \end{cases}$$

When we additionally assume Assumption 9.1, the structure of Ψ implies $\gamma_{i,S_\psi} = 0$. Thus $y - X\beta^* = \sum_{i=1}^n Z_{S_\psi}^i\gamma_{i,S_\psi} + \epsilon_i$, where $\gamma_{i,S_\psi} \in \mathrm{SGV}(c_1\|\Psi\|_2)$. Then following the same arguments as above, using the tail bound for sub-Gaussian random variables, and using the results in Lemma B.1.4.1, we have $t = O_p\left(\sqrt{\frac{q\log(p)}{mn}\left(1 \wedge \frac{m\log^2(n)}{q}\right)}\right)$, with probability at least

$$1 - 4\exp\{-cn\} - 6\exp\{-c\log(n)\} - 2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - \exp\left\{-cn\,(m/q)^{\mathbf{1}\{q>c_0m\}}\right\}.$$

Thus under Assumption 9.1, we obtain:

$$z_0^* \leq \begin{cases} c_1\sqrt{\frac{\log(p)\log^2(n)}{n}} & , q > c_0 m, \\ c_1\sqrt{\frac{\log(p)}{nm^2}} & , m > c_0 q \text{ and } p = q, \\ c_1\sqrt{\frac{\log(p)}{nm}} & , m > c_0 q \text{ and } p > q. \end{cases}$$

□

B.2 Inference Framework for β

Inference for β is based on the de-biased LASSO framework. We follow the same procedure as introduced in Section 3.2 of the main text, with slight modification in the context of a generic doubly high-dimensional LMM. To infer β_j, we first define the de-biased estimator $\hat{\beta}_j^{(db)}$:

$$\hat{\beta}_j^{(db)} = \hat{\beta}_j + \frac{\sum_{i=1}^n (\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2}(y^i - X^i\hat{\beta})}{\sum_{i=1}^n (\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2} X_j^i}. \tag{B.14}$$

Here, the modified proxy matrices $(\Sigma_b^i)^{-1/2}$ are defined as:

$$\Sigma_b^i = aZ_{-j}^i (Z_{-j}^i)^\top + I_m,$$
$$\Sigma_b = \text{diag}\left(\{\Sigma_b^i\}_{i=1}^n\right),$$

where with some abuse of notation, we define Z_{-j}^i, $j = 1, \ldots, q$ as the sub-matrix of Z^i obtained by dropping the jth column of Z^i, and $Z_{-j}^i = Z^i$ for $j > q$. The projection orthogonal vector w_j^i is defined as

$$\hat{w}_j^i = (\Sigma_b^i)^{-1/2}\left(X_j^i - X_{-j}^i\hat{\kappa}_j\right)$$

and the projection vector $\hat{\kappa}_j$ is defined as

$$\hat{\kappa}_j = \underset{\kappa_j \in \mathbb{R}^{p-1}}{\operatorname{argmin}} \left(2 \operatorname{tr}\left(\Sigma_b^{-1}\right)\right)^{-1} \|\Sigma_b^{-1/2}\left(X_j - X_{-j}\kappa_j\right)\|_2^2 + \lambda_j \|\kappa_j\|_1.$$

We state here the generalized version of Theorem 3.3.2 of the main text and its related assumptions for the generic doubly high-dimensional LMM in (B.1):

Assumption 10. *1.* $\log(p) = o(mn)$. $\forall\ j = 1, \ldots, p$, *conditional on* X^i_{-j}, *the random vector* X^i_j *has mean* $X^i_{-j}\kappa^*_j$ *for* $\kappa^*_j \in \mathbb{R}^{p-1}$ *and variance* G_j, *and* $X^i_j - X^i_{-j}\kappa^*_j \mid X^i_{-j} \in \mathrm{SGV}(c_1\|G_j\|_2)$. *The support for* κ^*_j *is* H_j *and its cardinality is* $|H_j|$. *Assume* $\|\kappa^*_j\|_1 \leq c_1|H_j|$.

2. $\frac{\|G_j\|_2}{\sigma_{\min}(G_j)} \log(n) \sqrt{1 \wedge \frac{\log(n)}{m \vee q}} \leq c_1 \sqrt{mnq^{-\mathbf{1}\{q > c_0 m\}}}$

3. 3.1. When $q > c_0 m$, $p = q$*:*

$$\frac{|H_j| q^2 \log(p)}{m^3 n} \ll \|G_j\|_2 \leq c_1 \frac{mn}{\log(n)\log(p)} \tag{B.15}$$

$$\frac{\|G_j\|_2}{\sigma^2_{\min}(G_j)} \ll \frac{mn}{|H_j|^3 \log(n)\log(p)} \wedge \frac{m^2 n^2}{s^2 |H_j| q \log(n) \log^2(p)} \tag{B.16}$$

$$\frac{\|G_j\|_2}{\sigma_{\min}(G_j)} \ll \frac{mn}{|H_j| \log(n)\log(p)} \tag{B.17}$$

3.2. when $q > c_0 m$, $p > q$*: assume the conditions* (B.15)–(B.17)*, and additionally assume* $\|G_j\|_2 \gg \frac{|H_j| \log(p) \log^3(n)}{mn}$

3.3. When $m > c_0 q$*:*

$$\frac{|H_j| \log(p)}{m^3 n} \left(m^3 \log^5(n)\right)^{\mathbf{1}\{p>q\}} \ll \|G_j\|_2 \leq c_1 \frac{mn}{\log(p)} \left(\frac{1}{m \log(n)}\right)^{\mathbf{1}\{p>q\}}$$

$$\frac{\|G_j\|_2}{\sigma^2_{\min}(G_j)} \ll \left(\frac{m^3 n}{|H_j|^3 \log(p)} \left(\frac{1}{m \log^2(n)}\right)^{\mathbf{1}\{p>q\}}\right) \wedge \left(\frac{m^3 n^2}{s^2 |H_j| \log^2(p)} \left(\frac{1}{m^2 \log^2(n)}\right)^{\mathbf{1}\{p>q\}}\right)$$

$$\frac{\|G_j\|_2}{\sigma_{\min}(G_j)} \ll \frac{m^2 n}{|H_j| \log(p)} \left(\frac{1}{m \log(n)}\right)^{\mathbf{1}\{p>q\}}$$

Assumption 11. *1. Condition 1: When Assumption 9.2 holds:*

$$\begin{cases} \frac{\|G_j\|_2}{\sigma_{\min}(G_j)} \ll \frac{n}{\log^6(n)} & , \ q > c_0 m \\ \frac{\|G_j\|_2}{\sigma_{\min}(G_j)} \ll \frac{n}{\log^5(n)} & , \ m > c_0 q \end{cases}$$

2. Condition 2: When Assumption 9.1 holds and $j \in S_\psi$:

$$\begin{cases} \frac{\|G_j\|_2}{\sigma^2_{\min}(G_j)} \ll \frac{n}{\log^6(n)} & , \ q > c_0 m \\ \frac{\|G_j\|_2}{\sigma^2_{\min}(G_j)} \ll \frac{mn}{\log^5(n)} & , \ m > c_0 q \end{cases}$$

3. Condition 3: When Assumption 9.1 holds and $j \notin S_\psi$:

$$\begin{cases} \frac{\|G_j\|_2}{\sigma_{\min}(G_j)} \ll \frac{n}{s_\psi \log^7(n)} \wedge \frac{n^2}{s_\psi |H_j|^2 \log(p) \log^8(n)} \wedge \frac{n}{\log^6(n)} & , \ q > c_0 m \\ \frac{\|G_j\|_2}{\sigma_{\min}(G_j)} \ll \frac{mn^2}{s_\psi |H_j|^2 \log(p) \log^7(n)} \wedge \frac{n}{\log^5(n)} & , \ m > c_0 q, \ p = q \\ \frac{\|G_j\|_2}{\sigma_{\min}(G_j)} \ll \frac{n^2}{s_\psi |H_j|^2 \log(p) \log^8(n)} \wedge \frac{n}{\log^5(n)} & , \ m > c_0 q, \ p > q \end{cases}$$

As discussed in the main text, the bound $\|\kappa_j^*\|_1 \leq c_1 |H_j|$ is satisfied when $\Sigma_X^i = \Sigma_X$. The variance G_j may take any form as long as its singular values satisfy Assumption 10 and Assumption 11. If we were to assume that G_j takes the same "sandwich" form as Σ_θ^i such that $G_j = Z_{-j} \Psi^j (Z_{-j})^\top + R^{i,j}$ for some matrix $\Psi^j \in \mathbb{R}^{(p-1)\times(p-1)}$ and $R^{i,j} \in \mathbb{R}^{m \times m}$, we could bound $\sigma(G_j)$ based on the rates of $\sigma\left(\Sigma_\theta^i\right)$: assuming that $\sigma\left(R^{i,j}\right) \asymp 1$, and Ψ^j takes the same form as Ψ (i.e., $\sigma\left(\Psi^j\right) \asymp 1$ as in Assumption 9.2, or Ψ^j is a sparse diagonal matrix as in Assumption 9.1), we would have $\sigma_{\min}(G_j) \asymp \sigma_{\min}(\Sigma_\theta^i)$ and $\|G_j\|_2 \asymp \|\Sigma_\theta^i\|_2$, and the following results bound $\sigma\left(\Sigma_\theta^i\right)$:

Lemma B.2.1. *For $\Sigma_\theta^i = Z^i \Psi (Z^i)^\top + R^i$, $i \in \{1, \ldots, n\}$, under Assumption 8, with probability at least $1 - c_1 \exp\{-c_2 m\}$, we have:*

1. When $m > c_0 q$: Under either Assumption 9.1 or Assumption 9.2,

$$c_3 \leq \sigma_{\min}(\Sigma_\theta^i) \leq \sigma_{\max}(\Sigma_\theta^i) \leq c_4 m.$$

2. When $q > c_0 m$: Under Assumption 9.2,

$$\sigma_{\min}(\Sigma_\theta^i) \asymp \sigma_{\max}(\Sigma_\theta^i) \asymp q.$$

Under Assumption 9.1,

$$c_5 \leq \sigma_{\min}(\Sigma_\theta) \leq \sigma_{\max}(\Sigma_\theta) \leq c_6 m.$$

Theorem B.2.2. *Under Assumption 8, Assumption 10 and Assumption 11, with probability at least* $1 - c_1 \exp\{-cn\} - c_2 \exp\{-c\log(n)\} - c_3 \exp\{-cmnq^{-\mathbf{1}\{q>c_0 m\}}\} - c_4 \exp\left\{-cn\,(m/q)^{\mathbf{1}\{q>c_0 m\}}\right\} - c_5 \exp\{-c\log(p)\} - c_6 \exp\{-cmn\}$*, we have that*

$$\frac{1}{\sqrt{V_j}}\left(\hat{\beta}_j^{(db)} - \beta_j^*\right) = R_j + o_p(1), \quad \textit{where } R_j \xrightarrow{d} N(0,1),$$

where the variance V_j is given by

$$V_j = \frac{\sum_{i=1}^n (\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2} \Sigma_{\theta^*}^i (\Sigma_b^i)^{-1/2} \hat{w}_j^i}{\left|\sum_{i=1}^n (\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2} X_j^i\right|^2}.$$

B.2.1 Related lemmas for Theorem B.2.2

Lemma B.2.3.

1. Define

$$z_j^* := \frac{1}{\mathrm{tr}(\Sigma_b^{-1})} \left\| X_{-j}^\top \Sigma_b^{-1} (X_j - X_{-j}\kappa_j^*) \right\|_\infty. \tag{B.18}$$

Under Assumption 8 and Assumption 10.1, with probability at least $1 - 2\exp\{-cmnq^{-\mathbf{1}\{q>c_0 m\}}\} - 4\exp\{-c\log(n)\} - 4\exp\{-cn\} - 2\exp\{-c\log(p)\} -$

$\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\}$, *we have that:*

$$z_j^* \leq \begin{cases} c_1\sqrt{\frac{q\log(p)\|G_j\|_2}{m^2n}\left(1\wedge\frac{m\log(n)}{q}\right)} & ,q>c_0m \text{ and } q=p, \\ c_1\sqrt{\frac{\log(p)\|G_j\|_2}{m^3n}} & ,m>c_0q \text{ and } q=p, \\ c_1\sqrt{\frac{q\log(p)\|G_j\|_2}{mn}\left(1\wedge\frac{\log(n)}{q}\right)} & ,q>c_0m \text{ and } q<p, \\ c_1\sqrt{\frac{\log(p)\|G_j\|_2}{mn}\left(1\wedge\frac{\log(n)}{m}\right)} & ,m>c_0q \text{ and } q<p. \end{cases}$$

2. *Under Assumption 8 and Assumption 10.1, we have the following results with probability at least* $1-c_1\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\}-c_2\exp\{-cn\}-c_3\exp\{-c\log(n)\}-c_4\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}-c_5\exp\{-c\log(p)\}$*:*

   2.1. *When* $q=p$ *and* $q>c_0m$*:* $\lambda_j=c_6\sqrt{\frac{q\log(p)\|G_j\|_2}{m^2n}\left(1\wedge\frac{m\log(n)}{q}\right)}$ *with suitably large* $c_6>0$*, and*

$$\|\hat{\kappa}_j-\kappa_j^*\|_2\leq c_7\sqrt{\frac{|H_j|q\log(p)\|G_j\|_2}{m^2n}\left(1\wedge\frac{m\log(n)}{q}\right)},$$
$$\|\hat{\kappa}_j-\kappa_j^*\|_1\leq c_7|H_j|\sqrt{\frac{q\log(p)\|G_j\|_2}{m^2n}\left(1\wedge\frac{m\log(n)}{q}\right)},$$
$$\|\Sigma_b^{-1/2}X_{-j}(\hat{\kappa}_j-\kappa_j^*)\|_2^2\leq c_7|H_j|\frac{\log(p)\|G_j\|_2}{m}\left(1\wedge\frac{m\log(n)}{q}\right).$$

   2.2. *When* $q=p$ *and* $m>c_0q$*:* $\lambda_j=c_6\sqrt{\frac{\log(p)\|G_j\|_2}{m^3n}}$ *with suitably large* $c_6>0$*, and*

$$\|\hat{\kappa}_j-\kappa_j^*\|_2\leq c_7\sqrt{\frac{|H_j|\log(p)\|G_j\|_2}{mn}},$$
$$\|\hat{\kappa}_j-\kappa_j^*\|_1\leq c_7|H_j|\sqrt{\frac{\log(p)\|G_j\|_2}{mn}},$$
$$\|\Sigma_b^{-1/2}X_{-j}(\hat{\kappa}_j-\kappa_j^*)\|_2^2\leq c_7|H_j|\frac{\log(p)\|G_j\|_2}{m}.$$

   2.3. *When* $q<p$ *and* $q>c_0m$*:* $\lambda_j=c_6\sqrt{\frac{q\log(p)\|G_j\|_2}{mn}\left(1\wedge\frac{\log(n)}{q}\right)}$ *with suitably large*

$c_6 > 0$, *and*

$$\|\hat{\kappa}_j - \kappa_j^*\|_2 \leq c_7 \sqrt{\frac{|H_j| q \log(p) \|G_j\|_2}{mn} \left(1 \wedge \frac{\log(n)}{q}\right)},$$
$$\|\hat{\kappa}_j - \kappa_j^*\|_1 \leq c_7 |H_j| \sqrt{\frac{q \log(p) \|G_j\|_2}{mn} \left(1 \wedge \frac{\log(n)}{q}\right)},$$
$$\|\Sigma_b^{-1/2} X_{-j} (\hat{\kappa}_j - \kappa_j^*)\|_2^2 \leq c_7 |H_j| \log(p) \|G_j\|_2 \left(1 \wedge \frac{\log(n)}{q}\right).$$

2.4. When $q < p$ *and* $m > c_0 q$*:* $\lambda_j = c_6 \sqrt{\frac{\log(p)\|G_j\|_2}{mn} \left(1 \wedge \frac{\log(n)}{m}\right)}$ *with suitably large* $c_6 > 0$, *and*

$$\|\hat{\kappa}_j - \kappa_j^*\|_2 \leq c_7 \sqrt{\frac{|H_j| m \log(p) \|G_j\|_2}{n} \left(1 \wedge \frac{\log(n)}{m}\right)},$$
$$\|\hat{\kappa}_j - \kappa_j^*\|_1 \leq c_7 |H_j| \sqrt{\frac{m \log(p) \|G_j\|_2}{n} \left(1 \wedge \frac{\log(n)}{m}\right)},$$
$$\|\Sigma_b^{-1/2} X_{-j} (\hat{\kappa}_j - \kappa_j^*)\|_2^2 \leq c_7 m \log(p) \|G_j\|_2 \left(1 \wedge \frac{\log(n)}{m}\right).$$

3. Under Assumption 8 and Assumption 10.1, with probability at least $1 - c_1 \exp\{-cmnq^{-\mathbf{1}\{q > c_0 m\}}\} - c_2 \exp\{-c \log(n)\} - c_3 \exp\{-cn\} - c_4 \exp\{-c \log(p)\} - c_5 \exp\{-cn(m/q)^{\mathbf{1}\{q > c_0 m\}}\} - c_6 \exp\{-cmn\}$*, we have that:*

$$\|X_{-j}^\top \Sigma_b^{-1} X_{-j} (\hat{\kappa}_j - \kappa_j^*)\|_\infty \leq \begin{cases} c_7 \sqrt{|H_j| \frac{n \log(p) \|G_j\|_2}{q} \left(1 \wedge \frac{m \log(n)}{q}\right)} & , q > c_0 m \text{ and } q = p, \\ c_7 \sqrt{|H_j| \frac{n \log(p) \|G_j\|_2}{m}} & , m > c_0 q \text{ and } q = p, \\ c_7 \sqrt{|H_j| \frac{mn \log(p) \|G_j\|_2}{q} \left(1 \wedge \frac{\log(n)}{q}\right)} & , q > c_0 m \text{ and } q < p, \\ c_7 \sqrt{|H_j| n \log(p) \|G_j\|_2 \left(m \wedge \log(n)\right)^2} & , m > c_0 q \text{ and } q < p, \end{cases}$$

4. Define

$$w_j = \Sigma_b^{-1/2}(X_j - X_{-j}\kappa_j^*)$$
$$w_j^i = (\Sigma_b^i)-\frac{1}{2}(X_j^i - X_{-j}^i\kappa_j^*).$$

Under Assumption 8, Assumption 10.1 and Assumption 10.2, we have $c_1\sigma_{\min}(G_j)nmq^{-\mathbf{1}\{q>c_0m\}} \leq \|w_j\|_2^2 \leq c_2\sigma_{\max}(G_j)nmq^{-\mathbf{1}\{q>c_0m\}}$ *with probability at least* $1-4\exp\{-cn\}-2\exp\{-c\log(n)\}$*, and*

$$\max_i \|w_j^i\|_2^2 \leq \begin{cases} c_1(m+\sqrt{m}\log(n))\left(1\wedge\frac{\log(n)}{q}\right)\|G_j\|_2 & ,q>c_0m, \\ c_1\left(m+\log(n)\sqrt{m\wedge\log(n)}\right)\|G_j\|_2 & ,m>c_0q \end{cases}$$

with probability at least $1-6\exp\{-c\log(n)\}$*.*

5. *Under Assumption 8, Assumption 10.1, Assumption 10.2, and Assumption 10.3, we have with probability at least* $1-c_1\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}-c_2\exp\{-c\log(n)\}-c_3\exp\{-cn\}-c_4\exp\{-c\log(p)\}-c_5\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\}-c_6\exp\{-cmn\}$ *that*

$$|\hat{w}_j^\top\Sigma_b^{-1/2}X_j| \geq c_7\sigma_{\min}(G_j)nmq^{-\mathbf{1}\{q>c_0m\}}.$$

6. *Under Assumption 8 and Assumption 10, we have*

$$c_1\sigma_{\min}(G_j)mnq^{-\mathbf{1}\{q>c_0m\}} \leq \|\hat{w}_j\|_2^2 \leq c_2\sigma_{\max}(G_j)nmq^{-\mathbf{1}\{q>c_0m\}},$$

and

$$\max_i \|\hat{w}_j^i\|_2^2 \leq \begin{cases} c_1\left(m+\sqrt{m}\log(n)\right)\left(1\wedge\frac{\log(n)}{q}\right)\|G_j\|_2 & ,q>c_0m, \\ c_1\left(m+\log(n)\sqrt{m\wedge\log(n)}\right)\|G_j\|_2 & ,m>c_0q, \end{cases}$$

with probability at least $1-c_1\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}-c_2\exp\{-c\log(n)\}-c_3\exp\{-cn\}-c_4\exp\{-c\log(p)\}-c_5\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\}-c_6\exp\{-cmn\}$*.*

7. *Under Assumption 8 and Assumption 10, with probability at least* $1 - c_1\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - c_2\exp\{-c\log(n)\} - c_3\exp\{-cn\} - c_4\exp\{-c\log(p)\} - c_5\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\} - c_6\exp\{-cmn\}$*, we have the following results hold:*

7.1. Under Condition 1 defined in Assumption 11, we have

$$\max_i \left\|(\Sigma^i_{\theta^*})^{1/2}(\Sigma^i_b)^{-1/2}\hat{w}^i_j\right\|_2^2 \leq \begin{cases} c_1\left(m+\sqrt{m}\log(n)\right)\left(1\wedge\frac{\log(n)}{q}\right)\log^2(n)\|G_j\|_2 & , q > c_0m, \\ c_1\left(m+\log(n)\sqrt{m\wedge\log(n)}\right)\log^2(n)\|G_j\|_2 & , m > c_0q, \end{cases}$$

$$\hat{w}_j^\top\Sigma_b^{-1/2}\Sigma_{\theta^*}\Sigma_b^{-1/2}\hat{w}_j \geq c_2nmq^{-\mathbf{1}\{q>c_0m\}}\sigma_{\min}(G_j).$$

7.2. Under Condition 2 defined in Assumption 11, we have

$$\max_i \left\|(\Sigma^i_{\theta^*})^{1/2}(\Sigma^i_b)^{-1/2}\hat{w}^i_j\right\|_2^2 \leq \begin{cases} c_1\left(m+\sqrt{m}\log(n)\right)\left(1\wedge\frac{\log(n)}{q}\right)\log^2(n)\frac{m}{q}\|G_j\|_2 & , q > c_0m, \\ c_1\left(m+\log(n)\sqrt{m\wedge\log(n)}\right)\log^2(n)\|G_j\|_2 & , m > c_0q, \end{cases}$$

$$\hat{w}_j^\top\Sigma_b^{-1/2}\Sigma_{\theta^*}\Sigma_b^{-1/2}\hat{w}_j \geq c_2nm^2q^{-2\times\mathbf{1}\{q>c_0m\}}\sigma^2_{\min}(G_j).$$

7.3. Under Condition 3 defined in Assumption 11, we have

$$\max_i \left\|(\Sigma^i_{\theta^*})^{1/2}(\Sigma^i_b)^{-1/2}\hat{w}^i_j\right\|_2^2 \leq c_1 \begin{cases} s_\psi\frac{m\log^4(n)}{q^2}\|G_j\|_2 + s_\psi|H_j|^2\frac{m\log(p)\log^5(n)}{q^2n}\|G_j\|_2 + \frac{m\log^3(n)}{q^2}\| \\ \qquad\qquad , q > c_0m, \\ s_\psi\frac{\log^2(n)}{m}\|G_j\|_2 + s_\psi|H_j|^2\frac{\log(p)\log^4(n)}{mn}\|G_j\|_2 + \log^2(n)\|G_j\|_2 \\ \qquad\qquad , m > c_0q,\ p = \\ s_\psi\frac{\log^2(n)}{m}\|G_j\|_2 + s_\psi|H_j|^2\frac{\log(p)\log^5(n)}{n}\|G_j\|_2 + \log^2(n)\|G_j\|_2 \\ \qquad\qquad , m > c_0q,\ p > \end{cases}$$

$$\hat{w}_j^\top\Sigma_b^{-1/2}\Sigma_{\theta^*}\Sigma_b^{-1/2}\hat{w}_j \geq \begin{cases} c_2\frac{nm\sigma_{\min}(G_j)}{q(q+\log(n))} & , q > c_0m, \\ c_2\frac{nm\sigma_{\min}(G_j)}{m+\log(n)} & , m > c_0q, \end{cases}$$

Remark B.2.1. If we only assume $\|\Psi\|_2 \leq c_1$ and do not restrict the structure of Ψ, we can still show

$$\max_i \left\| (\Sigma^i_{\theta^*})^{1/2} (\Sigma^i_b)^{-1/2} \hat{w}^i_j \right\|_2^2 \leq c_1 m \log^3(n) \left(\log(n)/q\right)^{\mathbf{1}\{q > c_0 m\}} \|G_j\|_2$$
$$\hat{w}_j^\top \Sigma_b^{-1/2} \Sigma_{\theta^*} \Sigma_b^{-1/2} \hat{w}_j \geq c_2 n m q^{-\mathbf{1}\{q > c_0 m\}} \sigma_{\min}(G_j)/(\log(n) + m \vee q)$$

under Assumption 8 and Assumption 10. Then Theorem B.2.2 still holds under the following additional assumption:

$$m \log^6(n) \left(q \log(n)/m\right)^{\mathbf{1}\{q > c_0 m\}} \|G_j\|_2 \ll n \sigma_{\min}(G_j).$$

B.2.2 Proof of Theorem B.2.2

Proof. Denote $\hat{u} = \hat{\beta} - \beta^*$ and $\hat{v} = \hat{\kappa}_j - \kappa_j^*$. Then we have:

$$\begin{aligned}
\hat{\beta}_j^{(db)} - \beta_j^* &= \hat{\beta}_j - \beta_j^* + \frac{\hat{w}_j^\top \Sigma_b^{-1/2}(y - X\hat{\beta})}{\hat{w}_j^\top \Sigma_b^{-1/2} X_j} \\
&= \left(e_j^\top - \frac{\hat{w}_j^\top \Sigma_b^{-1/2} X}{\hat{w}_j^\top \Sigma_b^{-1/2} X_j} \right) \hat{u} \qquad \text{(B.19)} \\
&\quad + \frac{\hat{w}_j^\top \Sigma_b^{-1/2}(y - X\beta^*)}{\hat{w}_j^\top \Sigma_b^{-1/2} X_j}. \qquad \text{(B.20)}
\end{aligned}$$

In the following, we will show (B.19) is $o_p(1)$, and (B.20) is asymptotically normal.

To show (B.19) is $o_p(1)$: First note that

$$\begin{aligned}
\left(e_j^\top - \frac{\hat{w}_j^\top \Sigma_b^{-1/2} X}{\hat{w}_j^\top \Sigma_b^{-1/2} X_j} \right) \hat{u} &\le \left\| e_j^\top - \frac{\hat{w}_j^\top \Sigma_b^{-1/2} X}{\hat{w}_j^\top \Sigma_b^{-1/2} X_j} \right\|_\infty \|\hat{u}\|_1 \\
&= \frac{\left\| e_j^\top \hat{w}_j^\top \Sigma_b^{-1/2} X_j - \hat{w}_j^\top \Sigma_b^{-1/2} X \right\|_\infty}{\left| \hat{w}_j^\top \Sigma_b^{-1/2} X_j \right|} \|\hat{u}\|_1 \\
&= \frac{\left\| \hat{w}_j^\top \Sigma_b^{-1/2} X_{-j} \right\|_\infty}{|\hat{w}_j^\top \Sigma_b^{-1/2} X_j|} \|\hat{u}\|_1,
\end{aligned}$$

where

$$\begin{aligned}
\left\| \hat{w}_j^\top \Sigma_b^{-1/2} X_{-j} \right\|_\infty &= \|X_{-j}^\top \Sigma_b^{-1}(X_j - X_{-j}\hat{\kappa}_j)\|_\infty \\
&\le \|X_{-j}^\top \Sigma_b^{-1}(X_j - X_{-j}\kappa_j^*)\|_\infty + \|X_{-j}^\top \Sigma_b^{-1} X_{-j}(\kappa_j^* - \hat{\kappa}_j)\|_\infty \\
&= z_j^* \operatorname{tr}(\Sigma_b^{-1}) + \|X_{-j}^\top \Sigma_b^{-1} X_{-j} \hat{v}\|_\infty
\end{aligned}$$

by the definition of z_j^* in (B.18). Then using Lemma B.1.2.8 to bound $\operatorname{tr}(\Sigma_b^{-1})$, Lemma

B.2.3.1 to bound z_j^* and Lemma B.2.3.3 to bound $\|X_{-j}^\top \Sigma_b^{-1} X_{-j}\hat{v}\|_\infty$, we have

$$\left\|\hat{w}_j^\top \Sigma_b^{-1/2} X_{-j}\right\|_\infty \leq \begin{cases} c_1\sqrt{|H_j|\frac{mn\log(p)\log(n)\|G_j\|_2}{q^2}} & , q > c_0 m, \\ c_1\sqrt{|H_j|\frac{n\log(p)\|G_j\|_2}{m}} & , m > c_0 q \text{ and } q = p, \\ c_1\sqrt{|H_j| n\log(p)\log^2(n)\|G_j\|_2} & , m > c_0 q \text{ and } q < p, \end{cases}$$

with probability at least

$$1-c_1\exp\{-cn\}-c_2\exp\{-c\log(n)\}-c_3\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}-c_4\exp\{-cn(m/q)^{-\mathbf{1}\{q>c_0m\}}\}-c_5\exp\{-c\log$$

Then using Lemma B.2.3.5 to bound $|\hat{w}_j^\top \Sigma_b^{-1/2} X_j|$ and Theorem B.1.1 to bound $\|\hat{u}\|_1$, we can obtain the following bounds for (B.19):

$$\left|\left(e_j^\top - \frac{\hat{w}_j^\top \Sigma_b^{-1/2} X}{\hat{w}_j^\top \Sigma_b^{-1/2} X_j}\right)\hat{u}\right| \leq \begin{cases} c_1 s\sqrt{|H_j|\frac{q\log^2(p)\log(n)\|G_j\|_2}{m^2n^2\sigma_{\min}^2(G_j)}} & , q > c_0 m, \\ c_1 s\sqrt{|H_j|\frac{\log^2(p)\|G_j\|_2}{m^3n^2\sigma_{\min}^2(G_j)}} & , m > c_0 q \text{ and } q = p, , \\ c_1 s\sqrt{|H_j|\frac{\log^2(p)\log^2(n)\|G_j\|_2}{mn^2\sigma_{\min}^2(G_j)}} & , m > c_0 q \text{ and } q < p. \end{cases}$$

with probability at least $1 - c_1\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - c_2\exp\{-c\log(n)\} - c_3\exp\{-cn\} - c_4\exp\{-c\log(p)\} - c_5\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\} - c_6\exp\{-cmn\}$. Moreover, under Assumption 10.3, we have (B.19) $= o_p(1)$.

Next we show that with random matrices X^i and Z^i, the term (B.20) is asymptotically normal. Note that conditional on X, the terms $\hat{w}_j$, Σ_b, Σ_{θ^*} are fixed quantities. Denote

$$\xi_i = \frac{(\Sigma_{\theta^*}^i)^{1/2}(\Sigma_b^i)^{-1/2}\hat{w}_j^i}{\|\Sigma_{\theta^*}^{1/2}\Sigma_b^{-1/2}\hat{w}_j\|_2}, \quad \tilde{\epsilon}_i = (\Sigma_{\theta^*}^i)^{-1/2}(y^i - X^i\beta^*), \quad i = 1,\ldots,n$$

and let ξ and $\tilde{\epsilon}$ be the vectors formed by vertically stacking ξ_i's and $\tilde{\epsilon}_i$'s, respectively. Then, $\|\xi\|_2^2 = 1$, $\mathbb{E}(\tilde{\epsilon}_i|X) = 0$, $\mathrm{Var}(\tilde{\epsilon}_i|X) = I_m$, and $\xi_i^\top\tilde{\epsilon}_i$ is independent of $\xi_j^\top\tilde{\epsilon}_j$ for $i \neq j$. Recalling

that

$$V_j = \frac{\left\|\Sigma_{\theta^*}^{1/2}\Sigma_b^{-1/2}\hat{w}_j\right\|_2^2}{\left|\hat{w}_j^\top \Sigma_b^{-1/2} X_j\right|^2},$$

we have for the term (B.20) that

$$\frac{1}{\sqrt{V_j}} \frac{\hat{w}_j^\top \Sigma_b^{-1/2}(y - X\beta^*)}{\hat{w}_j^\top \Sigma_b^{-1/2} X_j} = \sum_{i=1}^n \xi_i^\top \tilde{\epsilon}_i.$$

In the following, we first use the Lyapunov Central Limit Theorem to show the (conditional) asymptotic normality of $\sum_{i=1}^n \xi_i^\top \tilde{\epsilon}_i$ given X, and then establish the unconditional asymptotic normality.

We need to verify the following three conditions for the Lyapunov Central Limit Theorem to hold:

1. $\forall\ 1 \le i \le n$, $\mathbb{E}(\xi_i \tilde{\epsilon}_i) < \infty$.
2. $\forall\ 1 \le i \le n$, $\mathrm{Var}(\xi_i \tilde{\epsilon}_i \mid X) < \infty$.
3. Defining $s_n^2 = \sum_{i=1}^n \mathrm{Var}(\xi_i \tilde{\epsilon}_i \mid X)$, for some $\delta > 0$,

$$\lim_{n \to \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E}\left(\left|\xi_i \tilde{\epsilon}_i - \mathbb{E}(\xi_i \tilde{\epsilon}_i)\right|^{2+\delta} \mid X\right) = 0.$$

The first two conditions are trivially satisfied since $\mathbb{E}(\xi_i \tilde{\epsilon}_i) = 0$ and $\mathrm{Var}(\xi^\top \tilde{\epsilon} \mid X) = 1$. We show the third condition is satisfied at $\delta = 2$. Because γ_i's and ϵ_i's are independent sub-Gaussian vectors, the vectors $\tilde{\epsilon}_i$'s is also sub-Gaussian with parameter $c \asymp 1$. Then $\mathbb{E}(|\xi_i^\top \tilde{\epsilon}_i|^4) \le 16\Gamma(2)\|\xi_i\|_2^4$ based on Proposition 3.2 in [181]. Thus, we have

$$\begin{aligned}\frac{1}{s_n^4} \sum_{i=1}^n \mathbb{E}\left(\left|\xi_i^\top \tilde{\epsilon}_i - \mathbb{E}(\xi_i^\top \tilde{\epsilon}_i)\right|^4 \mid X\right) &= \sum_{i=1}^n \mathbb{E}\left(\left|\xi_i^\top \tilde{\epsilon}_i\right|^4 \mid X\right) \le c_1 \sum_{i=1}^n \|\xi_i\|_2^4 \le c_1 \max_i \|\xi_i\|_2^2 \times \|\xi\|_2^2 \\ &= c_1 \frac{\max_i \left\|(\Sigma_{\theta^*}^i)^{1/2}(\Sigma_b^i)^{-1/2}\hat{w}_j^i\right\|_2^2}{\left\|\Sigma_{\theta^*}^{1/2}\Sigma_b^{-1/2}\hat{w}_j\right\|_2^2}. \end{aligned} \tag{B.21}$$

Based on Lemma B.2.3.7, under Assumption 10.3 and Assumption 11, with probability at least $1 - c_1 \exp\{-cmnq^{-\mathbf{1}\{q>c_0 m\}}\} - c_2 \exp\{-c\log(n)\} - c_3 \exp\{-cn\} - c_4 \exp\{-c\log(p)\} - c_5 \exp\{-cn(m/q)^{\mathbf{1}\{q>c_0 m\}}\} - c_6 \exp\{-cmn\}$, we have (B.21) $= o_p(\log^{-2}(n))$, and thus the third condition of the Lyapunov Central Limit Theorem is verified.

Notice that here we use a different approach from that in [132] to show (B.21) $\ll 1$, and this difference is critical. [132] directly assumes $\sigma_{\min}\left(\left(\Sigma_b^i\right)^{-1/2}\Sigma_\theta^i\left(\Sigma_b^i\right)^{-1/2}\right)\mathrm{tr}\left((\Sigma_a)^{-1}\right) \gg m\log(n)$ in order to bound (B.21). However, under our settings, based on Lemma B.1.2, we can show that this assumption implies $n \gg p\log(n)$, which greatly restricts how fast p can grow relative to n. Therefore, in our proof, we explicitly discuss the rate of the terms related to this assumption in Lemma B.2.3.7. By imposing the structural assumptions on the matrix Ψ in Assumption 9, we establish the asymptotic normality of $\hat{\beta}_{j,k}^{(db)}$ with relatively mild sample size assumption. In particular, we allow p to grow faster than n and m.

Finally we show the unconditional asymptotic normality of the term (B.20): Based on [243], suppose we have a sequence of independent random variables U_l, $l = 1, \ldots, n, \ldots$, with $\mathbb{E}(U_l) = \mu_l$, $\mathbb{E}(|U_l - \mu_l|^2) = \sigma_l^2$, $\mathbb{E}(|U_l - \mu_l|^3) = b_l$, $F_n(t)$ being the cumulative density function of the variable

$$\frac{\sum_{l=1}^n (U_l - \mu_l)}{\sqrt{\sum_{l=1}^n \sigma_l^2}},$$

and $\Phi_0(t)$ being the cumulative density function of the standard normal distribution. Then,

$$\sup_t |F_n(t) - \Phi_0(t)| \leq c_1 \log(n) \frac{\sum_{l=1}^n b_l}{\sum_{l=1}^n \sigma_l^2 \sqrt{\sum_{l=1}^n \sigma_l^2}}.$$

In our case, for $l = 1, \ldots, n$, $U_l = \xi_l^\top \tilde{\epsilon}_l$, and we have $\mu_l = 0$, $\sum_{l=1}^n \sigma_l^2 = 1$. Then we have:

$$\begin{aligned}
&\sum_{l=1}^n \mathbb{E}(|U_l - \mu_l|^3 | X) \leq \sum_{l=1}^n \sqrt{\mathbb{E}(U_l^2|X)\mathbb{E}(U_l^4|X)} \quad \text{(Hoeffding's inequality)} \\
&\leq \sqrt{\sum_{l=1}^n \mathbb{E}(U_l^2|X) \sum_{l=1}^n \mathbb{E}(U_l^4|X)} \quad \text{(Cauchy's inequality)} \\
&\leq \sqrt{\sum_{l=1}^n \|\xi_l\|_2^4}.
\end{aligned}$$

As shown previously in (B.21), under under Assumption 10.3 and Assumption 11, we have $\sqrt{\sum_{l=1}^n \|\xi_l\|_2^4} = o_p\left(\log^{-1}(n)\right)$ for any X. Thus $\sup_t |F_n(t) - \Phi_0(t)| = o_p(1)$, indicating the unconditional asymptotic normality of $\xi^\top \tilde{\epsilon}$, i.e., the unconditional asymptotic normality of term (B.20). □

B.2.3 Proof of related lemmas for Theorem B.2.2

Proof of Lemma B.2.1.

First note that $\sigma_{\min}(\Psi)\sigma_{\min}\left(Z^i(Z^i)^\top\right) + \sigma_{\min}(R^i) \leq \sigma\left(\Sigma_\theta^i\right) \leq \|\Psi\|_2 \sigma_{\max}\left(Z^i(Z^i)^\top\right) + \|R^i\|_2$. Using Lemma B.1.2.4, with probability at least $1 - 2\exp\{-cm \vee q\}$, we have

$$\begin{aligned}
\sigma_{\min}\left(Z^i(Z^i)^\top\right) &\geq \begin{cases} c_1 q & , \text{ if } q > c_0 m, \\ 0 & , \text{ if } m > c_0 q, \end{cases} \\
\sigma_{\max}\left(Z^i(Z^i)^\top\right) &\leq c_2 m \vee q.
\end{aligned}$$

Then under Assumption 8 and Assumption 9.2, we have

$$\begin{cases} c_3 \leq \sigma\left(\Sigma_\theta^i\right) \leq c_4 m & , m > c_0 q, \\ \sigma\left(\Sigma_\theta^i\right) \asymp q & , q > c_0 m. \end{cases}$$

Under Assumption 9.1, note that we have $\min(\psi_{S_\psi}) Z^i_{S_\psi}(Z^i_{S_\psi})^\top \preceq Z^i \Psi (Z^i)^\top \preceq \max(\psi_{S_\psi}) Z^i_{S_\psi}(Z^i_{S_\psi})^\top$ with $s_\psi < c_5 m$. Thus, using Lemma B.1.2.4, with probability at

least $1-2\exp\{-cm\}$, we have

$$\sigma_{\max}\left(Z^i(Z^i)^\top\right) \leq c_6 m.$$

Then under Assumption 8 and Assumption 9.1, for both cases of $q > c_0 m$ and $m > c_0 q$ we get

$$c_7 \leq \sigma\left(\Sigma_\theta^i\right) \leq c_8 m.$$

□

Proof of Lemma B.2.3.

Lemma B.2.3.1)

Recalling the definition of z_j^* in (B.18), we can rewrite the expression of z_j^* as

$$z_j^* = \frac{1}{\mathrm{tr}(\Sigma_b^{-1})} \max_{l\neq j, 1\leq l\leq p} \left| X_l^\top \Sigma_b^{-1}(X_j - X_{-j}\kappa_j^*)\right|.$$

Under Assumption 10.1 and using the tail bound of sub-Gaussian random variables, we can get

$$\mathbb{P}\left(z_j^* > t \mid X_{-j}\right) \leq 2\exp\left\{\log(p) - c_1 \frac{t^2 \mathrm{tr}^2\left(\Sigma_b^{-1}\right)}{\|G_j\|_2 \max_{l\neq j,1\leq l\leq p}\|X_l^\top \Sigma_b^{-1}\|_2^2}\right\}. \tag{B.22}$$

For the term $\max_{l\neq j}\|X_l^\top \Sigma_b^{-1}\|_2^2$:

1. When $l \leq q$, $l \neq j$: $X_l^i = Z_l^i$. Then by Lemma B.1.2.10,

$$\begin{aligned}\max_{l\neq j}\|X_l^\top \Sigma_b^{-1}\|_2^2 &\leq \sigma_{\max}\left(\sum_{i=1}^n \left(Z_{-j}^i\right)^\top \left(\Sigma_b^i\right)^{-2} Z_{-j}^i\right) \max_l \|e_l\|_2^2 \\ &\leq \begin{cases} c_1 \frac{n}{q}\left(1 \wedge \frac{m\log(n)}{q}\right) & , q > c_0 m, \\ c_1 \frac{n}{m} & , m > c_0 q, \end{cases}\end{aligned}$$

with probability at least $1-2\exp\{-cn\}-2\exp\{-c\log(n)\}-\exp\left\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\right\}$.

2. When $l > q$, $l \neq j$: Since there exist a column X_l with $l > q$, it implies that $p > q$. Then based on Lemma B.1.2.7 and Lemma B.1.3.1, we have

$$\begin{aligned}\max_{l\neq j}\|X_l^\top\Sigma_b^{-1}\|_2^2 &\leq \max_l \|e_l\|_2^2\sigma_{\max}\left(X_{-j}^\top\Sigma_b^{-1}X_{-j}\right)\sigma_{\max}\left(\Sigma_b^{-1}\right)\\ &\leq \begin{cases} c_2\frac{mn}{q}\left(\frac{\log(n)}{q}\wedge 1\right) & ,q > c_0m,\\ c_2n(\log(n)\wedge m) & ,m > c_0q,\end{cases}\end{aligned}$$

with probability at least $1-4\exp\{-cn\}-4\exp\{-c\log(n)\}-\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\}-2\exp\{-cnmq^{-\mathbf{1}\{q>c_0m\}}\}$.

Combining the above two cases for $l \leq q$ and $l > q$, we obtain the following bounds:

$$\max_{l\neq j}\|X_l^\top\Sigma_b^{-1}\|_2^2 \leq \begin{cases} c_1\frac{n}{q}\left(1\wedge\frac{m\log(n)}{q}\right) & ,q > c_0m \text{ and } p = q,\\ c_1\frac{n}{m} & ,m > c_0q \text{ and } p = q,\\ c_1\frac{mn}{q}\left(\frac{\log(n)}{q}\wedge 1\right) & ,q > c_0m \text{ and } p > q,\\ c_1n(\log(n)\wedge m) & ,m > c_0q \text{ and } p > q.\end{cases} \tag{B.23}$$

Plugging (B.23) into (B.22) and taking $t = c_2\sqrt{\log(p)\max_{l\neq j}\|X_j^\top\Sigma_b^{-1}\|_2^2\|G_j\|_2}/\mathrm{tr}(\Sigma_b^{-1})$, we obtain the stated bounds for z_j^* with probability at least $1-2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}-4\exp\{-c\log(n)\}-4\exp\{-cn\}-2\exp\{-c\log(p)\}-\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\}$.

Lemma B.2.3.2)

Let $\hat{v} = \hat{\kappa}_j - \kappa_j^*$. Then based on the definition of $\hat{\kappa}_j$, we have

$$\frac{1}{2\,\mathrm{tr}\left(\Sigma_b^{-1}\right)}\left\|\Sigma_b^{-1/2}\left(X_j - X_{-j}\hat{\kappa}_j\right)\right\|_2^2 + \lambda_j\|\hat{\kappa}_j\|_1 \leq \frac{1}{2\,\mathrm{tr}\left(\Sigma_b^{-1}\right)}\left\|\Sigma_b^{-1/2}\left(X_j - X_{-j}\kappa_j^*\right)\right\|_2^2 + \lambda_j\|\kappa_j^*\|_1.$$

Rearranging the terms, we get

$$0 \leq \frac{1}{2\,\mathrm{tr}\left(\Sigma_b^{-1}\right)} \left\|\Sigma_b^{-1/2} X_{-j}\hat{v}\right\|_2^2 \leq (z_j^* + \lambda_j)\|\hat{v}_{H_j}\|_1 + (z_j^* - \lambda_j)\|\hat{v}_{H_j^c}\|_1$$

where z_j^* is defined in (B.18). Based on Lemma B.1.3.1, we have $\left\|\Sigma_b^{-1/2} X_{-j}\hat{v}\right\|_2^2 \geq \|\hat{v}\|_2^2 \sigma_{\min}(X_{-j}^\top \Sigma_b^{-1} X_{-j}) \geq c_1\|\hat{v}\|_2^2 n(m/q)^{\mathbf{1}\{q>c_0 m\}}$. Lemma B.2.3.1 gives the upper bound for z_j^*. Then following the same arguments in the proof of Theorem B.1.1, we can obtain the stated bounds with probability at least $1 - c_1\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0 m\}}\} - c_2\exp\{-cn\} - c_3\exp\{-c\log(n)\} - c_4\exp\{-cmnq^{-\mathbf{1}\{q>c_0 m\}}\} - c_5\exp\{-c\log(p)\}$.

Lemma B.2.3.3)

Let $\hat{v} = \hat{\kappa}_j - \kappa_j^*$. Then we have:

$$\max_{l\neq j}\left|X_l^\top \Sigma_b^{-1} X_{-j}\hat{v}\right| \leq \max_{l\neq j}\|\Sigma_b^{-1/2} X_l\|_2 \|\Sigma_b^{-1/2} X_{-j}\hat{v}\|_2.$$

We can first write

$$\begin{aligned}\max_{l\neq j}\|\Sigma_b^{-1/2} X_l\|_2^2 &\leq \min\left(\max_{l\neq j}\|X_l\|_2^2 \sigma_{\max}\left(\Sigma_b^{-1}\right), \max_{l\neq j}\|e_l\|_2^2 \sigma_{\max}\left(X_{-j}^\top \Sigma_b^{-1} X_{-j}\right)\right) \\ &\leq \begin{cases} c_1\frac{nm}{q} & , q > c_0 m, \\ c_1 n(m \wedge \log(n)) & , m > c_0 q \text{ and } q < p, \\ c_1 n & , m > c_0 q \text{ and } q = p, \end{cases}\end{aligned}$$

where we have used Lemma B.1.2.7 to bound $\sigma_{\max}\left(\Sigma_b^{-1}\right)$, Lemma B.1.3.1 to bound $\sigma_{\max}\left(X_{-j}^\top \Sigma_b^{-1} X_{-j}\right)$, and the tail bound for sum of square of independent Gaussian variables to bound $\max_{l\neq j}\|X_l\|_2^2$:

$$\mathbb{P}\left(\max_l \left|\|X_l\|_2^2 - \sum_{i=1}^n m\left(\Sigma_X^i\right)_{l,l}\right| < c_2 nm\right) > 1 - 2\exp\{-cmn\} \quad \left(\text{under Assumption 10.1 } \frac{\log(p)}{mn} = o(1)\right)$$

Then based on the bounds for $\|\Sigma_b^{-1/2} X_{-j}\hat{v}\|_2$ in Lemma B.2.3.2, we obtain the states bounds

in the lemma, with probability at least $1 - c_1 \exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - c_2 \exp\{-c\log(n)\} - c_3 \exp\{-cn\} - c_4 \exp\{-c\log(p)\} - c_5 \exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\} - c_6 \exp\{-cmn\}$.

Lemma B.2.3.4)

Conditioning on X_{-j}, we have $X_j - X_{-j}\kappa_j^* \in \mathrm{SGV}(c\|G_j\|_2)$ (Assumption 10.1). Then based on Corollary 2.8 of [244], we can get:

$$\mathbb{P}\left(\left|\|w_j\|_2^2 - \mathbb{E}\left(\|w_j\|_2^2 \mid X_{-j}\right)\right| > t \mid X_{-j}\right) \leq 2\exp\left\{-c_1 \min\left(\frac{t^2}{\|\Sigma_b^{-1}\|_F^2\|G_j\|_2^2}, \frac{t}{\|\Sigma_b^{-1}\|_F\|G_j\|_2}\right)\right\}.$$

Here based on Lemma B.1.2.7 and Lemma B.1.2.8, we have

$$\mathbb{E}\left(\|w_j\|_2^2 \mid X_{-j}\right) = \sum_{i=1}^{n} \mathrm{tr}\left((\Sigma_b^i)^{-1}G_j\right) \in [c_1\sigma_{\min}(G_j)nmq^{-\mathbf{1}\{q>c_0m\}}, c_2\sigma_{\max}(G_j)nmq^{-\mathbf{1}\{q>c_0m\}}],$$
$$\|\Sigma_b^{-1}\|_F^2 = \mathrm{tr}^2(\Sigma_b^{-1}) \leq \|\Sigma_b^{-1}\|_2 \,\mathrm{tr}(\Sigma_b^{-1}) \leq c_3 nmq^{-\mathbf{1}\{q>c_0m\}}\left(1 \wedge \frac{\log(n)}{m \vee q}\right).$$

Thus, under Assumption 10.2, we have $\mathbb{E}\left(\|w_j\|_2^2 \mid X_{-j}\right) \geq c_1\log(n)\|\Sigma_b^{-1}\|_F\|G_j\|_2$ for suitably small $c_1 > 0$. Therefore, we get $c_4\sigma_{\min}(G_j)nmq^{-\mathbf{1}\{q>c_0m\}} \leq \|w_j\|_2^2 \leq c_5\sigma_{\max}(G_j)nmq^{-\mathbf{1}\{q>c_0m\}}$ with probability at least $1 - 4\exp\{-cn\} - 2\exp\{-c\log(n)\}$.

For w_j^i, we similarly condition on X_{-j} and use Corollary 2.8 of [244] to get

$$\begin{aligned}&\mathbb{P}\left(\forall\, i, \left|\left\|w_j^i\right\|_2^2 - \mathbb{E}\left(\left\|w_j^i\right\|_2^2 \mid X_{-j}\right)\right| > t \mid X_{-j}\right) \\ &\quad\leq 2\exp\left\{\log(n) - c_1 \min\left(\frac{t^2}{\max_i \|(\Sigma_b^i)^{-1}\|_F^2\|G_j\|_2^2}, \frac{t}{\max_i \|(\Sigma_b^i)^{-1}\|_F\|G_j\|_2}\right)\right\}. \quad \text{(B.24)}\end{aligned}$$

Here, based on Lemma B.1.2.7 and Lemma B.1.2.8, we have

$$\mathbb{E}\left(\left\|w_j^i\right\|_2^2 \mid X_{-j}\right) = \mathrm{tr}\left((\Sigma_b^i)^{-1}G_j\right) \in \begin{cases} c_1 m\left(1 \wedge \frac{\log(n)}{q}\right)[\sigma_{\min}(G_j), \sigma_{\max}(G_j)] & , q > c_0 m, \\ c_1\left(m + q\left(1 \wedge \frac{\log(n)}{m}\right)\right)[\sigma_{\min}(G_j), \sigma_{\max}(G_j)] & , m > c_0 q, \end{cases}$$
$$\|(\Sigma_b^i)^{-1}\|_F^2 = \mathrm{tr}^2((\Sigma_b^i)^{-1}) \leq \|(\Sigma_b^i)^{-1}\|_2 \,\mathrm{tr}((\Sigma_b^i)^{-1}) \leq \begin{cases} c_2 m\left(1 \wedge \frac{\log(n)}{q}\right)\left(1 \wedge \frac{\log(n)}{q}\right) & , q > c_0 m, \\ c_2\left(m + q\left(1 \wedge \frac{\log(n)}{m}\right)\right)\left(1 \wedge \frac{\log(n)}{m}\right) & , m > c_0 q, \end{cases}$$

Then taking $t = c_3 \log(n) \max_i \|(\Sigma_b^i)^{-1}\|_F \|G_j\|_2$ in (B.24), we obtain

$$\max_i \|w_j^i\|_2^2 \leq \begin{cases} c_1(m + \sqrt{m}\log(n)) \left(1 \wedge \frac{\log(n)}{q}\right) \|G_j\|_2 & , q > c_0 m, \\ c_1 \left(m + \sqrt{m\left(1 \wedge \frac{\log(n)}{m}\right)} \log(n)\right) \|G_j\|_2 & , m > c_0 q \end{cases}$$

with probability at least $1 - 6\exp\{-c\log(n)\}$.

Lemma B.2.3.5)

Define $\hat{v} = \hat{\kappa}_j - \kappa_j^*$. Then $\hat{w}_j - w_j = -\Sigma_b^{-1/2} X_{-j} \hat{v}$, and by definition of z_j^* we have $\|X_{-j}^\top \Sigma_b^{-1/2} w_j\|_\infty = z_j^* \operatorname{tr}(\Sigma_b^{-1})$. Thus we can write

$$\begin{aligned} |\hat{w}_j^\top \Sigma_b^{-1/2} X_j| &= |\hat{w}_j^\top \Sigma_b^{-1/2} (X_j - X_{-j}\kappa_j^* + X_{-j}\kappa_j^*| \\ &\geq |\hat{w}_j^\top w_j| - |\hat{w}_j^\top \Sigma_b^{-1/2} X_j \kappa_j^*| \\ &\geq |w_j^\top w_j| - |(\hat{w}_j - w_j)^\top w_j| - |w_j^\top \Sigma_b^{-1/2} X_{-j}\kappa_j^*| - |(\hat{w}_j - w_j)^\top \Sigma_b^{-1/2} X_{-j}\kappa_j^*| \\ &= \|w_j\|_2^2 - \|w_j^\top \Sigma_b^{-1/2} X_{-j}\hat{v}| - \|w_j^\top \Sigma_b^{-1/2} X_{-j}\kappa_j^*| - \|\hat{v}^\top X_{-j}^\top \Sigma_b^{-1} X_{-j}\kappa_j^*| \\ &\geq \|w_j\|_2^2 - z_j^* \operatorname{tr}\left(\Sigma_b^{-1}\right) \left(\|\hat{v}\|_1 + \|\kappa_j^*\|_1\right) - \|X_{-j}^\top \Sigma_b^{-1} X_{-j}\hat{v}\|_\infty \|\kappa_j^*\|_1. \end{aligned}$$

Then plugging in the bound for $\|\kappa_j^*\|$ from Assumption 10.1, the bound for z_j^* from Lemma B.2.3.1, the bound for $\|X_{-j}^\top \Sigma_b^{-1} X_{-j}\hat{v}\|_\infty$ from Lemma B.2.3.3, the bound for $\|w_j\|_2^2$ from Lemma B.2.3.4, the bound for $\|\hat{v}\|_1$ from Lemma B.2.3.2, and the bound for $\operatorname{tr}\left(\Sigma_b^{-1}\right)$ from Lemma B.1.2.8, under Assumption 10:3, we obtain

$$|\hat{w}_j^\top \Sigma_b^{-1/2} X_j| \geq c_1 \sigma_{\min}(G_j) nmq^{-\mathbf{1}\{q > c_0 m\}}$$

with probability at least $1 - c_1\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - c_2\exp\{-c\log(n)\} - c_3\exp\{-cn\} - c_4\exp\{-c\log(p)\} - c_5\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\} - c_6\exp\{-cmn\}$.

Lemma B.2.3.6)

Define $\hat{v} = \hat{\kappa}_j - \kappa_j^*$. Then based on Lemma B.2.3.2 and Lemma B.2.3.4, and under Assumption 10.3, with probability at least $1-c_1\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}-c_2\exp\{-c\log(n)\}-c_3\exp\{-cn\}-c_4\exp\{-c\log(p)\}-c_5\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\}-c_6\exp\{-cmn\}$ we have

$$\|\hat{w}_j\|_2^2 = \|w_j - \Sigma_b^{-1/2}X_{-j}\hat{v}\|_2^2 \geq \|w_j\|_2^2 - \|\Sigma_b^{-1/2}X_{-j}\hat{v}\|_2^2 \geq c_1 nmq^{-\mathbf{1}\{q>c_0m\}}\sigma_{\min}(G_j), \tag{B.25}$$

$$\|\hat{w}_j\|_2^2 \leq \|w_j\|_2^2 + \|\Sigma_b^{-1/2}X_{-j}\hat{v}\|_2^2 \leq c_2 nmq^{-\mathbf{1}\{q>c_0m\}}\sigma_{\max}(G_j). \tag{B.26}$$

For $\max_i \|\hat{w}_j^i\|_2^2$, we follow the similar arguments as (B.25) and (B.26) and have:

$$\max_i \|\hat{w}_j^i\|_2^2 \leq \max_i \|w_j^i\|_2^2 + \|\hat{v}\|_2^2 \max_i \sigma_{\max}\left(\left(X_{-j}^i\right)^\top (\Sigma_b^i)^{-1}X_{-j}^i\right).$$

Then, based on Lemma B.1.3.2, Lemma B.2.3.2, Lemma B.2.3.4, and under Assumption 10.3, we have

$$\|\hat{v}\|_2^2 \max_i \sigma_{\max}\left(\left(X_{-j}^i\right)^\top (\Sigma_b^i)^{-1}X_{-j}^i\right) \ll \max_i \|w_j^i\|_2^2,$$

and thus

$$\max_i \|\hat{w}_j^i\|_2^2 \leq \begin{cases} c_1(m+\sqrt{m}\log(n))\left(1 \wedge \frac{\log(n)}{q}\right)\|G_j\|_2 & , q > c_0 m, \\ c_1\left(m+\log(n)\sqrt{m\wedge\log(n)}\right)\|G_j\|_2 & , m > c_0 q, \end{cases}$$

with probability at least $1 - c_1\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - c_2\exp\{-c\log(n)\} - c_3\exp\{-cn\} - c_4\exp\{-c\log(p)\} - c_5\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\} - c_6\exp\{-cmn\}$.

Lemma B.2.3.7)

1. For B.2.3.7.1: Recall that $\Sigma_b^i = aZ_{-j}^i(Z_{-j}^i)^\top + I_m$. We can show that with probability

at least $1 - c_1 \log(n)$, we have

$$\begin{aligned}
\sigma_{\max}\left(\Sigma^i_{\theta^*}\left(\Sigma^i_b\right)^{-1}\right) &\leq \|\Psi\|_2 \sigma_{\max}\left((Z^i)^\top (\Sigma^i_b)^{-1} Z^i\right) + \|R^i\|_2 \sigma_{\max}\left((\Sigma^i_b)^{-1}\right) \\
&\leq \begin{cases} \|\Psi\|_2 \sigma_{\max}\left((Z^i)^\top (\Sigma^i_a)^{-1} Z^i\right) + \|R^i\|_2 \sigma_{\max}\left((\Sigma^i_a)^{-1}\right) & , \text{if } j > q \\ \|\Psi\|_2 \sigma_{\max}\left((Z^i_{-j})^\top (\Sigma^i_b)^{-1} Z^i_{-j}\right) + \|R^i\|_2 \sigma_{\max}\left((\Sigma^i_b)^{-1}\right) & \\ \qquad + \|\Psi\|_2 (Z^i_j)^\top (\Sigma^i_b)^{-1} Z^i_j & , \text{if } j \leq q \end{cases} \\
&\leq \begin{cases} c_2 & , \text{if } j > q \\ c_2 + c_3 \|Z^i_j\|_2^2 \sigma_{\max}\left((\Sigma^i_b)^{-1}\right) & , \text{if } j \leq q \end{cases} \\
&\leq \begin{cases} c_2 & , \text{if } j > q, \\ c_2 + c_4 m \log(n) \left(1 \wedge \frac{\log(n)}{m \vee q}\right) & , \text{if } j \leq q. \end{cases} \\
&\leq c_5 \log^2(n)
\end{aligned}$$

where we have used Lemma B.1.2.7 to bound $\sigma_{\max}\left((\Sigma^i_b)^{-1}\right)$, Lemma B.1.3.2 to bound $\sigma_{\max}\left((Z^i)^\top (\Sigma^i_a)^{-1} Z^i\right)$ and $\sigma_{\max}\left((Z^i_{-j})^\top (\Sigma^i_b)^{-1} Z^i_{-j}\right)$, with $\|\Psi\|_2 \asymp \|R^i\|_2 \asymp 1$ under Assumption 8.2, and the following tail bound for sub-Gaussian random variables:

$$\mathbb{P}\left(\max_i \left| \|Z^i_j\|_2^2 - m(\Sigma^i_Z)_{j,j} \right| < c_1 m \log(n) \mid \Sigma^i_Z \right) \geq 1 - 2\exp\{-c \log(n)\}. \tag{B.27}$$

We can also show that, under Assumption 9.2,

$$\begin{aligned}
\sigma_{\min}\left(\Sigma^i_{\theta^*}(\Sigma^i_b)^{-1}\right) &\geq \sigma_{\min}\left(\left(\sigma_{\min}(\Psi)(Z^i_{-j})^\top Z^i_{-j} + \sigma_{\min}(R^i) I_m\right)(\Sigma^i_b)^{-1}\right) \\
&\geq c_1 \min_l \frac{\sigma_l^2(Z^i_{-j}) + c_2}{a\sigma_l^2(Z^i_{-j}) + 1} \\
&\geq c_3.
\end{aligned}$$

Then, by Lemma B.2.3.6, we can show that

$$\begin{aligned}
\hat{w}_j^\top \Sigma_b^{-1/2} \Sigma_{\theta^*} \Sigma_b^{-1/2} \hat{w}_j &\geq \|\hat{w}_j\|_2^2 \sigma_{\min}\left(\Sigma_{\theta^*}^i \left(\Sigma_b^i\right)^{-1}\right) \\
&\geq c_1 nmq^{-\mathbf{1}\{q > c_0 m\}} \sigma_{\min}(G_j) \\
\max_i \|(\Sigma_{\theta^*}^i)^{1/2}(\Sigma_b^i)^{-1/2}\hat{w}_j^i\|_2^2 &\leq \max_i \sigma_{\max}\left(\Sigma_{\theta^*}^i(\Sigma_b^i)^{-1}\right) \max_i \|\hat{w}_j^i\|_2^2 \\
&\leq \begin{cases} c_1(m + \sqrt{m}\log(n))\left(1 \wedge \frac{\log(n)}{q}\right)\log^2(n)\|G_j\|_2 & , q > c_0 m, \\ c_1(m + \log(n)\sqrt{m \wedge \log(n)})\log^2(n)\|G_j\|_2 & , m > c_0 q, \end{cases}
\end{aligned}$$

with probability at least $1 - c_1\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - c_2\exp\{-c\log(n)\} - c_3\exp\{-cn\} - c_4\exp\{-c\log(p)\} - c_5\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\} - c_6\exp\{-cmn\}$.

2. For B.2.3.7.2: First note that based on the arguments in Lemma B.1.4.1: $\sigma_{\max}(Z_{S_\psi}^i(Z_{S_\psi}^i)^\top) \leq c_1 m\log(n)$ with probability at least $1 - 2\exp\{-c\log(n)\}$. So we have

$$\begin{aligned}
&\sigma_{\max}((Z_{S_\psi}^i)^\top(\Sigma_b^i)^{-1}Z_{S_\psi}^i) \leq \sigma_{\max}(Z_{S_\psi}^i(Z_{S_\psi}^i)^\top)\sigma_{\max}((\Sigma_b^i)^{-1}) \leq c_1\frac{m\log^2(n)}{m \vee q} \\
&\sigma_{\max}\left(\Sigma_{\theta^*}^i(\Sigma_b^i)^{-1}\right) \leq \|\Psi\|_2\sigma_{\max}((Z_{S_\psi}^i)^\top(\Sigma_b^i)^{-1}Z_{S_\psi}^i) + c_2\sigma_{\max}((Z_{S_\psi}^i)^\top Z_{S_\psi}^i) \leq c_3\frac{m\log^2(n)}{m \vee q}.
\end{aligned}$$

Then, based on Lemma B.2.3.5, we can show that

$$\begin{aligned}
\max_i \|(\Sigma_{\theta^*}^i)^{1/2}(\Sigma_b^i)^{-1/2}\hat{w}_j^i\|_2^2 &\leq \max_i \sigma_{\max}\left(\Sigma_{\theta^*}^i(\Sigma_b^i)^{-1}\right) \max_i \|\hat{w}_j^i\|_2^2 \\
&\leq \begin{cases} c_1(m + \sqrt{m}\log(n))\left(1 \wedge \frac{\log(n)}{q}\right)\log^2(n)\frac{m}{q}\|G_j\|_2 & , q > c_0 m, \\ c_1(m + \log(n)\sqrt{m \wedge \log(n)})\log^2(n)\|G_j\|_2 & , m > c_0 q. \end{cases}
\end{aligned}$$

We can also write

$$\begin{aligned}
\hat{w}_j^\top \Sigma_b^{-1/2} \Sigma_{\theta^*} \Sigma_b^{-1/2} \hat{w}_j &= \sum_{i=1}^n (\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2} \left(\sum_{l \in S_\psi} \psi_l Z_l^i (Z_l^i)^\top + R^i \right) (\Sigma_b^i)^{-1/2} \hat{w}_j^i \\
&\geq c_1 \sum_{i=1}^n \left| (Z_l^i)^\top (\Sigma_b^i)^{-1/2} \hat{w}_j^i \right|^2 \\
&\geq \frac{c_1}{n} \left| \sum_{i=1}^n (Z_l^i)^\top (\Sigma_b^i)^{-1/2} \hat{w}_j^i \right|^2 \quad \text{(Cauchy's inequality)} \\
&\geq c_1 n m^2 q^{-2 \times \mathbf{1}\{q > c_0 m\}} \sigma_{\min}^2(G_j). \quad \text{(Lemma B.2.3.5)}
\end{aligned}$$

3. For B.2.3.7.3:

Using Lemma B.2.3.6 and Lemma B.1.2.7, we can show that

$$\begin{aligned}
\hat{w}_j^\top \Sigma_b^{-1/2} \Sigma_{\theta^*} \Sigma_b^{-1/2} \hat{w}_j &\geq \sum_{i=1}^n (\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2} R^i (\Sigma_b^i)^{-1/2} \hat{w}_j^i \\
&\geq c_1 \|\hat{w}_j\|_2^2 \sigma_{\min}\left(\Sigma_b^{-1}\right) \\
&\geq c_1 n m q^{-\mathbf{1}\{q > c_0 m\}} \sigma_{\min}(G_j) \frac{1}{\log(n) + m \vee q}.
\end{aligned}$$

Moreover, denoting $h_l^i = (\Sigma_b^i)^{-1} Z_l^i$, we can also show that

$$\begin{aligned}
\max_i \|(\Sigma_{\theta^*}^i)^{1/2} (\Sigma_b^i)^{-1/2} \hat{w}_j^i\|_2^2 &\leq c_2 \max_i \|R^i\|_2 (\hat{w}_j^i)^\top (\Sigma_b^i)^{-1} \hat{w}_j^i + c_2 s_\psi \max_i \max_{l \in S_\psi} \left| (Z_l^i)^\top (\Sigma_b^i)^{-1/2} \hat{w}_j^i \right|^2 \\
&\leq c_3 \max_i (\hat{w}_j^i)^\top (\Sigma_b^i)^{-1} \hat{w}_j^i \qquad \text{(B.28)} \\
&\quad + c_3 s_\psi \max_i \max_{l \in S_\psi} \left| (h_l^i)^\top (X_j^i - X_{-j}^i \kappa_j^*) \right|^2 \qquad \text{(B.29)} \\
&\quad + c_3 s_\psi \max_i \max_{l \in S_\psi} \left| (h_l^i)^\top X_{-j}^i \hat{v} \right|^2. \qquad \text{(B.30)}
\end{aligned}$$

To bound the term (B.28): Based on Lemma B.1.2.7 and Lemma B.2.3.6, we can get

$$\begin{aligned}\max_i (\hat{w}_j^i)^\top (\Sigma_b^i)^{-1} \hat{w}_j^i &\leq \max_i \sigma_{\max}((\Sigma_b^i)^{-1}) \|\hat{w}_j^i\|_2^2 \\ &\leq \begin{cases} c_1 \left(1 \wedge \frac{\log(n)}{q}\right)^2 (m + \sqrt{m}\log(n)) \|G_j\|_2 & , q > c_0 m, \\ c_1 \left(1 \wedge \frac{\log(n)}{m}\right) (m + \log(n)\sqrt{m \wedge \log(n)}) \|G_j\|_2 & , m > c_0 q, \end{cases}\end{aligned}$$

with probability at least $1 - c_1 \exp\{-cmnq^{-\mathbf{1}\{q>c_0 m\}}\} - c_2 \exp\{-c\log(n)\} - c_3 \exp\{-cn\} - c_4 \exp\{-c\log(p)\} - c_5 \exp\{-cn(m/q)^{\mathbf{1}\{q>c_0 m\}}\} - c_6 \exp\{-cmn\}$.

To bound the second term (B.29): Conditioning on X_{-j}, under Assumption 10.1, we have

$$\begin{aligned}&(h_l^i)^\top (X_j^i - X_{-j}^i \kappa_j^*) \mid X_{-j} \in \mathrm{SG}(c\|h_l^i\|_2^2 \|G_j\|_2) \\ &\left|(h_l^i)^\top (X_j^i - X_{-j}^i \kappa_j^*)\right|^2 - (h_l^i)^\top G_j h_l^i \in \mathrm{SE}\left(c_1 \|h_l^i\|_2^4 \|G_j\|_2^2, c_2 \|h_l^i\|_2^2 \|G_j\|_2\right).\end{aligned}$$

Then we have the following bound

$$\begin{aligned}&\mathbb{P}\left(\max_i \max_{l \in S_\psi} \left|(h_l^i)^\top (X_j^i - X_{-j}^i \kappa_j^*)\right|^2 \leq t + \max_i \max_{l \in S_\psi} \|G_j\|_2 \|h_l^i\|_2^2 \mid X_{-j}\right) \\ &\quad \geq 1 - 2\exp\left\{\log(n) + \log(s_\psi) - c_1 \min\left(\frac{t^2}{\max_i \max_{l \in S_\psi} \|G_j\|_2^2 \|h_l^i\|_2^4}, \frac{t}{\max_i \max_{l \in S_\psi} \|G_j\|_2 \|h_l^i\|_2^2}\right)\right\}\end{aligned}$$

where by Lemma B.1.2.7

$$\begin{aligned}\max_i \max_{l \in S_\psi} \|h_l^i\|_2^2 &= \max_i \max_{l \in S_\psi} (Z_l^i)^\top (\Sigma_b^i)^{-2} Z_l^i \\ &\leq \max_i \sigma_{\max}((\Sigma_b^i)^{-1}) \sigma_{\max}\left((Z_{S_\psi}^i)^\top (\Sigma_b^i)^{-1} Z_{S_\psi}^i\right) \\ &\leq c_1 \left(1 \wedge \frac{\log(n)}{m \vee q} \left(\frac{m\log^2(n)}{q}\right)^{\mathbf{1}\{q>c_0 m\}}\right). \qquad \text{(B.31)}\end{aligned}$$

The last inequality (B.31) holds because $\sigma_{\max}\left((Z_{S_\psi}^i)^\top (\Sigma_b^i)^{-1} Z_{S_\psi}^i\right) \leq c_1 m \log^2(n)/(m \vee q)$ (shown in Lemma B.2.3.7.2), and we also have $\sigma_{\max}\left((Z_{S_\psi}^i)^\top (\Sigma_b^i)^{-1} Z_{S_\psi}^i\right) \leq \sigma_{\max}\left((Z_{-j}^i)^\top (\Sigma_b^i)^{-1} Z_{-j}^i\right) \leq c$ by Lemma B.1.3.2. Thus, we can bound the term

(B.29) by $c_1 s_\psi \log(n) \left(1 \wedge \frac{\log(n)}{m \vee q}\right) \left(\frac{m \log^2(n)}{q}\right)^{\mathbf{1}\{q > c_0 m\}} \|G_j\|_2$ with probability at least $1 - 4\exp\{-c\log(n)\}$.

We finally bound the term (B.30). To this end, we use Lemma B.1.2.7, Lemma B.2.3.2, and the tail bounds for $\|Z^i_{l_1}\|_2^2$, $\|X^i_{l_2}\|_2^2$ based on (B.27) to show that

$$\begin{aligned}
\max_i \max_{l \in S_\psi} \left|(h^i_l)^\top X^i_{-j} \hat{v}\right|^2 &\leq \max_i \max_{l \in S_\psi} \left\|(Z^i_l)^\top (\Sigma^i_b)^{-1} X^i_{-j}\right\|_\infty^2 \|\hat{v}\|_1^2 \\
&= \max_i \max_{l_1 \in S_\psi, l_2 \in S_\psi} \left|(Z^i_{l_1})^\top (\Sigma^i_b)^{-1} X^i_{l_2}\right|^2 \|\hat{v}\|_1^2 \\
&\leq \max_i \max_{l_1 \in S_\psi, l_2 \in S_\psi} \|Z^i_{l_1}\|_2^2 \sigma^2_{\max}((\Sigma^i_b)^{-1}) \|X^i_{l_2}\|_2^2 \|\hat{v}\|_1^2 \\
&\leq \begin{cases} c_1 |H_j|^2 \frac{m \log(p) \log^5(n)}{nq^2} \|G_j\|_2 & , q > c_0 m, \\ c_1 |H_j|^2 \frac{\log(p) \log^4(n)}{mn} \|G_j\|_2 & , m > c_0 q \text{ and } q = p, \\ c_1 |H_j|^2 \frac{\log(p) \log^5(n)}{n} \|G_j\|_2 & , m > c_0 q \text{ and } q < p, \end{cases}
\end{aligned}$$

with probability at least $1 - c_1 \exp\{-cmnq^{-\mathbf{1}\{q > c_0 m\}}\} - c_2 \exp\{-c\log(n)\} - c_3 \exp\{-cn\} - c_4 \exp\{-c\log(p)\} - c_5 \exp\{-cn(m/q)^{\mathbf{1}\{q > c_0 m\}}\} - c_6 \exp\{-cmn\}$.

Therefore, under Assumption 10, we get the stated bounds for $\max_i \|(\Sigma^i_{\theta^*})^{1/2} (\Sigma^i_b)^{-1/2} \hat{w}^i_j\|_2^2$.

□

B.3 Sandwich Estimator for V_j

We propose a sandwich estimator $\hat{V}_j$ for the variance V_j:

$$\hat{V}_j = \frac{\sum_{i=1}^n \left|(\hat{w}^i_j)^\top (\Sigma^i_b)^{-1/2} (y^i - X^i \hat{\beta})\right|^2}{\left|\sum_{i=1}^n (\hat{w}^i_j)^\top (\Sigma^i_b)^{-1/2} X^i_j\right|^2}.$$

We state the generalized version of Theorem 3.3.3 of the main text and its related assumptions for the generic doubly high-dimensional LMM (B.1):

Assumption 12. *For the three conditions defined in Assumption 11:*

1. Under Condition 1:

$$\begin{cases} \frac{\|G_j\|_2}{\sigma_{\min}(G_j)} \ll \frac{n}{s\log^4(n)\log(p)} \wedge \frac{n}{s^2\log^2(p)} & , q > c_0 m \\ \frac{\|G_j\|_2}{\sigma_{\min}(G_j)} \ll \frac{n}{s^2\log(p)\log^3(n)} & , m > c_0 q,\ p = q \\ \frac{\|G_j\|_2}{\sigma_{\min}(G_j)} \ll \frac{n}{sm\log(p)\log^3(n)} \wedge \frac{n\log(n)}{s^2m^2\log(p)} & , m > c_0 q,\ p > q \end{cases}$$

2. Under Condition 2:

$$\begin{cases} \frac{\|G_j\|_2}{\sigma^2_{\min}(G_j)} \ll \frac{n}{s\log^5(n)\log(p)} \wedge \frac{n}{s^2\log^2(n)\log^2(p)} & , q > c_0 m \\ \frac{\|G_j\|_2}{\sigma^2_{\min}(G_j)} \ll \frac{mn}{s^2\log(p)\log^3(n)} & , m > c_0 q,\ p = q \\ \frac{\|G_j\|_2}{\sigma^2_{\min}(G_j)} \ll \frac{n}{s\log(p)\log^3(n)} \wedge \frac{n\log(n)}{s^2m\log(p)} & , m > c_0 q,\ p > q \end{cases}$$

3. Under Condition 3:

$$\begin{cases} \frac{\|G_j\|_2}{\sigma_{\min}(G_j)} \ll \frac{n}{sm\log^5(n)\log(p)} \wedge \frac{n}{s^2m\log^3(n)\log^2(p)} & , q > c_0 m \\ \frac{\|G_j\|_2}{\sigma_{\min}(G_j)} \ll \frac{n}{s^2m\log(p)\log^4(n)} & , m > c_0 q,\ p = q \\ \frac{\|G_j\|_2}{\sigma_{\min}(G_j)} \ll \frac{n}{sm^2\log(p)\log^4(n)} \wedge \frac{n}{s^2m^3\log(p)} & , m > c_0 q,\ p > q \end{cases}$$

Theorem B.3.1. *Under Assumption 8, Assumption 10, Assumption 11 and Assumption 12, with probability at least* $1 - c_1\exp\{-cn\} - c_2\exp\{-c\log(n)\} - c_3\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - c_4\exp\left\{-cn\,(m/q)^{\mathbf{1}\{q>c_0m\}}\right\} - c_5\exp\{-c\log(p)\} - c_6\exp\{-cmn\}$ *we have*

$$\frac{\hat{V}_j}{V_j} = 1 + o_p(1).$$

B.3.1 Proof of Theorem B.3.1

Proof. According to the definitions of $\hat{V}_j$ and V_j, we can write:

$$\begin{aligned}
\frac{\hat{V}_j}{V_j} &= \frac{\sum_{i=1}^n \left|(\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2}\left(y^i - X^i\hat{\beta}\right)\right|^2}{\hat{w}_j^\top \Sigma_b^{-1/2}\Sigma_{\theta^*}\Sigma_b^{-1/2}\hat{w}_j} \\
&= \frac{\sum_{i=1}^n \left|(\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2}\left(y^i - X^i\beta^*\right)\right|^2}{\hat{w}_j^\top \Sigma_b^{-1/2}\Sigma_{\theta^*}\Sigma_b^{-1/2}\hat{w}_j} && \text{(B.32)} \\
&\quad + \frac{\sum_{i=1}^n \left|(\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2} X^i\left(\hat{\beta} - \beta^*\right)\right|^2}{\hat{w}_j^\top \Sigma_b^{-1/2}\Sigma_{\theta^*}\Sigma_b^{-1/2}\hat{w}_j} && \text{(B.33)} \\
&\quad + 2\frac{\sum_{i=1}^n (\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2}\left(y^i - X^i\beta^*\right) \times (\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2} X^i\left(\beta^* - \hat{\beta}\right)}{\hat{w}_j^\top \Sigma_b^{-1/2}\Sigma_{\theta^*}\Sigma_b^{-1/2}\hat{w}_j}. && \text{(B.34)}
\end{aligned}$$

Denote $\xi_i = (\Sigma_{\theta^*}^i)^{1/2}(\Sigma_b^i)^{-1/2}\hat{w}_j^i$, $\delta_i = (\Sigma_a^i)^{1/2}(\Sigma_b^i)^{-1/2}\hat{w}_j^i$, $\hat{u} = \hat{\beta} - \beta^*$, and $\tilde{\epsilon}_i = (\Sigma_{\theta^*}^i)^{-1/2}(y^i - X^i\beta^*)$.

1. For term (B.32):

$$\frac{\sum_{i=1}^n \left|(\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2}(y^i - X^i\beta^*)\right|^2}{\hat{w}_j^\top \Sigma_b^{-1/2}\Sigma_{\theta^*}\Sigma_b^{-1/2}\hat{w}_j} = \frac{\sum_{i=1}^n (\xi_i^\top \tilde{\epsilon}_i)^2}{\sum_{i=1}^n \|\xi_i\|_2^2}.$$

Denote $E_1 := \frac{\sum_{i=1}^n (\xi_i^\top \tilde{\epsilon}_i)^2}{\sum_{i=1}^n \|\xi_i\|_2^2}$. Then conditioning on X^i's, we have

$$\sum_{i=1}^n (\xi_i^\top \tilde{\epsilon}_i)^2 - \sum_{i=1}^n \|\xi_i\|_2^2 \in \mathrm{SE}\left(c_1 \sum_{i=1}^n \|\xi_i\|_2^4, c_2 \max_i \|\xi_i\|_2^2\right).$$

By the tail bound of sub-exponential random variables, we have

$$\mathbb{P}(|E_1 - 1| > t \mid X) \le 2\exp\left\{-c_1 \min\left(\frac{t^2\left(\sum_{i=1}^n \|\xi_i\|_2^2\right)^2}{\sum_{i=1}^n \|\xi_i\|_2^4}, \frac{t\sum_{i=1}^n \|\xi_i\|_2^2}{\max_i \|\xi_i\|_2^2}\right)\right\},$$

where by Lemma B.2.3.7 and under the conditions of Theorem B.2.2,

$$\frac{\sum_{i=1}^n \|\xi_i\|_2^4}{\left(\sum_{i=1}^n \|\xi_i\|_2^2\right)^2} \leq \frac{\max_i \|\xi_i\|_2^2}{\sum_{i=1}^n \|\xi_i\|_2^2} = o_p\left(\log^{-2}(n)\right).$$

Thus, $|E_1 - 1| < \log^{-1/2}(n)$ with probability at least $1 - c_1 \exp\{-cmnq^{-\mathbf{1}\{q>c_0 m\}}\} - c_2 \exp\{-c\log(n)\} - c_3 \exp\{-cn\} - c_4 \exp\{-c\log(p)\} - c_5 \exp\{-cn(m/q)^{\mathbf{1}\{q>c_0 m\}}\} - c_6 \exp\{-cmn\}$.

2. For term (B.33): Recall that $\Sigma_a = \operatorname{diag}\left(\{\Sigma_a^i\}_{i=1}^n\right)$ and X is obtained by vertically stacking the X^i's. We can write

$$\begin{aligned}
&\frac{\sum_{i=1}^n \left|(\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2} X^i (\hat{\beta} - \beta^*)\right|^2}{\hat{w}_j^\top \Sigma_b^{-1/2} \Sigma_{\theta^*} \Sigma_b^{-1/2} \hat{w}_j} \\
&= \frac{\sum_{i=1}^n \left|(\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2} (\Sigma_a^i)^{1/2} (\Sigma_a^i)^{-1/2} X^i (\hat{\beta} - \beta^*)\right|^2}{\hat{w}_j^\top \Sigma_b^{-1/2} \Sigma_{\theta^*} \Sigma_b^{-1/2} \hat{w}_j} \\
&\leq \frac{\max_i \left|(\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2} (\Sigma_a^i)^{1/2} (\Sigma_b^i)^{-1/2} \hat{w}_j^i\right| \times \left\|\Sigma_a^{-1/2} X (\hat{\beta} - \beta^*)\right\|_2^2}{\hat{w}_j^\top \Sigma_b^{-1/2} \Sigma_{\theta^*} \Sigma_b^{-1/2} \hat{w}_j} \\
&\leq \frac{\max_i \left\|\hat{w}_j^i\right\|_2^2 \sigma_{\max}\left((\Sigma_b^i)^{-1} \Sigma_a^i\right) \left\|\Sigma_a^{-1/2} X (\hat{\beta} - \beta^*)\right\|_2^2}{\hat{w}_j^\top \Sigma_b^{-1/2} \Sigma_{\theta^*} \Sigma_b^{-1/2} \hat{w}_j}.
\end{aligned}$$

Using (B.27) and Lemma B.1.2.7, we can bound $\sigma_{\max}\left((\Sigma_b^i)^{-1}\Sigma_a^i\right)$ by

$$\begin{aligned}
\sigma_{\max}\left((\Sigma_b^i)^{-1}\Sigma_a^i\right) &= \begin{cases} \sigma_{\max}\left((\Sigma_a^i)^{-1}\Sigma_a^i\right) & , \text{ if } j > q, \\ \sigma_{\max}\left(I + aZ_j^i (Z_j^i)^\top (\Sigma_b^i)^{-1}\right) & , \text{ if } j \leq q \end{cases} \\
&\leq \begin{cases} 1 & , \text{ if } j > q, \\ 1 + (Z_j^i)^\top (\Sigma_b^i)^{-1} Z_j^i \leq 1 + c_1 m \log(n) \left(1 \wedge \frac{\log(n)}{m \vee q}\right) & , \text{ if } j \leq q \end{cases} \\
&\leq c_1 \log^2(n).
\end{aligned}$$

Then using Lemma B.2.3.6 to bound $\max_i \|\hat{w}_j^i\|_2^2$, Lemma B.2.3.7 to bound $\hat{w}_j^\top \Sigma_b^{-1/2} \Sigma_{\theta^*} \Sigma_b^{-1/2} \hat{w}_j$ and Theorem B.1.1 to bound $\|\Sigma_a^{-1/2} X(\hat{\beta} - \beta^*)\|_2^2$, under As-

sumption 12, we obtain

$$\frac{\sum_{i=1}^n \left|(\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2} X^i (\hat{\beta} - \beta^*)\right|^2}{\hat{w}_j^\top \Sigma_b^{-1/2} \Sigma_{\theta^*} \Sigma_b^{-1/2} \hat{w}_j} = o_p(1).$$

3. For term (B.34):

$$\begin{aligned}
&\frac{\sum_{i=1}^n (\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2} (y^i - X^i\beta^*) \times (\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2} X^i (\beta^* - \hat{\beta})}{\hat{w}_j^\top \Sigma_b^{-1/2} \Sigma_{\theta^*} \Sigma_b^{-1/2} \hat{w}_j} \\
&\leq \frac{\|\hat{u}\|_2^2}{\sum_{i=1}^n \|\xi_i\|_2^2} \left\| \sum_{i=1}^n \xi_i^\top \tilde{\epsilon}_i (X^i)^\top (\Sigma_b^i)^{-1/2} \hat{w}_j^i \right\|_\infty \\
&\leq \frac{\|\hat{u}\|_2^2}{\sum_{i=1}^n \|\xi_i\|_2^2} \max_{1 \leq l \leq p} \left| \sum_{i=1}^n (X_l^i)^\top (\Sigma_b^i)^{-1/2} \hat{w}_j^i \xi_i^\top \tilde{\epsilon}_i \right|. \qquad \text{(B.35)}
\end{aligned}$$

Conditioning on X^i's, $\tilde{\epsilon}_i \in \mathrm{SGV}(c_1)$, and we use the tail bound of sub-Gaussian random variables to get

$$\mathbb{P}\left(\max_{1 \leq l \leq p} \left| \sum_{i=1}^n (X_l^i)^\top (\Sigma_b^i)^{-1/2} \hat{w}_j^i \xi_i^\top \tilde{\epsilon}_i \right| > t \mid X \right) \leq 2 \exp\left\{ \log(p) - c_1 \frac{t^2}{\max_l \sum_{i=1}^n \|(X_l^i)^\top (\Sigma_b^i)^{-1/2} \hat{w}_j^i \xi_i\|_2^2} \right\}$$

Then, taking $t = c_2 \sqrt{\log(p)} \sqrt{\max_l \sum_{i=1}^n \|(X_l^i)^\top (\Sigma_b^i)^{-1/2} \hat{w}_j^i \xi_i\|_2^2}$ and plugging the bound into (B.35), with probability at least $1 - 2\exp\{-c\log(p)\}$, we have

$$\begin{aligned}
&\frac{\sum_{i=1}^n (\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2} (y^i - X^i\beta^*) \times (\hat{w}_j^i)^\top (\Sigma_b^i)^{-1/2} X^i (\beta^* - \hat{\beta})}{\hat{w}_j^\top \Sigma_b^{-1/2} \Sigma_{\theta^*} \Sigma_b^{-1/2} \hat{w}_j} \\
&\leq c_3 \frac{\|\hat{u}\|_2^2}{\sum_{i=1}^n \|\xi_i\|_2^2} \sqrt{\log(p)} \sqrt{\max_l \sum_{i=1}^n \|(X_l^i)^\top (\Sigma_b^i)^{-1/2} \hat{w}_j^i \xi_i\|_2^2} \\
&\leq c_3 \frac{\|\hat{u}\|_2^2}{\sum_{i=1}^n \|\xi_i\|_2^2} \sqrt{\log(p)} \sqrt{\max_l \sum_{i=1}^n \|\hat{w}_j^i\|_2^2 \|\xi_i\|_2^2 \|(\Sigma_b^i)^{-1/2} X_l^i\|_2^2} \\
&\leq c_3 \frac{\|\hat{u}\|_2^2}{\sum_{i=1}^n \|\xi_i\|_2^2} \sqrt{\log(p)} \max_i \|\hat{w}_j^i\|_2 \max_l \|\xi_i\|_2 \sqrt{\sigma_{\max}(X^\top \Sigma_b^{-1} X)}.
\end{aligned}$$

Then, under the conditions of Theorem B.2.2 and Assumption 12, we use Theorem

B.1.1 to bound $\|\hat{u}\|_2^2$, use Lemma B.1.3.1 to bound $\sigma_{\max}(X^\top \Sigma_b^{-1} X)$, use Lemma B.2.3.7 to bound $\sum_{i=1}^n \|\xi_i\|_2^2$, $\max_l \|\xi_i\|_2$ and Lemma B.2.3.6 to bound $\max_i \|\hat{w}_j^i\|_2$. Then it follows that (B.34) is $o_p(1)$ with probability at least $1 - c_1 \exp\{-cmnq^{-\mathbf{1}\{q>c_0 m\}}\} - c_2 \exp\{-c\log(n)\} - c_3 \exp\{-cn\} - c_4 \exp\{-c\log(p)\} - c_5 \exp\{-cn(m/q)^{\mathbf{1}\{q>c_0 m\}}\} - c_6 \exp\{-cmn\}$.

□

B.4 Variance Component Estimator $\hat{\theta}$

We propose a consistent variance component estimator $\hat{\theta}$ for the more general doubly high-dimensional LMM (B.1) with $\Psi = \mathrm{diag}(\psi)$ and $R^i = I_m$. To simplify the discussion, we assume the vector ψ satisfies Assumption 9.1. The estimator $\hat{\theta} = (\hat{\phi}, \hat{\sigma}_e^2)$ is defined in (3.7) and (3.8) in the main text. We introduce some notations to facilitate the rest of the discussion in this section. For $l = 1, \ldots, q$, $i = 1, \ldots, n$, we define

$$\begin{aligned} A_l^i &= Z_l^i (Z_l^i)^\top - \left(\mathrm{diag}(Z_l^i)\right)^2, \\ r_i &= y_i - X^i \beta^* = Z^i \gamma_i + \epsilon_i, \end{aligned} \tag{B.36}$$

and define the $q \times q$ matrix B such that

$$B_{j,k} = \sum_{i \in S_2} \mathrm{tr}\left(A_j^i A_k^i\right), \quad \text{for } j, k = 1, \ldots, q. \tag{B.37}$$

We use $A \circ B$ to denote the Hadamard product of the matrices A and B. We define

$$\begin{aligned} s_Z^i &= \max_j \sum_{k=1, k\neq j}^{q} \mathbf{1}\left\{(\Sigma_Z^i)_{j,k} \gg \log(nq^2)/\sqrt{m}\right\}, \\ s_Z &= \max_i s_Z^i, \end{aligned} \tag{B.38}$$

which represent the maximum number of entries that are not too small in each row of Σ_Z^i. When Σ_Z^i is a sparse matrix, s_Z^i is small.

We first state the assumptions under which the consistency of $\hat{\theta}$ holds.

Assumption 13. *1. The vectors γ_i and ϵ_i are normally distributed with mean zero and variance matrices Ψ, $\sigma_e^2 I_m$, respectively. The covariance matrix Ψ satisfies Assumption 9.1.*

2. The following conditions hold:

$$\sqrt{m} \gg \log(nq)$$
$$nm \gg \max\left\{q^{3/2}\log(q)\log^2(n), \log(q)\log(n)\left(s_Z\sqrt{m}\log(nq^2) + q\log^2(nq^2)\right)\right\}$$
$$nm^3 \gg q^2\log(q)\log^2(n)$$

3. $\sqrt{n}m \gg s s_\psi (q/m)^{\mathbf{1}\{q>c_0 m\}} m^{\mathbf{1}\{p>q, m>c_0 q\}} \log(q)\log(p)\log(n)\log(nmq)$.

We are now ready the state the theoretical properties of $\hat{\theta} = (\hat{\psi}, \hat{\sigma}_e^2)$:

Theorem B.4.1. *Under Assumption 8 and Assumption 13.1–13.2, with probability at least* $1 - c_1\exp\{-c\log(nq)\} - c_2\exp\{-cn\} - c_3\exp\{-c\log(n)\} - c_4\exp\{-cmnq^{-\mathbf{1}\{q>c_0 m\}}\} - c_5\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0 m\}}\} - c_6\exp\{-c\log(q)\}$, *we have*

$$\|\hat{\psi} - \psi^*\|_\delta \le s_\psi^{1/\delta}\frac{\log(n)\log(p)\log(nmq)}{\sqrt{n}} m^{\mathbf{1}\{m>c_0 q, p>q\}}(q/m)^{\mathbf{1}\{q>c_0 m\}}, \quad \delta = 1, 2.$$

Theorem B.4.2. *Under Assumption 8 and Assumption 13, we have* $|\hat{\sigma}_e^2 - \sigma_e^{2,*}| = o_p(1)$ *with probability at least*

$$1 - c_1\frac{s_\psi(m\vee q)\log^3(n)}{nm} - c_2\exp\{-cnm\} - c_3\exp\{-cn\} - c_4\exp\{-c\log(n)\} - c_5\exp\{-cmnq^{-\mathbf{1}\{q>c_0 m\}}\}$$
$$- c_6\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0 m\}}\} - c_7\exp\left\{-c\frac{n^2}{s\log^4(n)\log(p)}\left(\frac{m^2}{q^2}\right)^{\mathbf{1}\{q>c_0 m\}} m^{-\mathbf{1}\{m>c_0 q, p>q\}}\right\}$$

B.4.1 Related lemmas for Theorem B.4.1 and Theorem B.4.2

Lemma B.4.3. *1. For $1 \le j \le q$, define*

$$E_{1,j} = \sum_{i\in S_2}\mathrm{tr}\left(A_j^i\left(r_i r_i^\top - \sum_{k=1}^q A_k^i\psi_k^*\right)\right).$$

Under Assumption 8 and Assumption 13.1, with probability at least $1 - 2\exp\{-c(m \vee q)\log(n)\} - 2\exp\{-cm\log(nq)\} - 2\exp\{-c\log(nmq)\} - 2\exp\{-c\log(q)\}$*, we have*

$$\max_j |E_{1,j}| \leq c_1(m \vee q)\log(n)\log(nmq)\log(q)m\sqrt{n}.$$

2. *Define*

$$E_{2,j} = \sum_{i \in S_2} \operatorname{tr}\left(A_j^i X^i (\beta^* - \hat{\beta})(\beta^* - \hat{\beta})^\top (X^i)^\top\right).$$

Under Assumption 8 and Assumption 13.1, with probability at least $1 - 4\exp\{-cn\} - 12\exp\{-c\log(n)\} - 2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - \exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\} - 2\exp\{-cm\log(nq)\} - 2\exp\{-c\log(nmq)\} - 2\exp\{-c(m \vee q)\log(n)\}$*, we have*

$$\max_j |E_{2,j}| \leq \begin{cases} c_2 s\log(p)\log^3(n)\log(nq)m^2 & , q > c_0 m \\ c_2 s\log(p)\log(n)\log(nq)m^2 & , m > c_0 q,\ p = q \\ c_2 s\log(p)\log(n)\log(nq)m^3 & , m > c_0 q,\ p > q, \end{cases}$$

3. *Define*

$$E_{3,j} = \sum_{i \in S_2} \operatorname{tr}\left(A_j^i r_i (\beta^* - \hat{\beta})^\top (X^i)^\top\right).$$

Under Assumption 8 and Assumption 13.1, with probability at least $1 - 4\exp\{-cn\} - 12\exp\{-c\log(n)\} - 2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - \exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\} - 2\exp\{-cm\log(nq)\} - 2\exp\{-c\log(nmq)\} - 2\exp\{-c\log(q)\} - 2\exp\{-c(m \vee q)\log(n)\}$*, we have*

$$\max_j |E_{3,j}| \leq \begin{cases} c_3\sqrt{s\log(p)\log(q)\log^4(n)\log^2(nq)m^3q} & , q > c_0 m \\ c_3\sqrt{s\log(p)\log(q)\log^2(n)\log^2(nq)m^4} & , m > c_0 q,\ p = q \\ c_3\sqrt{s\log(p)\log(q)\log^2(n)\log^2(nq)m^5} & , m > c_0 q,\ p > q. \end{cases}$$

Lemma B.4.4. *Define* $G_1^i = \left((Z^i)^\top Z^i\right) \circ \left((Z^i)^\top Z^i\right)$ *and* $G_2^i = \left(Z^i \circ Z^i\right)^\top \left(Z^i \circ Z^i\right)$.

1. *Under Assumption 8, Assumption 13.1 and Assumption 13.2, with probability at least* $1 - 2\exp\{-c\log(nq)\} - 2\exp\{-c\log(nq^2)\}$, *we have* $\max_i \left\|G_1^i - \mathbb{E}\left(G_1^i\right)\right\|_2 \leq c_1 m^{3/2}\log(nq^2) + c_1 s_Z m^{3/2}\log(nq^2) + c_1 mq\log^2(nq^2)$.

2. *Under Assumption 8 and Assumption 13.1, we have* $\max_i \|G_2^i - \mathbb{E}(G_2^i)\|_2^2 \leq c_2 mq\log(n)$ *with probability at least* $1 - \exp\{-cq\log(n)\}$.

Lemma B.4.5. *Under Assumption 8 and Assumption 13, with probability at least* $1 - \exp\{-c\log(q)\} - 2\exp\{-c\log(nq)\} - \exp\{-cq\log(n)\}$, *for any* $v \in \mathbb{R}^q$ *we have* $v^\top Bv \geq c_1 nm^2\|v\|_2^2$, *where* B *is defined in* (B.37).

B.4.2 Proof of Theorem B.4.1

Proof of Theorem B.4.1.

We first show the consistency of $\hat{\psi}$.

Based on the definition of $\hat{\psi}$ in (3.7) in the main text and the definition of A_l^i in (B.36), we can write

$$\begin{aligned}\hat{\psi} &= \arg\min_{\psi} \sum_{i\in S_2} \left\| \hat{r}_i\hat{r}_i^\top - (\mathrm{diag}(\hat{r}_i))^2 - \sum_{l=1}^{q} \psi_l A_l^i \right\|_F^2 + \lambda_\theta\|\psi\|_1 \\ &= \arg\min_{\psi} \sum_{i\in S_2} \mathrm{tr}\left(\left(\hat{r}_i\hat{r}_i^\top - (\mathrm{diag}(\hat{r}_i))^2 - \sum_{l=1}^{q} \psi_l A_l^i \right)^2 \right) + \lambda_\theta\|\psi\|_1.\end{aligned}$$

Rearranging the terms, we can write

$$\sum_{i\in S_2}\operatorname{tr}\left(\left(\hat{r}_i\hat{r}_i^\top-(\operatorname{diag}(\hat{r}_i))^2-\sum_{l=1}^{q}\hat{\psi}_l A_l^i\right)^2\right)+\lambda_\theta\|\hat{\psi}\|_1\le\sum_{i\in S_2}\operatorname{tr}\left(\left(\hat{r}_i\hat{r}_i^\top-(\operatorname{diag}(\hat{r}_i))^2-\sum_{l=1}^{q}\psi_l^* A_l^i\right)^2\right)+$$

$$\Rightarrow\sum_{i\in S_2}\operatorname{tr}\left(\sum_{l_1=1}^{q}\sum_{l_2=1}^{q}\hat{\psi}_{l_1}\hat{\psi}_{l_2}A_{l_1}^iA_{l_2}^i-2\sum_{l=1}^{q}\hat{\psi}_l A_l^i\left(\hat{r}_i\hat{r}_i^\top-(\operatorname{diag}(\hat{r}_i))^2\right)\right)+\lambda_\theta\|\hat{\psi}\|_1$$

$$\le\sum_{i\in S_2}\operatorname{tr}\left(\sum_{l_1=1}^{q}\sum_{l_2=1}^{q}\psi_{l_1}^*\psi_{l_2}^*A_{l_1}^iA_{l_2}^i-2\sum_{l=1}^{q}\psi_l^* A_l^i\left(\hat{r}_i\hat{r}_i^\top-(\operatorname{diag}(\hat{r}_i))^2\right)\right)+\lambda_\theta\|\psi^*\|_1. \tag{B.39}$$

Recall the definition of B in (B.37). Define the length q vector δ with $\delta_k=\sum_{i\in S_2}\operatorname{tr}\left(A_k^i\left(\hat{r}_i\hat{r}_i^\top-(\operatorname{diag}(\hat{r}_i))^2\right)\right)$ for $k=1,\ldots,q$. Then the above inequality (B.39) implies:

$$\begin{aligned}(\hat{\psi}-\psi^*)^\top B(\hat{\psi}-\psi^*)&\le 2(\hat{\psi}-\psi^*)(\delta-B\psi^*)+\lambda_\theta\|\psi^*\|_1-\lambda_\theta\|\hat{\psi}\|_1\\&\le 2\|\hat{\psi}-\psi^*\|_1\|\delta-B\psi^*\|_\infty+\lambda_\theta\|\psi^*\|_1-\lambda_\theta\|\hat{\psi}\|_1.\end{aligned}$$

We next show that $\|\delta-B\psi^*\|_\infty$ is upper bounded by $(\hat{\psi}-\psi^*)^\top B(\hat{\psi}-\psi^*)$, such that we can follow similar arguments as those in the proof of Theorem B.1.1 to show the consistency of $\hat{\psi}$.

For the term $\|\delta - B\psi^*\|_\infty$, we can write

$$
\begin{aligned}
|\left(\delta - B\psi^*\right)_j| &= \left| \sum_{i \in S_2} \operatorname{tr}\left(A_j^i \left(\hat{r}_i \hat{r}_i^\top - (\operatorname{diag}(\hat{r}_i))^2 - \sum_{l=1}^{q} A_l^i \psi_l^* \right) \right) \right| \\
&= \Bigg| \sum_{i \in S_2} \operatorname{tr}\left(A_j^i \left(r_i r_i^\top - (\operatorname{diag}(r_i))^2 - \sum_{l=1}^{q} A_l^i \psi_l^* \right) \right) \\
&\quad + \sum_{i \in S_2} \operatorname{tr}\left(A_j^i \left[X^i(\beta^* - \hat{\beta})(\beta^* - \hat{\beta})^\top (X^i)^\top - \left(\operatorname{diag}\left(X^i(\beta^* - \hat{\beta}) \right) \right)^2 \right] \right) \\
&\quad + 2 \sum_{i \in S_2} \operatorname{tr}\left(A_j^i \left[r_i(\beta^* - \hat{\beta})^\top (X^i)^\top - \operatorname{diag}\left(r_i(\beta^* - \hat{\beta})^\top (X^i)^\top \right) \right] \right) \Bigg| \\
&= \Bigg| \sum_{i \in S_2} \operatorname{tr}\left(A_j^i \left(r_i r_i^\top - \sum_{l=1}^{q} A_l^i \psi_l^* \right) \right) && \text{(B.40)} \\
&\quad + \sum_{i \in S_2} \operatorname{tr}\left(A_j^i X^i (\beta^* - \hat{\beta})(\beta^* - \hat{\beta})^\top (X^i)^\top \right) && \text{(B.41)} \\
&\quad + 2 \sum_{i \in S_2} \operatorname{tr}\left(A_j^i r_i (\beta^* - \hat{\beta})^\top (X^i)^\top \right) \Bigg|, && \text{(B.42)}
\end{aligned}
$$

where the last equality holds because $\operatorname{tr}\left(A_j^i \operatorname{diag}(u) \right) = 0$ for any vector $u \in \mathbb{R}^m$. Then using Lemma B.4.3, we can bound the terms (B.40), (B.41) and (B.42) to get

$$
|\left(\delta - B\psi^*\right)_j| \leq c_1 \log(n) \log(p) \log(nmq) m^{1+\mathbf{1}\{m > c_0 q, p > q\}} (m \vee q) \sqrt{n} \leq \lambda_\theta / 4
$$

with probability at least $1 - 4\exp\{-cn\} - 12\exp\{-c\log(n)\} - 2\exp\{-cmnq^{-\mathbf{1}\{q > c_0 m\}}\} - \exp\{-cn(m/q)^{\mathbf{1}\{q > c_0 m\}}\} - 2\exp\{-cm\log(nq)\} - 2\exp\{-c\log(nmq)\} - 2\exp\{-c\log(q)\} - 2\exp\{-c(m \vee q)\log(n)\}$.

On the other hand, using Lemma B.4.5 we can show that with probability at least $1 - \exp\{-c\log(q)\} - 2\exp\{-c\log(nq)\} - \exp\{-cq\log(n)\}$, we have

$$
(\hat{\psi} - \psi^*)^\top B(\hat{\psi} - \psi^*) \geq c_2 n m^2 \|\hat{\psi} - \psi^*\|_2^2.
$$

Next, letting $v = \hat{\psi} - \psi^*$, we follow the similar arguments as those in the proof of Theorem B.1.1 to show that

$$c_2 nm^2 \|v\|_2^2 \leq \frac{3}{2}\lambda_\theta \|v_{S_\psi}\|_1 \leq \frac{3}{2}\lambda_\theta \sqrt{s_\psi}\|v\|_2.$$

Thus, with probability at least $1 - c_3\exp\{-c\log(nq)\} - c_4\exp\{-cn\} - c_5\exp\{-c\log(n)\} - c_6\exp\{-cmnq^{-\mathbf{1}\{q>c_0 m\}}\} - c_7\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0 m\}}\} - c_8\exp\{-c\log(q)\}$ we have

$$\|\hat{\psi} - \psi^*\|_1 \leq s_\psi \frac{\log(n)\log(p)\log(nmq)}{\sqrt{n}} m^{\mathbf{1}\{m>c_0 q, p>q\}}(q/m)^{\mathbf{1}\{q>c_0 m\}}$$
$$\|\hat{\psi} - \psi^*\|_2 \leq \sqrt{s_\psi} \frac{\log(n)\log(p)\log(nmq)}{\sqrt{n}} m^{\mathbf{1}\{m>c_0 q, p>q\}}(q/m)^{\mathbf{1}\{q>c_0 m\}}.$$

□

Proof of Theorem B.4.2. We can write the proposed estimator $\hat{\sigma}_e^2$ as:

$$\hat{\sigma}_e^2 = \sigma_e^{2,*} + \frac{1}{n_3 m}\sum_{i\in S_3}\left(\mathrm{tr}(\hat{r}_i\hat{r}_i^\top - \Sigma_{\theta^*}^i) - \mathrm{tr}(Z^i\,\mathrm{diag}(\hat{\psi})(Z^i)^\top - Z^i\,\mathrm{diag}(\psi)(Z^i)^\top)\right).$$

We can then show that

$$\begin{aligned} n_3 m\left|\hat{\sigma}_e^2 - \sigma_e^{2,*}\right| &= \sum_{i\in S_3}\|r_i\|_2^2 + \|X^i(\beta^* - \hat{\beta})\|_2^2 + 2(\beta^* - \hat{\beta})^\top (X^i)^\top r_i - \mathrm{tr}(\Sigma_{\theta^*}^i) - \mathrm{tr}\left(Z^i\,\mathrm{diag}(\hat{\psi} - \psi^*)(Z^i)^\top\right) \\ &\leq \sum_{i\in S_3}\|r_i\|_2^2 - \mathrm{tr}(\Sigma_{\theta^*}^i) && \text{(B.43)} \\ &\quad + \|X(\beta^* - \hat{\beta})\|_2^2 && \text{(B.44)} \\ &\quad + 2\sum_{i\in S_3}(\beta^* - \hat{\beta})^\top (X^i)^\top r_i && \text{(B.45)} \\ &\quad + \sum_{i\in S_3}\mathrm{tr}\left(Z^i\,\mathrm{diag}(\hat{\psi} - \psi^*)(Z^i)^\top\right). && \text{(B.46)} \end{aligned}$$

For the term (B.43): Conditioning on the design matrices Z^i, we have $r_i \mid Z^i \sim N(0, \Sigma_\theta^i)$ independently for $i = 1, \ldots, n$, and $\mathbb{E}\left(\|r_i\|_2 \mid Z^i\right) = \mathrm{tr}(\Sigma_{\theta^*}^i)$. Denote the (j,k) entry of $\Sigma_{\theta^*}^i$

by $w_{i,j,k}$. By Markov's inequality we can show that

$$\begin{aligned}
&\mathbb{P}\left(\left|\sum_{i\in_3}\|r_i\|_2^2-\operatorname{tr}(\Sigma^i_{\theta^*})\right|\geq t\mid Z^i\right)\leq\frac{\sum_{i\in S_3}\operatorname{Var}\left(\|r_i\|_2^2\right)}{t^2}\\
&=\frac{\sum_{i\in S_3}\left(\sum_{j=1}^m 3w_{i,j,j}^2+\sum_{k=1}^m\sum_{j=1,j\neq k}^m(w_{i,j,j}w_{i,k,k}+2w_{i,j,k}^2)-(\sum_{j=1}^m w_{i,j,j})^2\right)}{t^2}\\
&=\frac{\sum_{i\in S_3}\sum_{k=1}^m\sum_{j=1}^m 2w_{i,j,k}^2}{t^2}=\frac{\sum_{i\in S_3}\|\Sigma^i_{\theta^*}\|_F^2}{t^2}.
\end{aligned}\tag{B.47}$$

Note that we have $\Psi=\operatorname{diag}(\psi)$ with $\|\psi\|_0=s_\psi<c_1m\wedge n$ (Assumption 9.1), and we can thus show that

$$\begin{aligned}
\sum_{i\in S_3}\|\Sigma^i_{\theta^*}\|_F^2&\leq\sum_{i\in S_3}\operatorname{tr}\left(Z^i\operatorname{diag}(\psi)(Z^i)^\top+\sigma_e^2I_m\right)\sigma_{\max}\left(\Sigma^i_{\theta^*}\right)\\
&\leq c_1\sum_{i\in S_3}\left(\sum_{l\in S_\psi}\psi_l\|Z_l^i\|_2^2+m\sigma_e^2\right)\left(\sigma^2_{\max}\left(Z^i\right)+1\right)\\
&\leq c_2\left(nm+s_\psi\max_{l\in S_\psi}\sum_{i\in S_3}\|Z_l^i\|_2^2\right)\sigma^2_{\max}(Z^i).
\end{aligned}\tag{B.48}$$

Using Lemma B.1.2.4, we have $\sigma^2_{\max}(Z^i)\leq c_3(m\vee q)\log(n)$ with probability at least $1-2\exp\{-c\log(n)\}$. We can also use $Z_l^i\mid\Sigma_Z^i\sim N\left(0,\left(\Sigma_Z^i\right)_{l,l}I_m\right)$ to show that

$$\begin{aligned}
&\mathbb{P}\left(\max_{l\in S_\psi}\left|\sum_{i\in S_3}\|Z_l^i\|_2^2-m\left(\Sigma_Z^i\right)_{l,l}\right|\geq t\mid\Sigma_Z^i\right)\leq 2\exp\left\{\log(s_\psi)-c_4\min\left\{\frac{t^2}{4nm},\frac{t}{4}\right\}\right\}\\
&\Rightarrow\mathbb{P}\left(\max_{l\in S_\psi}\sum_{i\in S_3}\|Z_l^i\|_2^2\leq c_5nm\right)\geq 1-2\exp\left\{-cnm\right\}.
\end{aligned}\tag{B.49}$$

Plugging in the bounds into (B.48), we obtain $\sum_{i\in S_3}\|\Sigma^i_{\theta^*}\|_F^2\leq c_6s_\psi nm(m\vee q)\log(n)$. Then plugging into (B.47) and taking $t=c_7nm/\log(n)$, we obtain (B.43) $\leq nm/\log(n)$ with probability at least $1-c\frac{s_\psi(m\vee q)\log^3(n)}{nm}-2\exp\{-c\log(n)\}-2\exp\{-cnm\}$.

For the term (B.44): Using Theorem B.1.1 and Lemma B.1.2.4, we have with probability at least $1-c_1\exp\{-cn\}-c_2\exp\{-c\log(n)\}-c_3\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}-$

$c_4 \exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\}$ that

$$\|X(\beta^* - \hat{\beta})\|_2^2 \leq \|\Sigma_a^{-1/2} X(\beta^* - \hat{\beta})\|_2^2 \sigma_{\max}(\Sigma_a)$$
$$\leq c_5 \begin{cases} sq\log(p)\log(n) & , q > c_0 m \\ sm\log^3(n)\log(p) & , q > c_0 m \text{ and assuming Assumption 9.1} \\ sm\log(p)\log(n) & , m > c_0 q,\ p = q \\ sm^2\log(p)\log(n) & , m > c_0 q,\ p > q. \end{cases}$$

For the term (B.45): Conditioning on $\hat{\beta}$ and X and using the tail bound of normally distributed random variables, we can write

$$\mathbb{P}\left(\left|\sum_{i\in S_3} (\beta^* - \hat{\beta})^\top (X^i)^\top r_i\right| \geq t | X, \hat{\beta}\right) \leq 2\exp\left\{-\frac{t^2}{8\sum_{i\in S_3}(\beta^* - \hat{\beta})^\top (X^i)^\top \Sigma_{\theta^*}^i X^i (\beta^* - \hat{\beta})}\right\}$$
$$\leq 2\exp\left\{-\frac{t^2}{8\max_i \sigma_{\max}(\Sigma_{\theta^*}^i)\sigma_{\max}(\Sigma_a^i)\|\Sigma_a^{-1/2} X(\beta^* - \hat{\beta})\|_2^2}\right\}. \tag{B.50}$$

Using Theorem B.1.1 to bound $\|\Sigma_a^{-1/2} X(\beta^* - \hat{\beta})\|_2^2$ and Lemma B.1.2.4 to bound $\sigma_{\max}(\Sigma_{\theta^*}^i)\sigma_{\max}(\Sigma_a^i)$, we have with probability at least $1 - c_1\exp\{-cn\} - c_2\exp\{-c\log(n)\} - c_3\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - c_4\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\}$,

$$\max_i \sigma_{\max}(\Sigma_{\theta^*}^i)\sigma_{\max}(\Sigma_a^i)\|\Sigma_a^{-1/2} X(\beta^* - \hat{\beta})\|_2^2$$
$$\leq c_5 \begin{cases} sq^2\log(p)\log^2(n) & , q > c_0 m \\ sm^2\log(p)\log^2(n) & , m > c_0 q,\ p = q \\ sm^3\log(p)\log^2(n) & , m > c_0 q,\ p > q. \end{cases}$$

Plugging in the bound into (B.50) and taking $t = c_6 nm/\log(n)$, we have (B.45) $\leq c_6 nm/\log(n)$ with probability at least $1 - c_1\exp\{-cn\} - c_2\exp\{-c\log(n)\} - c_3\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - c_4\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\} -$

$2\exp\{-c\frac{n^2}{s\log^4(n)\log(p)}\left(\frac{m^2}{q^2}\right)^{\mathbf{1}\{q>c_0m\}}m^{-\mathbf{1}\{m>c_0q,p>q\}}\}$

For the term (B.46): We can write

$$\sum_{i\in S_3}\mathrm{tr}\left(Z^i\,\mathrm{diag}(\hat{\psi}-\psi^*)(Z^i)^\top\right)=\sum_{i\in S_3}\sum_{l=1}^{q}(\hat{\psi}_l-\psi_l^*)\|Z_l^i\|_2^2$$
$$\leq\|\psi-\psi^*\|_1\max_{l=1,\ldots,q}\sum_{i\in S_3}\|Z_l^i\|_2^2.$$

Using similar arguments as those in (B.49), we can show $\max_{l=1,\ldots,q}\sum_{i\in S_3}\|Z_l^i\|_2^2\leq c_1\log(q)nm$ with probability at least $1-2\exp\{-cnm\log(q)\}$. Then using Theorem B.4.1, we have

$$(\text{B.46})\leq s_\psi\sqrt{n}m\log(q)\log(n)\log(p)\log(nmq)m^{\mathbf{1}\{m>c_0q,p>q\}}(q/m)^{\mathbf{1}\{q>c_0m\}}$$

with probability at least $1-c_1\exp\{-c\log(nq)\}-c_2\exp\{-cn\}-c_3\exp\{-c\log(n)\}-c_4\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}-c_5\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\}-c_6\exp\{-c\log(q)\}-c_7\exp\{-cnm\log(q)\}$.

Therefore, plugging in the above bounds for (B.43)–(B.46), under Assumption 13.3, we obtain $|\hat{\sigma}_e^2-\sigma_e^{2,*}|=o_p(1)$ with probability at least

$$1-c_1\frac{s_\psi(m\vee q)\log^3(n)}{nm}-c_2\exp\{-cnm\}-c_3\exp\{-cn\}-c_4\exp\{-c\log(n)\}-c_5\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\}$$
$$-c_6\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\}-c_7\exp\left\{-c\frac{n^2}{s\log^4(n)\log(p)}\left(\frac{m^2}{q^2}\right)^{\mathbf{1}\{q>c_0m\}}m^{-\mathbf{1}\{m>c_0q,p>q\}}\right\}.$$

□

B.4.3 Proof of related lemmas for Theorem B.4.1 and Theorem B.4.2

Proof of Lemma B.4.3.

Lemma B.4.3.1)

We can write

$$E_{1,j} = \sum_{i \in S_2} \operatorname{tr}\left(A_j^i \left(r_i r_i^\top - \sum_{k=1}^{q} A_k^i \psi_k^* \right) \right)$$
$$= \sum_{i \in S_2} r_i^\top A_j^i r_i - \sum_{i \in S_2} \operatorname{tr}\left(A_j^i \sum_{k=1}^{q} A_k^i \psi_k^* \right)$$

Conditioning on the design matrices Z^i, we have that $r_i \mid Z^i \sim N(0, \Sigma_{\theta^*}^i)$, and

$$\mathbb{E}\left(r_i^\top A_j^i r_i \mid Z^i \right) = \operatorname{tr}(A_j^i \Sigma_{\theta^*}^i)$$
$$= \operatorname{tr}\left(A_j^i \left(\sigma_e^{2,*} I_m + \sum_{k=1}^{q} \left(A_k^i + \left(\operatorname{diag}\left(Z_k^i \right) \right)^2 \right) \psi_k^* \right) \right)$$
$$= \sum_{i \in S_2} \operatorname{tr}\left(A_j^i \sum_{k=1}^{q} A_k^i \psi_k^* \right),$$

where for the last equality we again use $\operatorname{tr}\left(A_j^i \operatorname{diag}(u) \right) = 0$ for any vector $u \in R^m$. Then using the Hanson-Wright inequality [184] and denoting $W_j = \operatorname{diag}\left(\left\{ \left(\Sigma_{\theta^*}^i \right)^{1/2} A_j^i \left(\Sigma_{\theta^*}^i \right)^{1/2} \right\}_{i \in S_2} \right)$, we have that

$$\mathbb{P}\left(\max_{1 \le j \le q} |E_{1,j}| \ge t \mid Z \right) \le 2 \exp\left\{ \log(q) - c_1 \min\left\{ \frac{t^2}{\max_j \|W_j\|_F^2}, \frac{t}{\max_j \|W_j\|_2} \right\} \right\}. \quad \text{(B.51)}$$

First, for the term $\|W_j\|_F^2$, we can show that

$$
\begin{aligned}
\|W_j\|_F^2 &= \operatorname{tr}\left(W_j^2\right) \\
&= \sum_{i\in S_2} \operatorname{tr}\left(\Sigma_{\theta^*}^i A_j^i \Sigma_{\theta^*}^i A_j^i\right) \\
&= \sum_{i\in S_2} \operatorname{tr}\left(\sigma_e^{4,*} A_j^i A_j^i + 2\sigma_e^{2,*} Z^i \Psi (Z^i)^\top A_j^i A_j^i + Z^i \Psi (Z^i)^\top A_j^i Z^i \Psi (Z^i)^\top A_j^i\right) \\
&= \sum_{i\in S_2} \operatorname{tr}\Big(\sigma_\epsilon^{4,*} Z_j^i (Z_j^i)^\top Z_j^i (Z_j^i)^\top - \sigma_e^{4,*} \operatorname{diag}(Z_j^i)^2 Z_j^i (Z_j^i)^\top \\
&\qquad +2\sigma_e^{2,*} Z^i \Psi (Z^i)^\top Z_j^i (Z_j^i)^\top Z_j^i (Z_j^i)^\top \\
&\qquad +2\sigma_e^{2,*} Z^i \Psi (Z^i)^\top \operatorname{diag}(Z_j^i)^4 \\
&\qquad -4\sigma_e^{2,*} Z^i \Psi (Z^i)^\top \operatorname{diag}(Z_j^i)^2 Z_j^i (Z_j^i)^\top \\
&\qquad +Z^i \Psi (Z^i)^\top Z_j^i (Z_j^i)^\top Z^i \Psi (Z^i)^\top Z_j^i (Z_j^i)^\top \\
&\qquad +\ Z^i \Psi (Z^i)^\top \operatorname{diag}(Z_j^i)^2 Z^i \Psi (Z^i)^\top \operatorname{diag}(Z_j^i)^2 \\
&\qquad -\ 2\ Z^i \Psi (Z^i)^\top Z_j^i (Z_j^i)^\top Z^i \Psi (Z^i)^\top \operatorname{diag}(Z_j^i)^2\Big) \\
&\le \left(\max_i \sigma_{\max}^2\left(Z^i \Psi (Z^i)^\top\right) + 2\sigma_e^{2,*} \max_i \sigma_{\max}\left(Z^i \Psi (Z^i)^\top\right) + \sigma_e^{4,*}\right) \sum_{i\in S_2} \|Z_j^i\|_2^4 \\
&\quad + \left(3\max_i \sigma_{\max}^2\left(Z^i \Psi (Z^i)^\top\right) + 6\sigma_e^{2,*} \max_i \sigma_{\max}\left(Z^i \Psi (Z^i)^\top\right) + \sigma_e^{4,*}\right) \sum_{i\in S_2} \sum_{l=1}^m \left(\left(Z_j^i\right)_l\right)^4 .
\end{aligned}
\tag{B.52}
$$

Using Lemma B.1.2.4 we can show that $\max_i \sigma_{\max}\left(Z^i \Psi (Z^i)^\top\right) \le c_1 (m \vee q)\log(n)$ with probability at least $1 - 2\exp\{-c_2 (m\vee q)\log(n)\}$. Note that $\sum_{l=1}^m \left(\left(Z_j^i\right)_l\right)^4 \le \|Z_j^i\|_2^4$. Based on the distribution $Z_j^i \mid \Sigma_Z^i \sim N\left(0, \left(\Sigma_Z^i\right)_{j,j} I_m\right)$ and Assumption 8.2 that $\left(\Sigma_Z^i\right)_{j,j} \asymp 1$, we can show that

$$
\mathbb{P}\left(\max_j \max_i \left|\|Z_j^i\|_2^2 - c_3 m\right| \ge t \mid \Sigma_Z^i\right) \le 2\exp\left\{\log(q) + \log(n_2) - c_4 \min\left\{\frac{t^2}{8m}, \frac{t}{8}\right\}\right\}.
\tag{B.53}
$$

Taking $t = c_5 m\log(nq)$, we have $\max_{i,j} \|Z_j^i\|_2^2 \le c_6 m\log(nq)$ with probability at least $1 - 2\exp\{-c_7 m\log(nq)\}$. Thus, plugging in the bounds for $\max_{i,j} \|Z_j^i\|_2^2$ and

$\max_i \sigma_{\max}\left(Z^i \Psi (Z^i)^\top\right)$ into (B.52), we can show that with probability at least $1 - 2\exp\{-c(m \vee q)\log(n)\} - 2\exp\{-cm\log(nq)\}$, we have

$$\max_j \|W_j\|_F^2 \leq c(m^2 \vee q^2)\log^2(n) n m^2 \log^2(nq). \tag{B.54}$$

Next, for the term $\|W_j\|_2$, we can write

$$\begin{aligned}\max_j \|W_j\|_2 &= \max_i \max_j \left\| \Sigma^i_{\theta^*} Z^i_j (Z^i_j)^\top - \Sigma^i_{\theta^*} \operatorname{diag}(Z^i_j)^2 \right\|_2 \\ &\leq \max_i \max_j \|\Sigma^i_{\theta^*}\|_2 \|Z^i_j\|_2^2 + \max_i \max_j \|\Sigma^i_{\theta^*}\|_2 \left\|Z^i_j\right\|_\infty^2 .\end{aligned}$$

Then by the normality of Z^i_j, we can write

$$\mathbb{P}\left(\max_i \max_j \left| \left\|Z^i_j\right\|_\infty^2 - \left(\Sigma^i_Z\right)_{j,j} \right| \geq t \mid \Sigma^i_Z \right) \leq 2\exp\{\log(m) + \log(n) + \log(q) - \min\{t^2/8, t/8\}\}, \tag{B.55}$$

which implies $\max_i \max_j \left\|Z^i_j\right\|_\infty^2 \leq c_1 \log(nmq)$ with probability at least $1 - 2\exp\{-c_2 \log(nmq)\}$. Then using Lemma B.1.2.4 to bound $\max_i \|\Sigma^i_{\theta^*}\|_2$ and inequality (B.53) to bound $\max_i \max_j \|Z^i_j\|_2^2$, we have

$$\begin{aligned}\max_j \|W_j\|_2 &\leq c_3(m \vee q)\log(n) m \log(nq) + c_4(m \vee q)\log(n)\log(nmq) \\ &\leq c_5(m \vee q)\log(n) m \log(nmq)\end{aligned} \tag{B.56}$$

which holds with probability at least $1 - 2\exp\{-c(m \vee q)\log(n)\} - 2\exp\{-cm\log(nq)\} - 2\exp\{-c\log(nmq)\}$.

Plugging in the bounds in (B.54) and (B.56) into (B.51), and taking $t = c_6(m \vee q) m\sqrt{n}\log(n)\log(nmq)\log(q)$, we have

$$\max_j |E_{1,j}| \leq c_6(m \vee q)\log(n)\log(nmq)\log(q) m\sqrt{n}$$

with probability at least $1 - 2\exp\{-c(m \vee q)\log(n)\} - 2\exp\{-cm\log(nq)\} -$

$2\exp\{-c\log(nmq)\} - 2\exp\{-c\log(q)\}$.

Lemma B.4.3.2)

We can write

$$\begin{aligned}
\max_j |E_{2,j}| &= \max_j \left| \sum_{i\in S_2} \operatorname{tr}\left(A_j^i X^i (\beta^* - \hat{\beta})(\beta^* - \hat{\beta})^\top (X^i)^\top \right) \right| \\
&\leq \max_j \left| \max_i \left(\sigma_{\max}\left(A_j^i \Sigma_a^i \right) \right) \|\Sigma_a^{-1/2} X(\beta^* - \hat{\beta})\|_2^2 \right| \\
&= \|\Sigma_a^{-1/2} X(\beta^* - \hat{\beta})\|_2^2 \max_j \max_i \sigma_{\max}\left(\Sigma_a^i\right) \sigma_{\max}\left(Z_j^i \left(Z_j^i\right)^\top - \operatorname{diag}\left(Z_j^i\right)^2 \right) \\
&\leq \|\Sigma_a^{-1/2} X(\beta^* - \hat{\beta})\|_2^2 \max_i \sigma_{\max}(\Sigma_a^i) \left(\max_{i,j} \|Z_j^i\|_2^2 + \max_{i,j} \left\|Z_j^i\right\|_\infty^2 \right).
\end{aligned}$$

Thus, using Theorem B.1.1 to bound $\|\Sigma_a^{-1/2} X(\beta^* - \hat{\beta})\|_2^2$, using Lemma B.1.2.4 to bound $\sigma_{\max}(\Sigma_a^i)$, and using (B.53) and (B.55), we have

$$\max_j |E_{2,j}| \leq \begin{cases} c_1 s\log(p)\log(n)\log(nq)mq & , q > c_0 m \\ c_1 s\log(p)\log^3(n)\log(nq)m^2 & , q > c_0 m \text{ and assuming Assumption 9.1} \\ c_1 s\log(p)\log(n)\log(nq)m^2 & , m > c_0 q,\ p = q \\ c_1 s\log(p)\log(n)\log(nq)m^3 & , m > c_0 q,\ p > q, \end{cases}$$

with probability at least $1 - 4\exp\{-cn\} - 12\exp\{-c\log(n)\} - 2\exp\{-cmnq^{-\mathbf{1}\{q>c_0 m\}}\} - \exp\{-cn(m/q)^{\mathbf{1}\{q>c_0 m\}}\} - 2\exp\{-cm\log(nq)\} - 2\exp\{-c\log(nmq)\} - 2\exp\{-c(m \vee q)\log(n)\}$.

Lemma B.4.3.3)

We can write

$$E_{3,j} = \sum_{i\in S_2} \left(\left(\beta^* - \hat{\beta}\right)^\top (X^i)^\top A_j^i \left(\Sigma_{\theta^*}^i\right)^{1/2} \right) \left(\Sigma_{\theta^*}^i\right)^{-1/2} r_i.$$

Conditioning on X and $\hat{\beta}$, we have $\left(\Sigma^i_{\theta^*}\right)^{-1/2} r_i \mid X, \hat{\beta} \sim N(0, I_m)$. Thus,

$$\mathbb{P}\left(\max_{1\le j\le q} |E_{3,j}| \ge t | X, \hat{\beta}\right) \le 2\exp\left\{\log(q) - \frac{t^2}{8\max_j \sum_{i\in S_2} (\beta^* - \hat{\beta})^\top (X^i)^\top A^i_j \Sigma^i_{\theta^*} A^i_j X^i (\beta^* - \hat{\beta})}\right\}. \tag{B.57}$$

Then, with probability at least $1 - 4\exp\{-cn\} - 12\exp\{-c\log(n)\} - 2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} - \exp\{-cn(m/q)^{\mathbf{1}\{q>c_0m\}}\} - 2\exp\{-cm\log(nq)\} - 2\exp\{-c\log(nmq)\} - 2\exp\{-c(m\vee q)\log(n)\}$, we have

$$\begin{aligned}
&\max_j \sum_{i\in S_2} (\beta^* - \hat{\beta})^\top (X^i)^\top A^i_j \Sigma^i_{\theta^*} A^i_j X^i (\beta^* - \hat{\beta}) \\
&\le \|\Sigma_a^{-1/2} X(\beta^* - \hat{\beta})\|_2^2 \max_{i,j} \sigma_{\max}\left(A^i_j \Sigma^i_{\theta^*} A^i_j \Sigma^i_a\right) \\
&\le \|\Sigma_a^{-1/2} X(\beta^* - \hat{\beta})\|_2^2 \max_{i,j} \sigma^2_{\max}\left(A^i_j\right) \sigma_{\max}\left(\Sigma^i_{\theta^*}\right) \sigma_{\max}\left(\Sigma^i_a\right) \\
&\le c_1 \|\Sigma_a^{-1/2} X(\beta^* - \hat{\beta})\|_2^2 \left(\max_{i,j} \|Z^i_j\|_2^2 + \max_{i,j} \|Z^i_j\|_\infty^2\right)^2 \max_i \sigma_{\max}\left(\Sigma^i_{\theta^*}\right) \sigma_{\max}\left(\Sigma^i_a\right) \\
&\le c_2 \begin{cases} s\log(p)\log^2(n)\log^2(nq)m^2q^2 & , q > c_0 m \\ s\log(p)\log^4(n)\log^2(nq)m^3q & , q > c_0 m \text{ and assuming Assumption 9.1} \\ s\log(p)\log^2(n)\log^2(nq)m^4 & , m > c_0 q,\ p = q \\ s\log(p)\log^2(n)\log^2(nq)m^5 & , m > c_0 q,\ p > q, \end{cases}
\end{aligned} \tag{B.58}$$

where the last inequalities (B.58) use Theorem B.1.1, Lemma B.1.2.4, and the inequalities (B.53) and (B.55). Plugging (B.58) into (B.57) and we can obtain

$$\max_j |E_{3,j}| \le c_3 \begin{cases} \sqrt{s\log(p)\log(q)\log^2(n)\log^2(nq)m^2q^2} & , q > c_0 m \\ \sqrt{s\log(p)\log(q)\log^4(n)\log^2(nq)m^3q} & , q > c_0 m \text{ and assuming Assumption 9.1} \\ \sqrt{s\log(p)\log(q)\log^2(n)\log^2(nq)m^4} & , m > c_0 q,\ p = q \\ \sqrt{s\log(p)\log(q)\log^2(n)\log^2(nq)m^5} & , m > c_0 q,\ p > q, \end{cases}$$

with probability at least $1 - 4\exp\{-cn\} - 12\exp\{-c\log(n)\} - 2\exp\{-cmnq^{-\mathbf{1}\{q>c_0m\}}\} -$

$\exp\{-cn(m/q)^{\mathbf{1}\{q>c_0 m\}}\} - 2\exp\{-cm\log(nq)\} - 2\exp\{-c\log(nmq)\} - 2\exp\{-c\log(q)\} - 2\exp\{-c(m\vee q)\log(n)\}$. □

Proof of Lemma B.4.4.

For simpler notation, we denote the (j,k) entry of Σ_Z^i by $\sigma_{i,j,k}$ for the rest of the proof. We denote the lth entry of the vector Z_j^i by $Z_{l,j}^i$.

Lemma B.4.4.1)

Defining matrix $H^i = (Z^i)^\top Z^i$ for $i = 1,\ldots,n$, we can write $H_{j,k}^i = (Z_j^i)^\top Z_k^i = \sum_{l=1}^m Z_{l,j}^i Z_{l,k}^i$. We will focus on the entry-wise concentration of the matrices H^i.

First, for the diagonal terms $1 \le j = k \le q$, we have $H_{j,j}^i = \sum_{l=1}^m \left(Z_{l,j}^i\right)^2$. Since $Z^i \mid \Sigma_Z^i \sim MN_{m\times q}(0, I_m, \Sigma_Z^i)$, we have $(Z_{l,j}^i)^2 - \sigma_{i,j,j} \mid \Sigma_Z^i \in \mathrm{SE}(4\sigma_{i,j,j}^2, 4\sigma_{i,j,j})$, and thus $H_{j,j}^i - m\sigma_{i,j,j} \mid \Sigma_Z^i \in \mathrm{SE}(4m\sigma_{i,j,j}^2, 4\sigma_{i,j,j})$. Therefore, based on the tail bound of sub-exponential random variables,

$$\mathbb{P}\left(\max_{i,j}\left|H_{j,j}^i - m\sigma_{i,j,j}\right| > t\right) \le 2\exp\left\{\log(nq) - \min\left\{\frac{t^2}{8m\sigma_{i,j,j}^2}, \frac{t}{8\sigma_{i,j,j}}\right\}\right\}.$$

Taking $t = 16\sigma_{i,j,j}\sqrt{m}\log(nq)$, we obtain $\max_{i,j}\left|H_{j,j}^i - m\sigma_{i,j,j}\right| \le 16\sigma_{i,j,j}\sqrt{m}\log(nq)$ with probability at least $1 - 2\exp\{-c\log(nq)\}$. Under Assumption 13.2, this implies that $\forall i, j$,

$$m^2 + 16^2 m\log^2(nq) - 32m^{3/2}\log(nq) \le \frac{1}{\sigma_{i,j,j}^2}(H_{j,j}^i)^2 \le m^2 + 16^2 m\log^2(nq) + 32m^{3/2}\log(nq). \tag{B.59}$$

Next, for the off-diagonal terms $1 \le j \ne k \le q$, we have the fact that for any two normally distributed random variables U, V with mean 0, variance 1 and correlation ρ,

$$UV = \frac{(U+V)^2}{4} - \frac{(U-V)^2}{4} = \frac{1+\rho}{2}Q_1 - \frac{1-\rho}{2}Q_2,$$

where Q_1 and Q_2 are independent and follow a χ_1^2 distribution. Thus, the random variable

UV is sub-exponential with mean ρ, and hence $UV-\rho \in \mathrm{SE}(2\rho^2+2, 2+2|\rho|)$. We thus have

$$\begin{aligned} &Z^i_{l,k}Z^i_{l,j} - \sigma_{i,j,k} \in \mathrm{SE}(2\sigma^2_{i,j,k} + 2\sigma_{i,j,j}\sigma_{z,k,k}, 2\sqrt{\sigma_{i,j,j}\sigma_{i,k,k}} + 2|\sigma_{i,j,k}|) \\ \Rightarrow &H^i_{j,k} - m\sigma_{i,j,k} \in \mathrm{SE}(2m\sigma^2_{i,j,k} + 2m\sigma_{i,j,j}\sigma_{i,k,k}, 2\sqrt{\sigma_{i,j,j}\sigma_{i,k,k}} + 2|\sigma_{i,j,k}|). \end{aligned}$$

Denoting $\rho_{i,j,k} = \frac{\sigma_{i,j,k}}{\sqrt{\sigma_{i,j,j}\sigma_{i,k,k}}}$, based on the tail bound of sub-exponential random variables, we have that

$$\begin{aligned} &\mathbb{P}\left(\max_{i,j,k}\left|H^i_{j,k} - m\sigma_{i,j,k}\right| > t\right) \\ &\leq 2\exp\left\{\log(q^2n) - \min\left\{\frac{t^2}{m\sigma_{i,j,j}\sigma_{i,k,k}\left(4\rho^2_{i,j,k}+4\right)}, \frac{t}{\sqrt{\sigma_{i,j,j}\sigma_{i,k,k}}\left(4+4\left|\rho_{i,j,k}\right|\right)}\right\}\right\}. \end{aligned}$$

Thus, taking $t = \sqrt{8m\sigma_{i,j,j}\sigma_{i,k,k}\left(|\rho_{i,j,k}|+1\right)}\log(nq^2)$, with probability at least $1 - 2\exp\{-c\log(nq^2)\}$ we have

$$\max_{i,j,k}\left|H^i_{j,k} - m\sigma_{i,j,k}\right| \leq \sqrt{8m\sigma_{i,j,j}\sigma_{i,k,k}\left(|\rho_{i,j,k}|+1\right)}\log(nq^2). \tag{B.60}$$

To study the lower and upper bounds of $|H^i_{j,k}|^2$ based on (B.60), we consider two cases:

1. If for given i,j,k, $m|\sigma_{i,j,k}| \gg \sqrt{\sigma_{i,k,k}\sigma_{i,j,j}(|\rho_{j,k}|+1)m}\log(nq^2)$, i.e., $\frac{|\rho_{i,j,k}|+1}{\rho^2_{i,j,k}} \ll \frac{m}{\log^2(nq^2)}$: The correlation $|\rho_{i,j,k}|$ is not too small, and (B.60) implies that $H^i_{j,k}$ is bounded away from zero with high probability. Then with probability at least $1-2\exp\{-c\log(nq^2)\}$, we have that

$$\begin{aligned} \frac{(H^i_{j,k})^2}{\sigma_{z,j,j}\sigma_{z,k,k}} &\leq m^2\rho^2_{i,j,k} + 8m(|\rho_{i,j,k}|+1)\log^2(nq^2) + 2m|\rho_{i,j,k}|\sqrt{8m(|\rho_{i,j,k}|+1)}\log(nq^2) \\ &\leq m^2\rho^2_{i,j,k} + 16m\log^2(nq^2) + 8m|\rho_{i,j,k}|\sqrt{m}\log(nq^2), \qquad \text{(B.61)} \\ \frac{(H^i_{j,k})^2}{\sigma_{z,j,j}\sigma_{z,k,k}} &\geq m^2\rho^2_{i,j,k} + 8m(|\rho_{i,j,k}|+1)\log^2(nq^2) - 2m|\rho_{i,j,k}|\sqrt{8m(|\rho_{i,j,k}|+1)}\log(nq^2) \\ &\geq m^2\rho^2_{i,j,k} + 8m\log^2(nq^2) - 8m|\rho_{i,j,k}|\sqrt{m}\log(nq^2). \qquad \text{(B.62)} \end{aligned}$$

Note that the condition $\frac{|\rho_{i,j,k}|+1}{\rho_{i,j,k}^2} \ll \frac{m}{\log^2(nq^2)}$ implies $\rho_{i,j,k}^2 \gg \frac{\log^2(nq^2)}{m}$, which then implies

$$m^{3/2}|\rho_{i,j,k}| \log(nq^2) \gg m \log^2(nq^2). \tag{B.63}$$

2. If for given i, j, k, $m|\sigma_{i,j,k}| = O\left(\sqrt{\sigma_{i,k,k}\sigma_{i,j,j}(|\rho_{i,j,k}|+1)m}\log(nq^2)\right)$, i.e., $\frac{\rho_{i,j,k}^2}{|\rho_{i,j,k}|+1} = O\left(\frac{\log^2(nq^2)}{m}\right)$: The correlation $\rho_{i,j,k}$ is very small under Assumption 13.2, and (B.60) implies that $H_{j,k}^i$ is not bounded away from zero. Then with probability at least $1 - 2\exp\{-c\log(nq^2)\}$, we have

$$0 \leq \frac{(H_{j,k}^i)^2}{\sigma_{i,j,j}\sigma_{i,k,k}} \leq m^2\rho_{j,k}^2 + 16m\log^2(nq^2) + 8m|\rho_{i,j,k}|\sqrt{m}\log(nq^2). \tag{B.64}$$

Note that the condition $\frac{\rho_{i,j,k}^2}{|\rho_{i,j,k}|+1} = O\left(\frac{\log^2(nq^2)}{m}\right)$ implies $\rho_{i,j,k}^2 \leq c\log^2(nq^2)/m$ for some constant $c > 0$, which then implies

$$m^{3/2}|\rho_{i,j,k}| \log(nq^2) \leq c_1 m \log^2(nq^2) \tag{B.65}$$

for some constant $c_1 > 0$.

Note that for $v \in \mathbb{R}^q$, we can write

$$v^\top \left(G_1^i - \mathbb{E}\left(G_1^i\right)\right) v = v^\top \left(H^i \circ H^i - \mathbb{E}\left(H^i \circ H^i\right)\right) v \leq \sum_{j,k=1}^{q} |v_j||v_k| \left|\left(H_{j,k}^i\right)^2 - \mathbb{E}\left(\left(H_{j,k}^i\right)^2\right)\right| \leq \|v\|_2^2 \|W^i\|_2$$

$$v^\top \left(G_1^i - \mathbb{E}\left(G_1^i\right)\right) v \geq -\sum_{j,k=1}^{q} |v_j||v_k| \left|\left(H_{j,k}^i\right)^2 - \mathbb{E}\left(\left(H_{j,k}^i\right)^2\right)\right| \geq -\|v\|_2^2 \|W^i\|_2$$

where we define the matrix W^i such that $W_{j,k}^i = \left|\left(H_{j,k}^i\right)^2 - \mathbb{E}\left(\left(H_{j,k}^i\right)^2\right)\right|$. We compute

the values of $\mathbb{E}\left(\left(H^i_{j,k}\right)^2\right)$ based on the distribution $Z^i \mid \Sigma^i_Z \sim MN_{m\times q}(0, I_m, \Sigma^i_Z)$:

$$\mathbb{E}\left(\left(H^i_{j,k}\right)^2\right) = \begin{cases} (m^2+2m)\sigma^2_{i,j,j} & , j = k \\ (m^2+m)\sigma^2_{i,j,k} + m\sigma_{i,j,j}\sigma_{i,k,k} & , j \neq k. \end{cases}$$

Thus, based on (B.59)–(B.65), with probability at least $1 - 2\exp\{-c\log(nq)\} - 2\exp\{-c\log(nq^2)\}$, for a given i we have

$$W^i_{j,k} \leq \begin{cases} c_2 m^{3/2}\log(nq) & , \quad 1 \leq j = k \leq q; \\ c_3 m^{3/2}|\rho_{i,j,k}|\log(nq^2) & , \quad 1 \leq j \neq k \leq q, \text{ and } \frac{\rho^2_{i,j,k}}{|\rho_{i,j,k}|+1} \gg \frac{\log^2(nq^2)}{m} \\ c_4 m\log^2(nq^2) & , \quad 1 \leq j \neq k \leq q, \text{ and } \frac{\rho^2_{i,j,k}}{|\rho_{i,j,k}|+1} = O\left(\frac{\log^2(nq^2)}{m}\right) \end{cases}.$$

Using Gershgorin circle theorem [98], we can obtain the following bounds for the eigenvalues of the matrices W^i with probability at least $1 - 2\exp\{-c\log(nq)\} - 2\exp\{-c\log(nq^2)\}$:

$$\begin{aligned} \max_i |\sigma(W^i)| &\leq \max_{i,j}\left(|W^i_{j,j}| + \sum_{k=1,k\neq j}^{q} |W^i_{j,k}|\right) \\ &\leq \max_i \left(c_5 m^{3/2}\log(nq^2) + c_5 s^i_Z m^{3/2}\log(nq^2) + c_5 m(q - s^i_Z)\log^2(nq^2)\right) \\ &\leq c_5 m^{3/2}\log(nq^2) + c_5 s_Z m^{3/2}\log(nq^2) + c_5 mq\log^2(nq^2), \end{aligned}$$

where s^i_Z and s_Z are defined in B.38. Thus, with probability at least $1 - 2\exp\{-c\log(nq)\} - 2\exp\{-c\log(nq^2)\}$, we have $\left\|G^i_1 - \mathbb{E}\left(G^i_1\right)\right\|_2 \leq c_5 m^{3/2}\log(nq^2) + c_5 s_Z m^{3/2}\log(nq^2) + c_5 mq\log^2(nq^2)$.

Lemma B.4.4.2)

The matrix $Z^i \circ Z^i$ has independent rows, and its entries follow scaled χ^2_1 distributions. Thus, by Theorem 5.44 (random matrices with heavy-tailed non-isotropic independent rows) in

[223], with probability at least $1 - \exp\{\log(nq) - ct^2\}$ we have

$$\begin{aligned}\|Z^i \circ Z^i\|_2 &\leq \|2\Sigma_Z^i \circ \Sigma_Z^i + u_Z^i (u_Z^i)^\top\|_2^{1/2} \sqrt{m} + t\sqrt{m} \\ &\leq \left(\|2\Sigma_Z^i \circ \Sigma_Z^i\|_2^{1/2} + \|u_Z^i\|_2\right) \sqrt{m} + t\sqrt{m},\end{aligned}$$

where $u_Z^i = \left(\left(\Sigma_Z^i\right)_{1,1}, \ldots, \left(\Sigma_Z^i\right)_{q,q}\right)^\top$. Note that under Assumption 8.2, we have $\|\Sigma_Z^i \circ \Sigma_Z^i\|_2 \leq \|\Sigma_Z^i\|_2^2 \asymp 1$ [189, 116], and $\|u_Z^i\|_2 \asymp \sqrt{q}$. Thus, taking $t = c_1\sqrt{q\log(n)}$, with probability at least $1 - \exp\{-cq\log(n)\}$ we have

$$\|Z^i \circ Z^i\|_2 \leq c_2\sqrt{mq\log(n)}. \tag{B.66}$$

Since $Z^i \mid \Sigma_Z^i \sim MN_{m\times q}(0, I_m, \Sigma_Z^i)$, we have $\mathbb{E}\left(G_2^i\right) = 2m\Sigma_Z^i \circ \Sigma_Z^i + mu_Z^i(u_Z^i)^\top$, which gives

$$\left\|\mathbb{E}\left(G_2^i\right)\right\|_2 \leq 2m\|\Sigma_Z^i\|_2^2 + m\|u_Z^i\|_2^2 \leq cmq. \tag{B.67}$$

Thus, by (B.66) and (B.67), we have $\|G_2^i - \mathbb{E}(G_2^i)\|_2^2 \leq \|Z^i \circ Z^i\|_2^2 + \|\mathbb{E}(G_2^i)\|_2 \leq c_3 mq\log(n)$ with probability at least $1 - \exp\{-cq\log(n)\}$.

□

Proof of Lemma B.4.5. Recalling the definition of $B_{j,k} = \sum_{i\in S_2} \operatorname{tr}\left(A_j^i A_k^i\right)$ in (B.37), for a fixed vector $v \in \mathbb{R}^q$, we can write

$$\begin{aligned}v^\top B v &= \sum_{i\in S_2}\sum_{j,k=1}^{q} \operatorname{tr}\left(\left(Z_j^i(Z_j^i)^\top - \operatorname{diag}(Z_j^i)^2\right)\left(Z_k^i(Z_k^i)^\top - \operatorname{diag}(Z_k^i)^2\right)\right) v_j v_k \\ &= \sum_{i\in S_2}\sum_{j,k=1}^{q} \operatorname{tr}\left(Z_j^i(Z_j^i)^\top Z_k^i(Z_k^i)^\top\right) v_j v_k - \operatorname{tr}\left(Z_j^i(Z_j^i)^\top \operatorname{diag}(Z_k^i)^2\right) v_j v_k \\ &= \sum_{i\in S_2} v^\top\left((Z^i)^\top Z^i\right) \circ \left((Z^i)^\top Z^i\right) v - v^\top\left(Z^i \circ Z^i\right)^\top\left(Z^i \circ Z^i\right) v.\end{aligned} \tag{B.68}$$

Then, recalling the definitions of $G_1^i = \left((Z^i)^\top Z^i\right) \circ \left((Z^i)^\top Z^i\right)$, $G_2^i = \left(Z^i \circ Z^i\right)^\top\left(Z^i \circ Z^i\right)$ in

Lemma B.4.4, we can write (B.68) as

$$v^\top Bv = v^\top \left(\sum_{i\in S_2} \left(G_1^i - \mathbb{E}\left(G_1^i\right)\right) - \sum_{i\in S_2} \left(G_2^i - \mathbb{E}\left(G_2^i\right)\right) + \sum_{i\in S_2} \left(\mathbb{E}\left(G_1^i\right) - \mathbb{E}\left(G_2^i\right)\right) \right) v. \tag{B.69}$$

We study each of the terms in the parenthesis in (B.69) separately.

First, since $Z^i \mid \Sigma_Z^i \sim MN_{m\times q}(0, I_m, \Sigma_Z^i)$, we have $\mathbb{E}\left(G_1^i\right) - \mathbb{E}\left(G_2^i\right) = m(m-1)\left(\Sigma_Z^i \circ \Sigma_Z^i\right)$. Since we know that for two matrices A_1 and A_2, $A_1 \circ A_2$ is a principle submatrix of the Kronecker product $A_1 \otimes A_2$, we can show that $\sigma_{\min}(A_1)\sigma_{\min}(A_2) \le \sigma_{\min}\left(A_1 \otimes A_2\right) \le \sigma_{\min}\left(A_1 \circ A_2\right)$ [189, 116]. Thus, since $\sigma(\Sigma_Z^i) \asymp 1$ by Assumption 8.2, we have

$$\sigma_{\min}\left(\sum_{i\in S_2} \mathbb{E}\left(G_1^i\right) - \mathbb{E}\left(G_2^i\right) \right) \ge c_1 n m^2. \tag{B.70}$$

Next, for the term $\sum_{i\in S_2}\left(G_1^i - \mathbb{E}\left(G_1^i\right)\right)$, using Lemma B.4.4.1, we have $\max_i \|G_1^i - \mathbb{E}\left(G_1^i\right)\|_2 \le c_1 m^{3/2}\log(nq^2) + c_1 s_Z m^{3/2}\log(nq^2) + c_1 mq\log^2(nq^2)$ with probability at least $1 - 2\exp\{-c\log(nq)\} - 2\exp\{-c\log(nq^2)\}$. Then, by matrix Bernstein inequality [219], we get

$$\mathbb{P}\left(\left\| \sum_{i\in S_2} G_1^i - \mathbb{E}\left(G_1^i\right) \right\|_2 \ge t \right) \le \exp\left\{ \log(2q) - \frac{t^2}{2} \frac{1}{\nu_1 + c_2\left(s_Z m^{3/2}\log(nq^2) + mq\log^2(nq^2)\right)t} \right\}, \tag{B.71}$$

where $\nu_1 = n_2 \left\| \mathbb{E}(G_1^i (G_1^i)^\top) - \mathbb{E}(G_1^i)\mathbb{E}(G_1^i)^\top \right\|_2$. Denote the (j,k) entry of Σ_Z^i by $\sigma_{i,j,k}$.

Based on $Z^i \mid \Sigma_Z^i \sim MN_{m\times q}(0, I_m, \Sigma_Z^i)$, with some careful calculations, we can get

$$\begin{aligned}
\left(\mathbb{E}(G_1^i G_1^{i,\top}) - \mathbb{E}(G_1^i)\mathbb{E}(G_1^i)^\top\right)_{j,j} &= \sum_{l=1}^{q} \big(2m(m+3)\sigma_{i,j,j}^2\sigma_{i,l,l}^2 + 4m(m^2+7m+9)\sigma_{i,j,j}\sigma_{i,l,l}\sigma_{i,j,l}^2 \\
&\quad + 2(2m^2+5m+3)m\sigma_{i,j,l}^4\big) \\
\left(\mathbb{E}(G_1^i G_1^{i,\top}) - \mathbb{E}(G_1^i)\mathbb{E}(G_1^i)^\top\right)_{j,k} &= \sum_{l=1}^{q} (m(m+2)(m+3)4\sigma_{i,j,k}\sigma_{i,j,l}\sigma_{i,k,l}\sigma_{i,l,l} \\
&\quad + 2m(2m+3)\left(\sigma_{i,j,l}^2\sigma_{i,k,k}\sigma_{i,l,l} + \sigma_{i,j,j}\sigma_{i,k,l}^2\sigma_{i,l,l}\right) \\
&\quad + 2m(m+1)(2m+3)\sigma_{i,j,l}^2\sigma_{i,k,l}^2 + 2m(m+2)\sigma_{i,j,k}^2\sigma_{i,l,l}^2 \\
&\quad + 2m\sigma_{i,j,j}\sigma_{i,k,k}\sigma_{i,l,l}^2\big)
\end{aligned}$$

Then, using Gershgorin circle theorem [98], and under Assumption 8.2 which implies $\sigma_{i,j,k} = O(1)$, we can bound the maximum eigenvalue of $\mathbb{E}(G_1^i G_1^{i,\top}) - \mathbb{E}(G_1^i)\mathbb{E}(G_1^i)^\top$ by

$$\begin{aligned}
&\left\|\mathbb{E}(G_1^i G_1^{i,\top}) - \mathbb{E}(G_1^i)\mathbb{E}(G_1^i)^\top\right\|_2 \\
&\le \max_j \left(\left|\mathbb{E}(G_1^i G_1^{i,\top}) - \mathbb{E}(G_1^i)\mathbb{E}(G_1^i)^\top\right|_{j,j} + \sum_{k=1,k\neq j}^{q}\left|\mathbb{E}(G_1^i G_1^{i,\top}) - \mathbb{E}(G_1^i)\mathbb{E}(G_1^i)^\top\right|_{j,k}\right) \\
&\le c_1 \max_j \left(m^3\left(\sum_{l=1}^{q}\sigma_{i,j,l}^2 + \sum_{k,l=1}^{q}|\sigma_{i,k,l}|\right) + qm^2\left(2 + \sum_{l=1}^{q}\sigma_{i,j,l}^2\right) + m^2\sum_{k,l=1}^{q}\sigma_{i,k,l}^2\right) \\
&\le c_2 m^3 q^{3/2} \qquad \text{(B.72)}
\end{aligned}$$

where the last inequality follows from $\sum_{k=1}^{q}\sigma_{i,k,l}^2 \le c_3\sum_{k=1}^{q}|\sigma_{i,k,l}| \le c_3\|\Sigma_Z^i\|_1 \le c_3\sqrt{q}\|\Sigma_Z^i\|_2$, with matrix 1-norm defined by $\|A\|_1 = \max_j \sum_{k=1}^{q}|A_{j,k}|$ for matrix $A \in \mathbb{R}^{q\times q}$.

Therefore, taking $t = c_4 nm^2/\log(n)$ in (B.71) for some suitably large constant $c_4 > 0$, under Assumption 13.2, we have

$$\left\|\sum_{i\in S_2}\left(G_1^i - \mathbb{E}(G_1^i)\right)\right\|_2 \le c_5 nm^2/\log(n) \qquad \text{(B.73)}$$

with probability at least $1 - \exp\{-c\log(q)\} - 2\exp\{-c\log(nq)\}$.

Finally, for the term $\sum_{i\in S_2}\left(G_2^i-\mathbb{E}\left(G_2^i\right)\right)$, using Lemma B.4.4.2, with probability at least $1-\exp\{-cq\log(n)\}$ we have $\left\|G_2^i-\mathbb{E}\left(G_2^i\right)\right\|_2\leq c_1mq\log(n)$. Then by matrix Bernstein inequality [219],

$$\mathbb{P}\left(\left\|\sum_{i\in S_2}G_2^i-\mathbb{E}\left(G_2^i\right)\right\|_2\geq t\right)\leq\exp\left\{\log(2q)-\frac{t^2}{2}\frac{1}{\nu_2+c_2mq\log(n)t}\right\},\tag{B.74}$$

where $\nu_2=n_2\left\|\mathbb{E}\left(G_2^i(G_2^i)^\top\right)-\mathbb{E}\left(G_2^i\right)\mathbb{E}\left(G_2^i\right)^\top\right\|_2$. Based on $Z^i\mid\Sigma_Z^i\sim MN_{m\times q}(0,I_m,\Sigma_Z^i)$, with some careful calculation, we can get

$$\begin{aligned}\left(\mathbb{E}\left(G_2^i(G_2^i)^\top\right)-\mathbb{E}\left(G_2^i\right)\mathbb{E}\left(G_2^i\right)^\top\right)_{j,j}&=\sum_{l=1}^{q}m\left(20\sigma_{i,j,l}^4+8\sigma_{i,j,j}^2\sigma_{i,l,l}^2+68\sigma_{i,j,j}\sigma_{i,j,l}^2\sigma_{i,l,l}\right)=O_p(qm)\\
\left(\mathbb{E}\left(G_2^i(G_2^i)^\top\right)-\mathbb{E}\left(G_2^i\right)\mathbb{E}\left(G_2^i\right)^\top\right)_{j,k}&=\sum_{l=1}^{q}m\left(48\sigma_{i,j,k}\sigma_{i,j,l}\sigma_{i,k,l}\sigma_{i,l,l}+20\sigma_{i,j,l}^2\sigma_{i,k,l}^2+10\sigma_{i,j,l}^2\sigma_{i,k,k}\sigma_{i,l,l}\right.\\
&\quad\left.+6\sigma_{i,j,k}^2\sigma_{i,l,l}^2+2\sigma_{i,j,j}\sigma_{i,k,k}\sigma_{i,l,l}^2+10\sigma_{i,j,j}\sigma_{i,l,l}\sigma_{i,k,l}^2\right)\\
&=O_p(qm).\end{aligned}$$

Then by Gershgorin circle theorem, we can follow the similar arguments as those in (B.72) to obtain $\nu_2\leq c_3nmq^2$. Thus, taking $t=c_4nm^2/\log(n)$ in (B.74), under Assumption 13.2 we have

$$\left\|\sum_{i\in S_2}G_2^i-\mathbb{E}\left(G_2^i\right)\right\|_2\leq c_4nm^2/\log(n)\tag{B.75}$$

with probability at least $1-\exp\{-c\log(q)\}-\exp\{-cq\log(n)\}$.

Plugging the bounds (B.70), (B.73), and (B.75) into (B.69), we obtain

$$\begin{aligned}v^\top Bv&\geq\|v\|_2^2\left(\sigma_{\min}\left(\sum_{i\in S_2}\mathbb{E}\left(G_1^i\right)-\mathbb{E}\left(G_2^i\right)\right)-\left\|\sum_{i\in S_2}\left(G_1^i-\mathbb{E}\left(G_1^i\right)\right)\right\|_2-\left\|\sum_{i\in S_2}\left(G_2^i-\mathbb{E}\left(G_2^i\right)\right)\right\|_2\right)\\
&\geq c_5nm^2\|v\|_2^2\end{aligned}$$

with probability at least $1-\exp\{-c\log(q)\}-2\exp\{-c\log(nq)\}-\exp\{-cq\log(n)\}$. □

B.5 Additional Simulation Results

B.5.1 Results for the main simulation study

In this section, we present additional results for the simulation study in the main text. We include results for both $p = 20$ and $p = 60$ settings.

Figure B.1 presents separately the computation time for estimating and inferring β, and the computation time for estimating the variance components, for the proposed method, *LCL* and *dblasso*. Figure B.2 presents the selection consistency of the variance component corresponding to a single random effect. The numerical results of test power, type I error and confidence interval coverage for selected β_j's under each simulation setting, and the estimation accuracy in terms of MSE are illustrated in Table B.1–B.4.

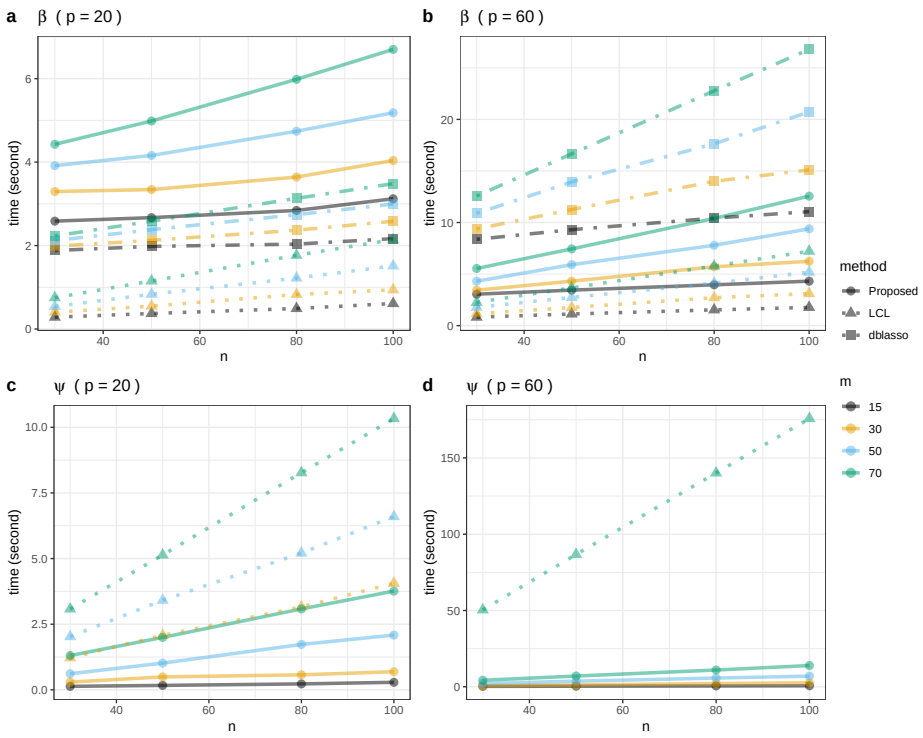

Figure B.1: **a** and **b**: Computation time to estimate and infer the fixed effect coefficients β. **c** and **d**: Computation time to estimate the variance components.

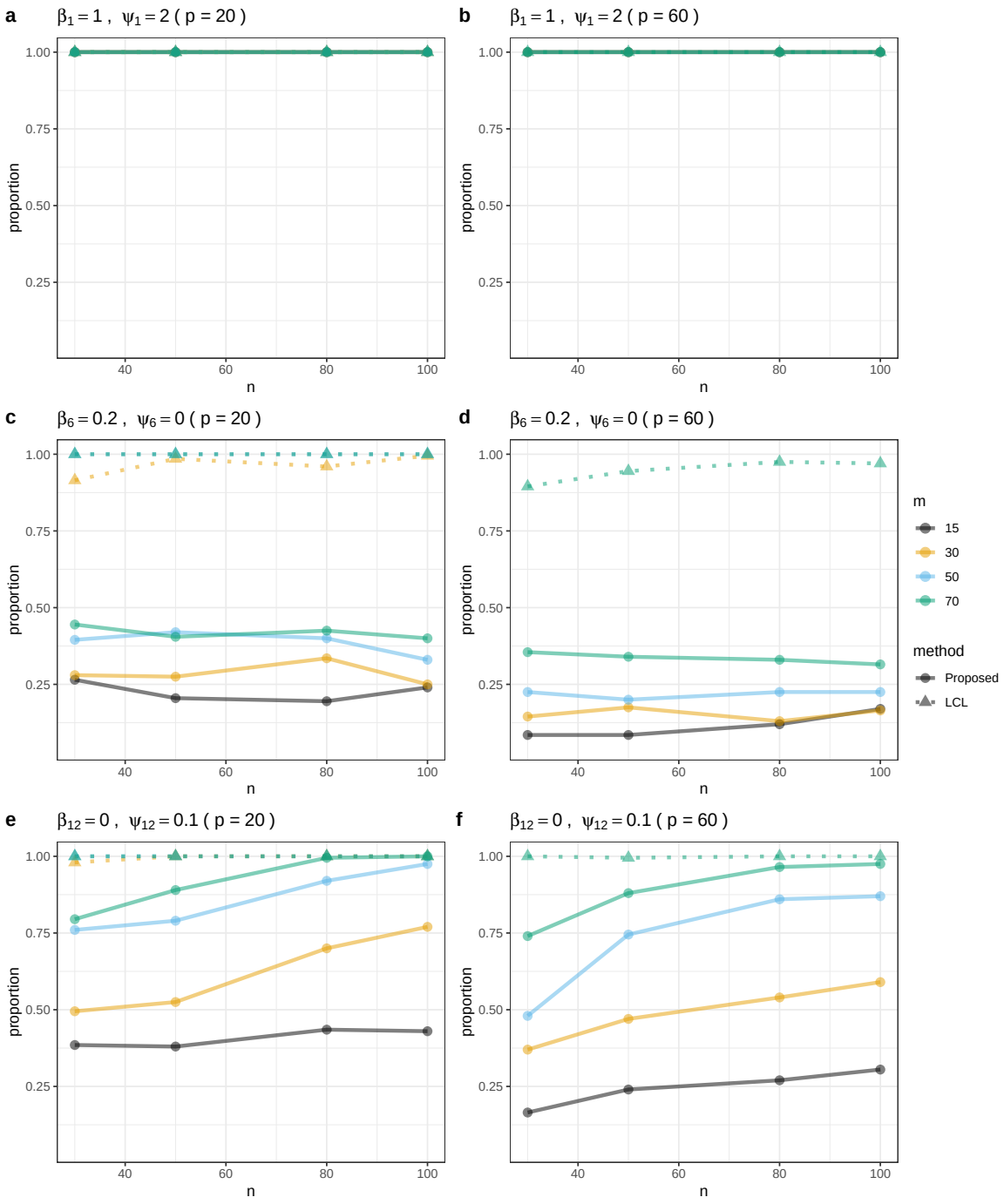

Figure B.2: .

Proportion of 200 Monte Carlo simulations that provide non-zero estimate for the selected variance component. The title of each subplot shows the true values of the targeted fixed effect coefficient β_j, the corresponding random effect variance component ψ_j, and the value of p in the setting. LCL does not estimate variance components when $m < q$, and thus for those m the results for LCL are missing.

Table B.1: Type-I error for testing zero β_l's at 0.05 significance level under the simulation settings in the main text. Bold face highlights type-I errors exceeding the normal range considering a Monte Carlo error of 0.03 with 200 replications.

			$\beta_{10}=0$ $\psi_{10}=4$			$\beta_{11}=0$ $\psi_{11}=0$			$\beta_{12}=0$ $\psi_{12}=0.1$		
p	m	n	Proposed	LCL	dblasso	Proposed	LCL	dblasso	Proposed	LCL	dblasso
20	15	30	0.06	0.04	**0.41**	0.045	0.035	0.04	0.045	0.055	0.06
		50	0.05	0.04	**0.44**	0.04	0.03	0.035	0.045	0.045	**0.09**
		80	0.055	0.05	**0.465**	0.03	0.035	0.065	0.08	0.065	0.04
		100	0.035	0.055	**0.445**	0.04	0.065	0.04	0.04	0.03	0.06
	30	30	0.04	0.04	**0.535**	0.035	0.08	0.065	0.035	**0.095**	0.07
		50	0.03	0.01	**0.525**	0.055	0.075	0.065	0.07	**0.085**	0.055
		80	0.03	0.055	**0.57**	0.03	0.08	0.05	0.03	0.035	0.075
		100	0.06	0.055	**0.525**	0.06	0.07	0.05	0.035	0.06	0.08
	50	30	0.075	**0.085**	**0.695**	0.07	**0.17**	0.035	0.05	**0.125**	**0.125**
		50	0.05	**0.085**	**0.655**	0.06	**0.16**	0.07	0.06	**0.11**	0.08
		80	0.04	0.04	**0.67**	0.055	**0.115**	0.065	0.03	0.075	**0.105**
		100	0.04	0.06	**0.65**	0.03	**0.155**	0.075	0.055	**0.125**	**0.125**
	70	30	0.06	0.06	**0.74**	0.055	**0.28**	0.055	0.065	**0.175**	**0.115**
		50	0.055	0.07	**0.72**	0.035	**0.225**	0.05	0.045	**0.135**	**0.105**
		80	0.04	0.07	**0.68**	0.04	**0.19**	0.04	0.05	**0.095**	**0.16**
		100	0.065	0.055	**0.71**	0.055	**0.22**	0.045	0.055	0.075	**0.125**
	120	30	0.045	0.07	**0.755**	0.035	**0.35**	0.03	0.025	**0.215**	**0.115**
		50	0.031	0.066	**0.74**	0.041	**0.311**	0.066	**0.087**	**0.199**	**0.173**
		80	0.056	0.061	**0.756**	0.03	**0.259**	0.046	0.056	**0.112**	**0.193**
		100	0.061	0.071	**0.77**	0.051	**0.27**	0.041	0.036	**0.107**	**0.179**
60	15	30	0.08	0.075	**0.425**	0.035	0.045	0.045	0.05	0.045	0.065
		50	0.03	0.03	**0.45**	0.045	0.05	0.06	0.06	0.07	0.04
		80	0.035	0.045	**0.465**	0.065	0.055	0.06	0.04	0.045	0.05
		100	0.045	0.05	**0.435**	0.04	0.05	0.035	0.065	0.06	0.08
	30	30	**0.09**	**0.1**	**0.525**	0.03	0.04	0.03	0.07	0.065	**0.105**
		50	0.065	0.05	**0.555**	0.04	0.06	0.06	0.04	0.04	0.065
		80	0.055	0.055	**0.585**	0.035	0.045	0.045	0.045	0.045	0.07
		100	0.08	0.07	**0.575**	0.06	0.055	0.055	0.055	0.055	0.08
	50	30	0.02	0.03	**0.705**	0.035	0.065	0.035	0.07	0.065	0.075
		50	0.055	0.055	**0.65**	0.055	0.045	0.045	0.07	0.055	0.08
		80	0.055	0.075	**0.665**	0.045	0.05	0.07	0.04	0.025	**0.09**
		100	**0.09**	0.075	**0.64**	0.07	0.05	0.05	0.03	0.035	**0.085**
	70	30	0.045	0.05	**0.665**	0.065	**0.09**	0.045	0.05	0.07	**0.11**
		50	0.06	0.055	**0.68**	0.065	**0.09**	0.035	0.035	0.04	0.08
		80	0.065	0.06	**0.715**	0.07	0.07	0.035	0.065	0.08	**0.11**
		100	0.065	0.06	**0.61**	0.065	0.045	0.045	0.025	0.03	**0.095**
	120	30	0.045	0.055	**0.725**	0.055	**0.235**	0.05	0.03	**0.145**	**0.115**
		50	0.06	0.06	**0.744**	0.04	**0.216**	0.05	0.06	**0.156**	**0.181**
		80	0.06	0.08	**0.745**	0.06	**0.205**	0.055	0.06	**0.13**	**0.17**
		100	**0.09**	0.08	**0.765**	0.04	**0.12**	0.04	0.025	**0.085**	**0.145**

Table B.2: Confidence interval coverage for the β_l's under the simulation settings in the main text. Bold face highlights type-I errors exceeding the normal range considering a Monte Carlo error of 0.03 with 200 replications.

			$\beta_1 = 1$ $\psi_1 = 2$			$\beta_2 = 0.5$ $\psi_2 = 0$			$\beta_6 = 0.2$ $\psi_6 = 0$			$\beta_7 = 0.1$ $\psi_7 = 0.1$			$\beta_9 = 0.05$ $\psi_9 = 0.1$			$\beta_{10} = 0$ $\psi_{10} = 4$			$\beta_{11} = 0$ $\psi_{11} = 0$			$\beta_{12} = 0$ $\psi_{12} = 0.1$		
p	m	n	Proposed	LCL	dblasso	Proposed	LCL	dblasso	Proposed	LCL	dblasso	Proposed	LCL	dblasso	Proposed	LCL	dblasso	Proposed	LCL	dblasso	Proposed	LCL	dblasso	Proposed	LCL	dblasso
20	15	30	0.945	0.945	**0.72**	0.92	0.955	0.935	0.92	**0.915**	0.96	0.935	0.935	**0.91**	0.925	0.95	**0.915**	0.94	0.96	**0.59**	0.955	0.965	0.96	0.955	0.945	0.94
		50	0.94	0.94	**0.725**	0.925	0.945	0.985	0.935	0.96	**0.915**	0.94	0.945	0.965	0.935	0.93	0.925	0.95	0.96	**0.56**	0.96	0.97	0.965	0.955	0.955	**0.91**
		80	0.925	0.96	**0.755**	0.945	0.93	0.92	0.93	0.95	0.96	0.925	0.965	0.965	0.965	0.95	0.935	0.945	0.95	**0.535**	0.97	0.965	0.935	0.92	0.935	0.96
		100	0.965	0.965	**0.78**	0.945	0.94	0.96	0.975	0.965	0.97	0.945	0.96	0.935	0.93	0.94	0.925	0.965	0.945	**0.555**	0.96	0.935	0.96	0.96	0.97	0.94
	30	30	0.93	**0.915**	**0.605**	0.985	0.94	0.955	0.98	**0.905**	0.965	0.935	**0.91**	0.955	0.945	**0.875**	0.92	0.96	0.96	**0.465**	0.965	0.92	0.935	0.965	**0.905**	0.93
		50	0.945	0.925	**0.605**	0.975	0.955	0.94	0.945	0.96	0.94	0.94	**0.915**	**0.9**	0.93	**0.89**	**0.915**	0.97	0.99	**0.475**	0.945	0.925	0.935	0.93	**0.915**	0.945
		80	0.95	0.94	**0.545**	0.975	0.95	0.95	0.975	0.945	**0.915**	0.94	0.935	0.92	0.965	**0.915**	0.92	0.97	0.945	**0.43**	0.97	0.92	0.95	0.97	0.965	0.925
		100	0.935	0.935	**0.57**	0.98	0.96	0.945	0.94	**0.915**	0.95	0.965	0.92	0.935	0.93	**0.91**	0.935	0.94	0.945	**0.475**	0.94	0.93	0.95	0.965	0.94	0.92
	50	30	**0.915**	**0.9**	**0.49**	0.985	**0.89**	0.98	0.99	**0.87**	0.98	0.93	**0.82**	0.925	0.94	**0.85**	**0.895**	0.925	**0.915**	**0.305**	0.93	**0.83**	0.965	0.95	**0.875**	**0.875**
		50	0.925	0.92	**0.505**	0.985	0.935	0.95	0.975	**0.905**	0.96	0.95	**0.905**	**0.885**	0.93	**0.865**	**0.885**	0.95	**0.915**	**0.345**	0.94	**0.84**	0.93	0.94	**0.89**	0.92
		80	0.945	0.935	**0.54**	0.99	0.92	0.965	0.985	0.92	0.94	0.965	0.92	0.92	0.93	**0.91**	0.935	0.96	0.96	**0.33**	0.945	**0.885**	0.935	0.97	0.925	**0.895**
		100	0.95	0.945	**0.475**	0.94	0.925	0.955	0.945	**0.88**	0.94	0.97	0.925	**0.905**	0.95	0.92	**0.905**	0.96	0.94	**0.35**	0.97	**0.845**	0.925	0.945	**0.875**	**0.875**
	70	30	0.95	**0.91**	**0.43**	0.985	**0.905**	0.965	1	**0.86**	0.975	0.94	**0.86**	**0.88**	0.94	**0.81**	**0.845**	0.94	0.94	**0.26**	0.945	**0.72**	0.945	0.935	**0.825**	**0.885**
		50	0.955	0.92	**0.345**	0.995	0.92	0.93	0.995	**0.88**	0.96	0.92	**0.89**	**0.885**	0.97	**0.91**	**0.895**	0.945	0.93	**0.28**	0.965	**0.775**	0.95	0.955	**0.865**	**0.895**
		80	0.955	0.93	**0.46**	0.985	0.945	0.965	0.99	**0.915**	0.95	0.975	0.945	0.925	0.965	0.94	0.92	0.96	0.93	**0.32**	0.96	**0.81**	0.96	0.95	**0.905**	**0.84**
		100	0.97	0.965	**0.5**	0.995	0.93	0.93	0.975	0.94	0.94	**0.915**	**0.885**	**0.86**	0.945	**0.905**	**0.895**	0.935	0.945	**0.29**	0.945	**0.78**	0.955	0.945	0.925	**0.875**
	120	30	0.97	0.955	**0.305**	1	**0.91**	0.93	1	**0.8**	0.92	0.98	**0.835**	**0.86**	0.96	**0.76**	**0.81**	0.955	0.93	**0.245**	0.965	**0.65**	0.97	0.975	**0.785**	**0.885**
		50	0.959	0.923	**0.372**	0.995	**0.878**	0.949	1	**0.852**	0.939	0.959	**0.852**	**0.842**	0.959	**0.878**	**0.842**	0.969	0.934	**0.26**	0.959	**0.689**	0.934	**0.913**	**0.801**	**0.827**
		80	0.924	0.924	**0.325**	0.99	**0.888**	0.964	0.995	**0.868**	0.949	0.98	**0.909**	**0.843**	0.964	**0.909**	**0.868**	0.944	0.939	**0.244**	0.97	**0.741**	0.954	0.944	**0.888**	**0.807**
		100	0.923	0.934	**0.281**	1	**0.872**	0.929	0.995	**0.913**	0.934	0.959	**0.918**	**0.847**	0.923	**0.908**	**0.842**	0.939	0.929	**0.23**	0.949	**0.73**	0.959	0.964	**0.893**	**0.821**
60	15	30	**0.91**	**0.915**	**0.74**	0.955	0.985	0.99	0.945	0.97	0.96	**0.91**	0.925	0.92	0.965	0.97	0.94	0.92	0.925	**0.575**	0.965	0.955	0.955	0.95	0.955	0.935
		50	0.945	0.965	**0.77**	0.93	0.96	0.96	0.93	0.93	0.945	0.925	0.94	0.945	0.95	0.94	**0.91**	0.97	0.97	**0.55**	0.955	0.95	0.94	0.94	0.93	0.96
		80	0.94	0.93	**0.735**	0.93	0.95	0.975	0.97	0.97	0.965	0.945	0.965	0.945	0.955	0.945	0.955	0.965	0.955	**0.535**	0.935	0.945	0.94	0.96	0.955	0.95
		100	0.96	0.955	**0.78**	0.975	0.955	0.99	0.95	0.955	0.925	0.955	0.92	0.97	0.94	0.945	0.95	0.955	0.95	**0.565**	0.96	0.95	0.965	0.935	0.94	0.92
	30	30	0.94	0.93	**0.705**	0.94	0.965	0.94	0.925	0.94	0.94	0.945	0.935	0.965	0.945	0.945	**0.915**	**0.91**	**0.9**	**0.475**	0.97	0.96	0.97	0.93	0.935	**0.895**
		50	0.945	0.945	**0.625**	0.93	0.96	0.95	0.945	0.97	0.935	0.96	0.935	0.945	0.975	0.98	0.945	0.935	0.95	**0.445**	0.96	0.94	0.94	0.96	0.96	0.935
		80	0.945	0.955	**0.63**	0.95	0.95	0.935	0.98	0.975	0.975	0.96	0.965	0.96	0.955	0.945	0.95	0.945	0.945	**0.415**	0.965	0.955	0.955	0.955	0.955	0.93
		100	0.965	0.96	**0.66**	0.95	0.97	0.945	0.94	0.94	0.92	0.94	0.95	0.94	0.94	0.965	0.975	0.92	0.93	**0.425**	0.94	0.945	0.945	0.945	0.945	0.92
	50	30	0.945	0.955	**0.52**	0.945	0.975	0.945	0.94	0.965	0.96	0.96	0.97	0.935	0.955	0.95	**0.875**	0.98	0.97	**0.295**	0.965	0.935	0.965	0.93	0.935	0.925
		50	**0.91**	0.935	**0.495**	0.955	0.965	0.96	0.95	0.97	0.95	0.95	**0.91**	0.93	0.935	0.95	0.92	0.945	0.945	**0.35**	0.945	0.955	0.955	0.93	0.945	0.92
		80	0.955	0.945	**0.535**	0.98	0.985	0.96	0.955	0.99	0.955	0.935	0.96	0.94	0.94	0.945	0.93	0.945	0.925	**0.335**	0.955	0.95	0.93	0.96	0.975	**0.91**
		100	**0.915**	0.925	**0.525**	0.94	0.96	0.925	0.965	0.965	0.94	0.94	0.935	0.95	0.94	0.965	**0.9**	**0.91**	0.925	**0.36**	0.93	0.95	0.95	0.97	0.965	**0.915**
	70	30	0.94	0.925	**0.51**	0.975	0.98	0.955	0.965	0.96	0.955	0.95	0.94	**0.915**	0.995	0.96	0.92	0.955	0.95	**0.335**	0.935	**0.91**	0.955	0.95	0.93	**0.89**
		50	0.975	0.97	**0.535**	0.95	0.965	0.925	0.975	0.955	0.96	0.945	0.935	**0.895**	0.95	0.955	**0.905**	0.94	0.945	**0.32**	0.935	**0.91**	0.965	0.965	0.96	0.92
		80	0.945	0.96	**0.455**	0.955	0.965	0.94	0.955	0.94	0.975	0.96	0.93	**0.875**	0.935	0.93	**0.9**	0.935	0.94	**0.285**	0.93	0.93	0.965	0.935	0.92	**0.89**
		100	0.96	0.97	**0.42**	0.935	0.955	0.935	0.94	0.965	0.945	0.93	0.94	0.93	0.935	0.94	**0.88**	0.935	0.94	**0.39**	0.935	0.955	0.955	0.975	0.97	**0.905**
	120	30	0.935	0.935	**0.365**	0.995	**0.91**	0.945	0.995	**0.83**	0.94	0.935	**0.87**	**0.89**	**0.895**	**0.82**	**0.89**	0.955	0.945	**0.275**	0.945	**0.765**	0.95	0.97	**0.855**	**0.885**
		50	0.945	**0.915**	**0.407**	0.995	0.935	**0.915**	0.98	**0.884**	**0.905**	0.935	**0.894**	**0.894**	0.97	0.92	**0.874**	0.94	0.94	**0.256**	0.96	**0.784**	0.95	0.94	**0.844**	**0.819**
		80	0.95	0.955	**0.395**	0.995	0.935	0.92	0.985	0.935	0.935	0.975	0.935	**0.885**	0.935	**0.9**	**0.87**	0.94	0.92	**0.255**	0.94	**0.795**	0.945	0.94	**0.87**	**0.83**
		100	0.945	0.96	**0.345**	0.99	0.985	0.925	0.995	0.96	0.94	0.945	0.935	**0.9**	0.96	**0.9**	**0.885**	**0.91**	0.92	**0.235**	0.96	**0.88**	0.96	0.975	**0.915**	**0.855**

B.5.2 Main results, effect of a

As discussed in Section 3.3, the framework works with any constant $a > 0$. We treated a as a tuning parameter in the main text and used cross-validation to select a in the main simulation study. In this section, we fix a at a constant value for $a \in \{0.01, 1, 10, 50\}$ and investigate the effect of a on the finite sample performance of the proposed method. The rest of the simulation settings are the same as those in the main text.

We illustrate the performance of the proposed method for $m = 30$ in Figure B.3 and Figure B.4. We find that regardless of the value of the constant $a > 0$, the proposed method can provide correct confidence interval coverage and control type-I error (Figure B.3a,b,c,d). However, different choices of a may impact the power of detecting non-zero β_l's (Figure B.3e,f), and the estimation accuracy of β and the variance components (Figure B.4).

B.5.3 Comparison with Two-stage Approaches and Identity Proxy Covariance Matrix

We present additional simulation study results under the same simulation settings as in Section 3.5 of the main paper. Here, we illustrate the power and type-I error rate for the following methods: the two-stage approach with t-test in the second stage (*two-stage t-test*), the two-stage approach with Fisher's p-val aggregation in the second stage (*two-stage Fisher*), the proposed framework (*Proposed*) and its variant with an identity matrix as the proxy covariance matrix (*Proposed-Identity Proxy*). For the two-stage approaches, we use the classic de-biased LASSO method to estimate the fixed effect coefficients in the first step. During the second stage, the individual-level de-biased estimates were collected for the t-test for the *two-stage t-test* approach. The *two-stage Fisher* method employed Fisher's method to aggregate the p-values from testing the individual-level estimates with the classic de-biased LASSO framework. We also compare the 95% confidence interval coverage of the *Proposed-Identity Proxy* method with the *Proposed* method.

Figure B.5 illustrates the power associated with selected methods for testing fixed effect coefficients. We observe no single method consistently exhibits the highest power across all scenarios. Generally, the *Proposed* approach demonstrates superior power compared to

Table B.3: Power for testing non-zero β_l's under the simulation settings in the main text.

			$\beta_1 = 1$ $\psi_1 = 2$		$\beta_2 = 0.5$ $\psi_2 = 0$		$\beta_6 = 0.2$ $\psi_6 = 0$		$\beta_7 = 0.1$ $\psi_7 = 0.1$		$\beta_9 = 0.05$ $\psi_9 = 0.1$	
p	m	n	Proposed	LCL	Proposed	LCL	Proposed	LCL	Proposed	LCL	Proposed	LCL
20	15	30	0.865	0.950	0.975	0.975	0.380	0.340	0.105	0.130	0.090	0.065
		50	0.975	1.000	1.000	1.000	0.700	0.620	0.145	0.170	0.085	0.085
		80	0.995	1.000	1.000	1.000	0.825	0.820	0.275	0.330	0.065	0.065
		100	1.000	1.000	1.000	1.000	0.870	0.880	0.290	0.305	0.145	0.115
	30	30	0.920	0.935	1.000	0.945	0.965	0.820	0.250	0.270	0.105	0.170
		50	1.000	0.995	1.000	0.995	0.995	0.955	0.340	0.360	0.120	0.155
		80	1.000	1.000	1.000	1.000	1.000	0.985	0.515	0.530	0.135	0.135
		100	1.000	1.000	1.000	1.000	1.000	0.985	0.635	0.640	0.160	0.175
	50	30	0.975	0.975	1.000	0.985	1.000	0.900	0.270	0.325	0.095	0.200
		50	1.000	0.990	1.000	0.985	1.000	0.955	0.485	0.440	0.125	0.220
		80	1.000	1.000	1.000	1.000	1.000	0.975	0.720	0.655	0.175	0.255
		100	1.000	1.000	1.000	1.000	1.000	0.985	0.805	0.770	0.250	0.285
	70	30	0.955	0.910	1.000	0.975	1.000	0.880	0.255	0.310	0.105	0.220
		50	0.995	0.995	1.000	0.995	1.000	0.955	0.475	0.470	0.155	0.205
		80	1.000	1.000	1.000	1.000	1.000	0.990	0.720	0.680	0.210	0.290
		100	1.000	1.000	1.000	1.000	1.000	0.995	0.775	0.680	0.230	0.245
	120	30	0.995	0.975	1.000	0.985	1.000	0.925	0.315	0.380	0.120	0.235
		50	0.995	0.995	1.000	0.980	1.000	0.944	0.531	0.500	0.133	0.235
		80	1.000	1.000	1.000	0.995	1.000	0.995	0.761	0.711	0.223	0.259
		100	1.000	1.000	1.000	0.995	1.000	0.985	0.862	0.735	0.250	0.311
60	15	30	0.825	0.870	0.810	0.940	0.235	0.250	0.115	0.110	0.035	0.025
		50	0.990	1.000	0.930	0.990	0.405	0.410	0.115	0.110	0.100	0.090
		80	1.000	1.000	0.995	1.000	0.555	0.645	0.130	0.065	0.065	0.075
		100	1.000	1.000	1.000	1.000	0.645	0.690	0.080	0.065	0.120	0.130
	30	30	0.925	0.950	0.995	1.000	0.490	0.525	0.135	0.100	0.095	0.090
		50	0.985	0.995	1.000	1.000	0.750	0.790	0.125	0.100	0.045	0.060
		80	1.000	1.000	1.000	1.000	0.910	0.935	0.210	0.210	0.155	0.205
		100	1.000	1.000	1.000	1.000	0.950	0.965	0.200	0.230	0.090	0.145
	50	30	0.925	0.960	1.000	1.000	0.685	0.780	0.110	0.110	0.080	0.070
		50	0.990	1.000	1.000	1.000	0.905	0.970	0.180	0.195	0.100	0.125
		80	1.000	1.000	1.000	1.000	0.985	1.000	0.295	0.310	0.110	0.160
		100	1.000	1.000	1.000	1.000	1.000	1.000	0.275	0.370	0.160	0.155
	70	30	0.935	0.945	1.000	0.995	0.920	0.840	0.140	0.150	0.060	0.105
		50	1.000	1.000	1.000	0.995	0.995	0.955	0.285	0.290	0.120	0.140
		80	1.000	1.000	1.000	1.000	1.000	0.990	0.410	0.405	0.130	0.165
		100	1.000	1.000	1.000	1.000	1.000	0.995	0.520	0.455	0.210	0.255
	120	30	0.980	0.965	1.000	0.975	1.000	0.875	0.265	0.290	0.110	0.185
		50	0.995	1.000	1.000	0.995	1.000	0.970	0.482	0.492	0.146	0.146
		80	1.000	1.000	1.000	1.000	1.000	0.990	0.590	0.555	0.215	0.250
		100	1.000	1.000	1.000	1.000	1.000	0.995	0.700	0.680	0.270	0.315

Table B.4: Point estimation accuracy for β, ψ and σ_e^2 under the simulation settings in the main text.

			β total RMSE			ψ total RMSE		σ_e^2 RMSE	
p	m	n	Proposed	LCL	dblasso	Proposed	LCL	Proposed	LCL
20	15	30	0.342	0.510	0.242	0.506	-	0.858	-
		50	0.257	0.403	0.188	0.494	-	0.729	-
		80	0.221	0.356	0.159	0.379	-	0.696	-
		100	0.199	0.321	0.152	0.285	-	0.585	-
	30	30	0.396	0.570	0.173	0.208	0.157	0.381	0.014
		50	0.274	0.441	0.114	0.239	0.141	0.334	0.003
		80	0.216	0.363	0.100	0.189	0.074	0.324	0.002
		100	0.193	0.329	0.099	0.208	0.081	0.318	0.003
	50	30	0.364	0.557	0.113	0.202	0.217	0.263	0.004
		50	0.278	0.459	0.086	0.158	0.161	0.216	0.01
		80	0.224	0.386	0.077	0.117	0.081	0.135	0.001
		100	0.188	0.338	0.066	0.076	0.075	0.164	0.001
	70	30	0.397	0.595	0.105	0.250	0.227	0.143	0.002
		50	0.287	0.484	0.097	0.143	0.14	0.191	0.002
		80	0.202	0.373	0.064	0.113	0.109	0.091	0.004
		100	0.185	0.352	0.058	0.061	0.084	0.066	0.001
	120	30	0.389	0.577	0.077	0.137	0.201	0.061	0.002
		50	0.272	0.485	0.056	0.118	0.181	0.063	0.001
		80	0.204	0.368	0.044	0.058	0.08	0.033	0.002
		100	0.189	0.361	0.048	0.120	0.125	0.040	0.001
60	15	30	0.431	0.575	0.402	0.912	-	1.245	-
		50	0.333	0.464	0.309	0.553	-	0.914	-
		80	0.272	0.384	0.251	0.480	-	0.805	-
		100	0.246	0.345	0.224	0.367	-	0.711	-
	30	30	0.332	0.461	0.295	0.511	-	0.544	-
		50	0.245	0.358	0.216	0.370	-	0.463	-
		80	0.195	0.298	0.176	0.220	-	0.420	-
		100	0.189	0.281	0.164	0.224	-	0.421	-
	50	30	0.278	0.396	0.190	0.297	-	0.380	-
		50	0.226	0.337	0.181	0.161	-	0.361	-
		80	0.166	0.264	0.137	0.176	-	0.268	-
		100	0.167	0.253	0.127	0.194	-	0.260	-
	70	30	0.329	0.487	0.173	0.166	0.203	0.249	0.002
		50	0.264	0.393	0.146	0.164	0.081	0.203	0.006
		80	0.195	0.323	0.123	0.124	0.089	0.160	0
		100	0.187	0.296	0.121	0.103	0.048	0.156	0.007
	120	30	0.361	0.568	0.130	0.164	0.211	0.133	0.005
		50	0.261	0.451	0.106	0.163	0.145	0.111	0.001

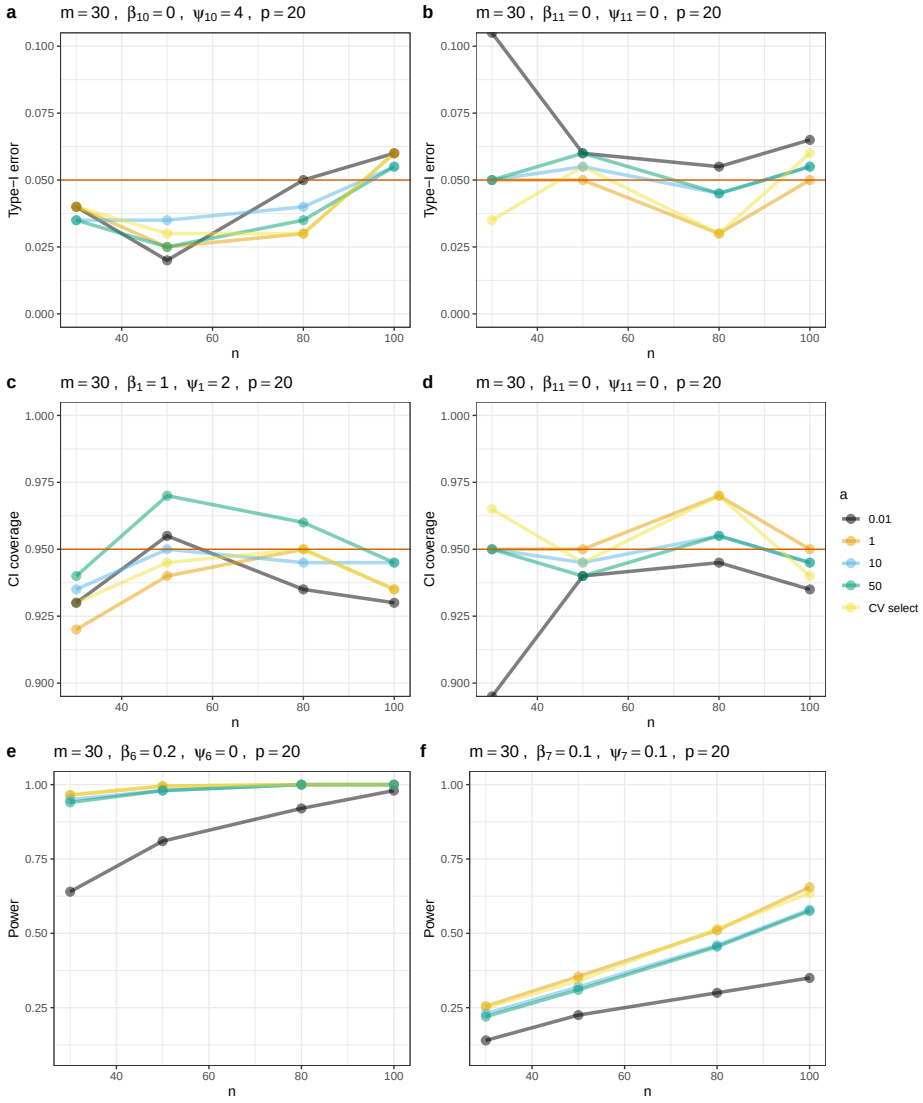

Figure B.3: Simulation results for the proposed method with different values of a, where "a = CV select" corresponds to using cross-validation to choose a for the proposed method. **a**,**b**: Type-I error for testing β_l's at the 0.05 significance level (0.05 marked by red solid line). **c**,**d**: 95% confidence interval coverage (0.95 marked by red solid line) for fixed effect coefficients. **e**,**f**: Power for testing fixed effect coefficients at the 0.05 significance level. All results are computed for $p = 20$ and $m = 30$ based on 200 Monte Carlo simulations, and are plotted separately for each value of a, and against increasing n. The title of each subplot shows the true values of the targeted fixed effect coefficient β_l, and the corresponding random effect variance component ψ_l.

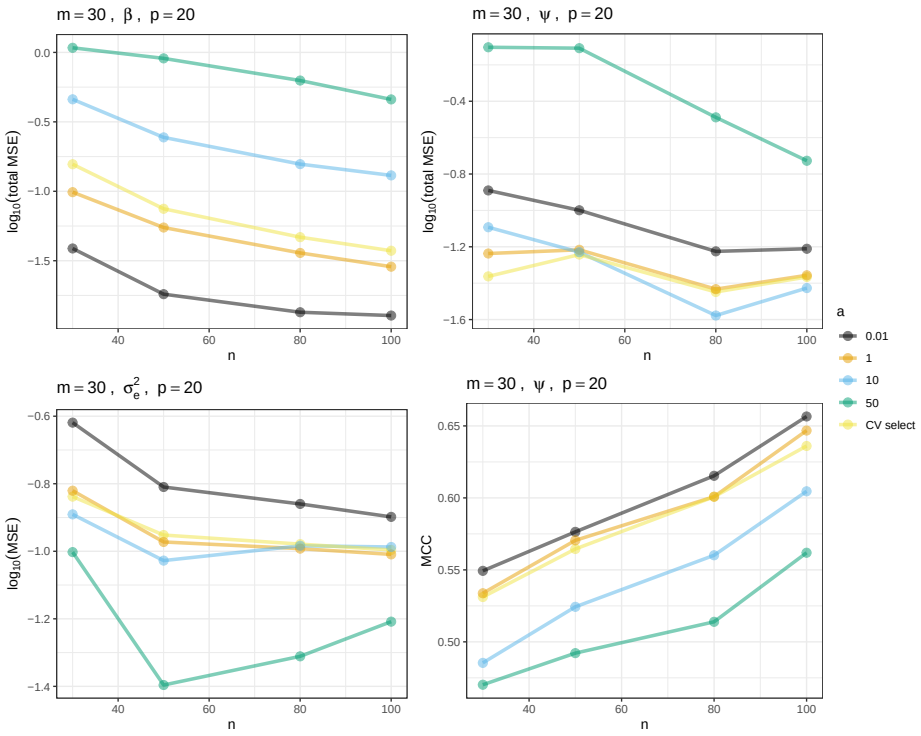

Figure B.4: Simulation results for the proposed method with different values of a, where "a = CV select" corresponds to using cross-validation to choose a for the proposed method. **a**: Total MSE on the $\log_{10}$ scale for estimating the fixed effect coefficients β. **b**: MSE on the $\log_{10}$ scale for estimating the noise variance σ_e^2. **c**: Total MSE on the $\log_{10}$ scale for estimating the random effect variance components ψ. **d**: Average MCC for identifying non-zero variance components, where we use non-zero estimates to select non-zero variance components. Values are computed for $p = 20$ and $m = 30$ based on 200 Monte Carlo simulations, are plotted separately for each a, and are against increasing n.

the *Proposed-Identity Proxy* approach. The *Proposed* method surpasses the *two-stage t-test* method in power in some scenarios, as observed when testing β_6 at p= 20 (Figure B.5c); its power is relatively lower than that of the *two-stage t-test* method in other instances, such as when testing β_6 at p= 60 with m$\geq$ 30 (Figure B.5d). In general, the *two-stage Fisher* approach has the highest power when assessing covariates with non-zero random effects, exemplified in the case of testing β_7 with $\psi_7 = 0.1$ (Figure B.5a, b).

In Figure B.6, we compare the type-I error rates across the four approaches. Both the *two-stage Fisher* and *two-stage t-test* approaches exhibit inflated type-I errors in various scenarios. The *Proposed-Identity Proxy* approach controls the type-I error rate at the nominal level in the majority of scenarios. However, slight inflation of the type-I error is observed in specific cases, such as at $p = 20$ when testing β_{12}, where a type-I error of 0.11 is recorded for $n = 100$ and $m = 30$ (Figure B.6c).

Finally, we examine the 95% confidence interval coverage achieved by the *Proposed* and the *Proposed-Identity Proxy* methods (Figure B.7). The *Proposed-Identity Proxy* approach yields confidence intervals with commendable coverage in the majority of cases. However, a few instances of under-coverage are noted, such as the confidence interval coverage of only 0.89 for β_{12} when $p = 20$, $n = 100$, and $m = 30$ (Figure B.7c); and coverage of 0.885 for β_2 when $p = 60$, $n = 15$, and $m = 70$ (Figure B.7b).

B.6 Extension to High-Dimensional Heterogeneous VAR models

B.6.1 Simulation Study

Simulation Settings

We conduct simulation studies to compare the proposed estimator and inference procedures (referred to as *MEVAR*) with the standard lasso approach ([246], referred to as *db-lasso*). The standard lasso approach ignores the correlations among the observations. We compare the performance in terms of VAR coefficient estimation mean squared error (MSE), 95% confidence interval coverage, type-I error of testing zero coefficients and power of testing non-zero coefficients at 5% significance level.

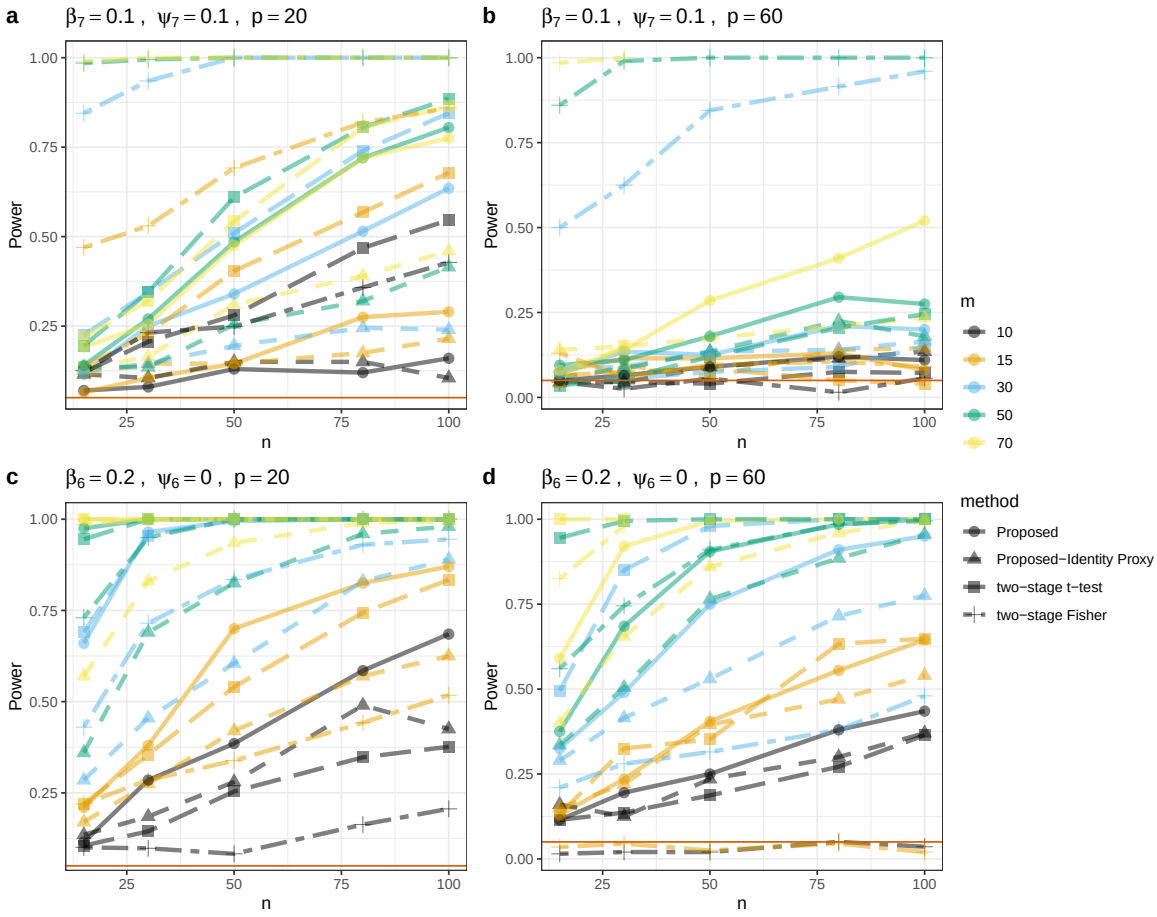

Figure B.5: Simulation results for comparing the power of testing fixed effect coefficients. We compare the *Proposed*, the *Proposed-Identity Proxy*, the *two-stage t-test*, and the *two-stage Fisher* methods. Each subplot presents the power with increasing n at different levels of m. The title of each subplot indicates the values of the tested fixed effect coefficient (β), the corresponding random effect (ψ) and the total number of covariates in the setting (p).

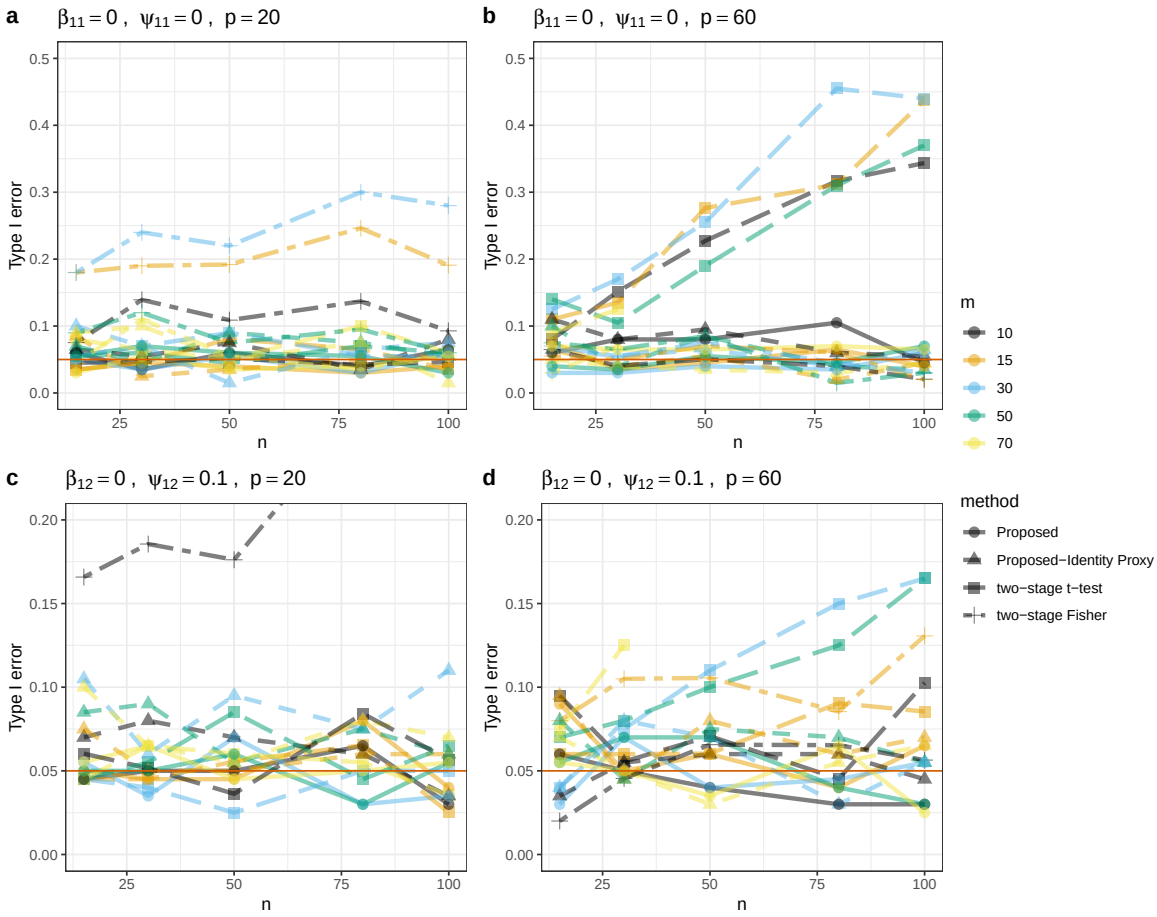

Figure B.6: Simulation results for comparing the type-I error of testing fixed effect coefficients. We compare the *Proposed*, the *Proposed-Identity Proxy*, the *two-stage t-test*, and the *two-stage Fisher* methods. Each subplot presents the type-I error with increasing n at different levels of m. The title of each subplot indicates the values of the tested fixed effect coefficient (β), the corresponding random effect (ψ) and the total number of covariates in the setting (p). We limit the range of the y-axis in subplot c and d to provide a clearer comparison among the *Proposed*, *Proposed-Identity Proxy* and *two-stage t-test* methods.

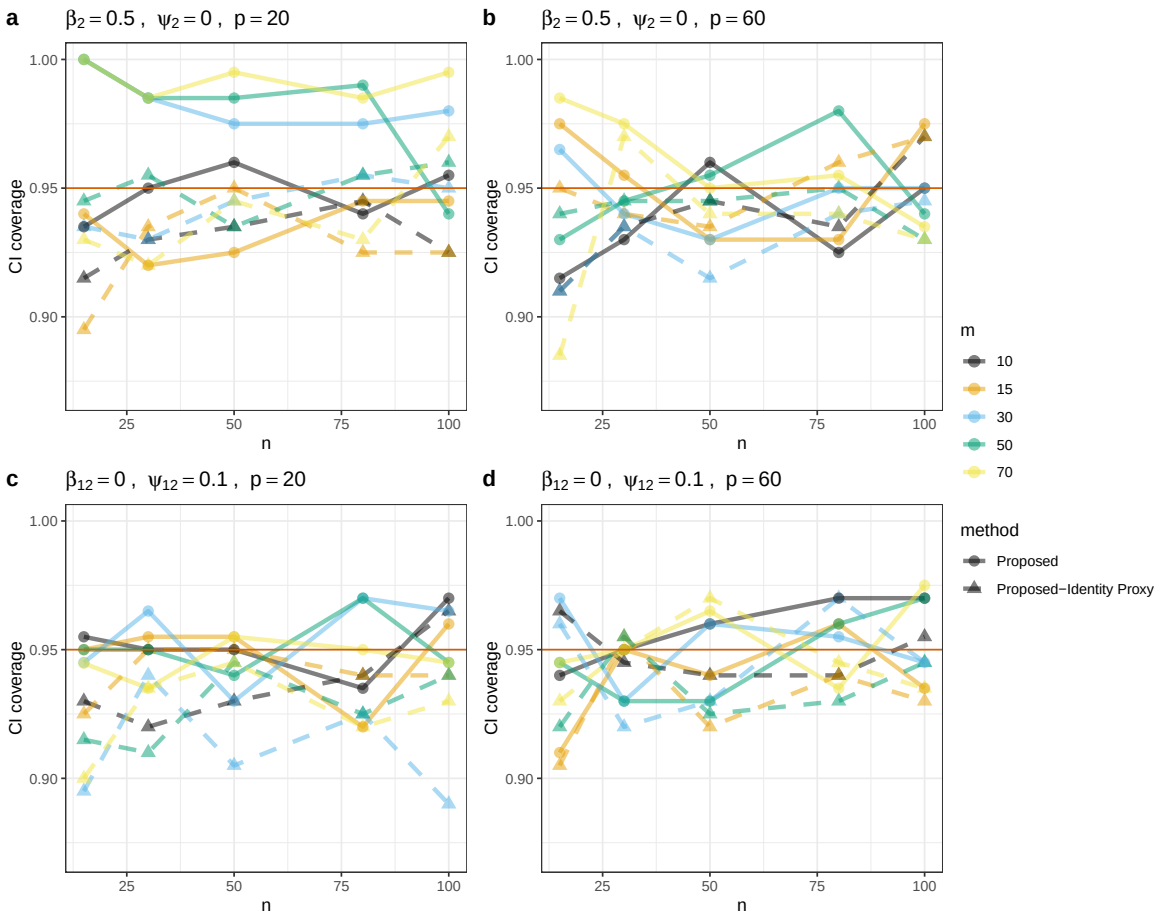

Figure B.7: Simulation results for comparing the 95% confidence interval coverage of the fixed effect coefficients. We compare the *Proposed* and the *Proposed-Identity Proxy* methods. Each subplot presents the confidence interval coverage with increasing n at different levels of m. The title of each subplot indicates the values of the tested fixed effect coefficient (β), the corresponding random effect (ψ) and the total number of covariates in the setting (p). The confidence interval coverage is calculated based on 200 Monte Carlo simulations.

We simulated data from the MEVAR(1) model (3.9) in the main text. We reiterate the model formula below:

$$Y^i(t) = (\Phi + \Gamma^i)Y^i(t-1) + \epsilon^i(t),$$

with $\Phi \in \mathbb{R}^{p\times p}$, $\text{vec}(\Gamma^i) \sim \text{N}(0, \Sigma_\Gamma)$, and $\epsilon^i(t) \sim \text{N}(0, \Sigma_\epsilon)$. We generated diagonal covariance matrices $\Sigma_\Gamma = \text{diag}(\sigma_\Gamma^2)$ and $\Sigma_\epsilon = \text{diag}(\sigma_\epsilon^2)$. The length p^2 vector σ_Γ^2 was sparse with each entry having a 0.1 probability of being non-zero, and the non-zero values were generated independently from $\text{Unif}(0.05, 0.15)$. The length p vector σ_ϵ^2 had a constant value of 0.5 for all entries. We generated a sparse group-level coefficient matrix Φ, with each entry generated independently. The diagonal entries of Φ were generated from $\text{Unif}(0.2, 0.8)$; the off-diagonal entries were generated from a mixture distribution: each off-diagonal entry had a 0.8 probability of taking value 0, and had a 0.2 probability of following $N(0, 0.04)$. We set $p = 30$, $n \in \{20, 40, 60, 80\}$, and $T \in \{25, 50, 100, 150\}$. For each combination of (n, T), we replicated 200 independent Monte Carlo simulations.

We used the same procedure as in the main text to select the tuning parameters, except we followed the cross-validation procedure in [185] for time-series observations. The results are presented based on the optimal values of the tuning parameters. We used the hdi R-package [51] to implement the *db-lasso* approach.

Simulation Results

Without loss of generality, we focus on comparing the results for the first row of the coefficient matrix Φ. Figure B.8 presents the performance of the inference procedures for selected coefficients. The confidence intervals constructed by the *MEVAR* approach have good coverage for all coefficients (Figure B.8a,b). In addition, the *MEVAR* approach always controls the type-I error rate at the nominal level (Figure B.8c,d), and maintains reasonable power for detecting non-zero coefficients (Figure B.8e,f). The standard *db-lasso* approach has inflated type-I error as high as 0.36, and poor confidence interval coverage as low as 0.19 for some of the coefficients (Figure B.8a,d). We noticed that for those coefficients with the corresponding random effect variance being non-zero, i.e., when there is subject-level

heterogeneity in the specific connection, the standard *db-lasso* would fail drastically. While for those coefficients that are fixed across subjects, the proposed *MEVAR* approach and the *db-lasso* approach provide similar results.

In terms of estimation, the proposed *MEVAR* approach consistently estimates the VAR coefficients with decreasing total MSE (Figure B.9), even though its total MSE is slightly higher than the MSE yielded by the *db-lasso* approach.

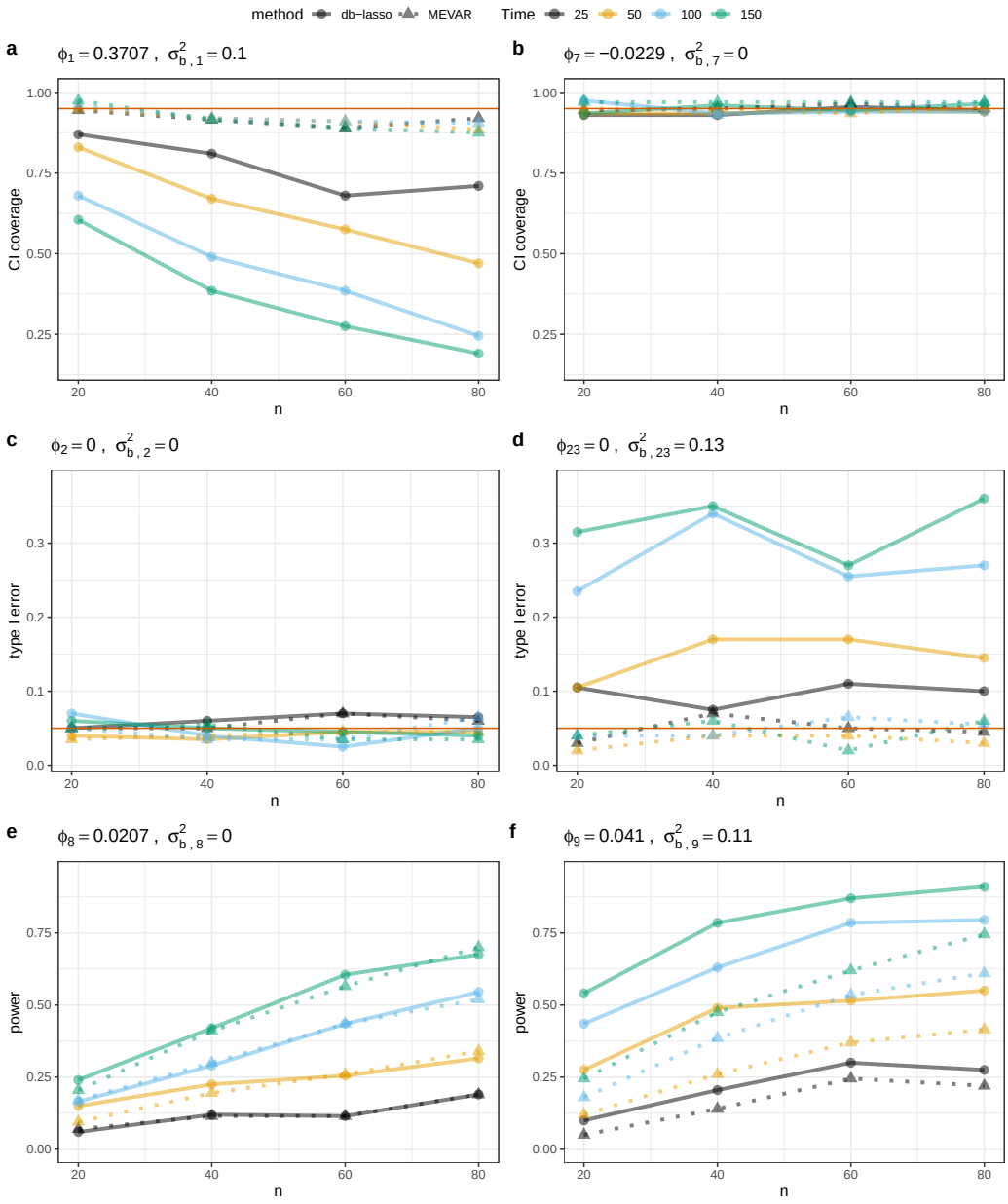

Figure B.8: The performance of the inference procedures by *MEVAR* and *db-lasso* under different values of n and T. **a,b**: the 95% confidence interval coverage for each coefficient (0.95 marked by red solid line); **c,d**: the type-I error of testing the zero coefficient at 5% significance level (0.05 marked by red solid line); **e,f**: the power of testing the non-zero coefficient at 5% significance level. The title of each subplot indicates the true value of the selected coefficient ϕ_j and its corresponding random effect variance $\sigma^2_{b,j}$.

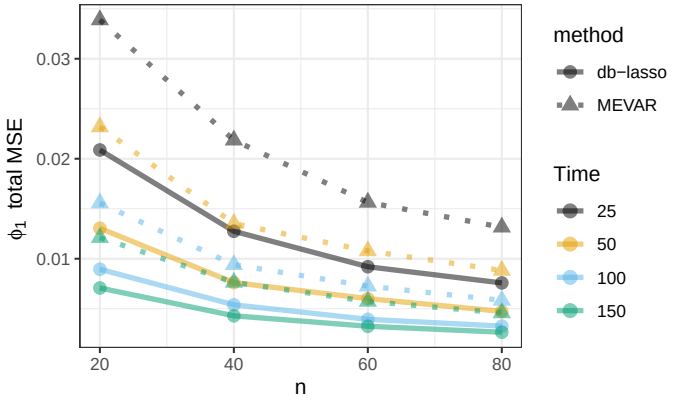

Figure B.9: The total MSE of estimating the first row of Φ, for the *MEVAR* approach and the *db-lasso* approach.

B.6.2 Theoretical Results

Without loss of generality, we establish the results for the first row of the matrix Φ. For notational convenience, we omit the subscripts when the reference to quantities related to the first row of Φ is clear. Denote the first row of Φ by ϕ. We can rewrite the model (3.10) for inferring ϕ as:

$$Y^i = X^i\phi + X^i\gamma^i + \epsilon^i, \quad i = 1, \ldots, n, \tag{B.76}$$

where Y^i, X^i, γ^i are sub-matrices/sub-vectors of $\tilde{Y}^i$, $\tilde{X}^i$ and $\text{vec}(\Gamma^i)$ that corresponds to inferring ϕ. Here, $\gamma^i \sim \text{N}(0, \Sigma_\gamma)$ and $\epsilon^i \sim \text{N}(0, \Sigma_e)$, with $\Sigma_\gamma = \text{diag}(\sigma^2_{\Gamma,1:p})$ and $\Sigma_e = \text{diag}(\sigma^2_{\epsilon,1:p})$. We allow for either $p > c_0 T$ or $T > c_0 p$ for some constant $c_0 > 1$.

We first state the pivotal lemma that connects the singular values of X^i to the singular values of a standard Gaussian matrix. Suppose $\sigma(X)$ denotes a non-zero singular value of the matrix X. We write $A \prec B$ if $B - A$ is a positive semi-definite matrix.

Lemma B.6.1. *Suppose $X \in \mathbb{R}^{T\times p}$ $(T \neq p)$ is a Gaussian matrix with* $\text{vec}(X) \sim \text{N}(0, \Xi)$.

Define $Z \in \mathbb{R}^{T\times p}$ *such that* $\operatorname{vec}(Z) = (\Xi)^{-1/2}\operatorname{vec}(X) \sim \mathrm{N}(0, I_{Tp})$. *We have that*

$$\sqrt{\sigma_{\min}(\Xi)}\sigma_{\min}(Z) \leq \sigma(X) \leq \sqrt{\sigma_{\max}(\Xi)}\sigma_{\max}(Z).$$

We also have

$$X^\top(aXX^\top + I)^{-1}X \succ \sigma_{\min}(\Xi)Z^\top(a\sigma_{\max}(\Xi)ZZ^\top + I)^{-1}Z$$
$$X^\top(aXX^\top + I)^{-1}X \prec \sigma_{\max}(\Xi)Z^\top(a\sigma_{\min}(\Xi)ZZ^\top + I)^{-1}Z.$$

Then in order to make the proposed doubly high-dimensional LMM framework work for the MEVAR(1) case, a sufficient condition is $\sigma(\Xi^i) \asymp 1$ where $\Xi^i = \operatorname{Var}(X^i)$. This is in fact a mild condition under the following assumption:

Assumption 14. *For the ith subject, conditioning on* Γ^i*, the observations* $\{Y^i(t)\}_{t=1}^T$ *are realizations of a stationary Gaussian process.*

Under Assumption 14, Proposition 2.3 of [11] bounds $\sigma(\Xi^i)$ by the extreme eigenvalues of the matrix-valued spectral density function over the unit circle. According to Lemma 5.5 of [252], we have $\sigma(\Xi^i) \asymp 1$ when $\|\Phi + \Gamma^i\|_2 \leq 1 - \Delta$ holds for some $\Delta \in (0, 1)$, which is a common condition for VAR process [252, 161].

The rest of the proof directly follows the proof for inferring the fixed effect coefficients in a standard doubly high-dimensional LMM.

Proof for Lemma B.6.1. We first work on the case when $T > p$:

Since the set of singular values $\{\sigma(X)\}$ is identical to the set $\left\{\sqrt{\sigma(X^\top X)}\right\}$, we have $\min_{\nu\in\mathbb{R}^p}\|X\nu\|_2^2 \leq \sigma_{\min}^2(X) \leq \sigma_{\max}^2(X) \leq \max_{\nu\in\mathbb{R}^p}\|X\nu\|_2^2$ for $\|\nu\|_2^2 = 1$. We can also show

that:

$$\begin{aligned}\|X\nu\|_2^2 &= \left\|\left(\nu^\top \otimes I_T\right)\operatorname{vec}(X)\right\|_2^2 \\ &= \left\|\left(\nu^\top \otimes I_T\right)\Xi^{1/2}\operatorname{vec}(Z)\right\|_2^2 \\ &\in \left[\sigma_{\min}(\Xi)\left\|\left(\nu^\top \otimes I_T\right)\operatorname{vec}(Z)\right\|_2^2, \sigma_{\max}(\Xi)\left\|\left(\nu^\top \otimes I_T\right)\operatorname{vec}(Z)\right\|_2^2\right]\end{aligned}$$

Thus with $\left\|\left(\nu^\top \otimes I_T\right)\operatorname{vec}(Z)\right\|_2^2 = \|Z\nu\|_2^2 \in [\sigma_{\min}^2(Z), \sigma_{\max}^2(Z)]$, we obtain

$$\sigma_{\min}(\Xi)\sigma_{\min}^2(Z) \leq \sigma_{\min}^2(X) \leq \sigma_{\max}^2(X) \leq \sigma_{\max}(\Xi)\sigma_{\max}^2(Z).$$

We next look at the case when $T < p$. Note that we now have $\{\sigma(X)\} = \{\sigma(X^\top)\} = \left\{\sqrt{\sigma(XX^\top)}\right\}$. Since the covariance matrix $\operatorname{Var}(\operatorname{vec}(X^\top))$ can be obtained by imposing the same permutations on the columns and the rows of the matrix $\Xi = \operatorname{Var}(\operatorname{vec}(X))$, the two covariance matrices are similar and have the same set of singular values. Therefore, we can just follow the same arguments used for the case of $T > p$ to establish the bounds for $\sigma(X)$ under the setting of $T < p$.

Based on the above bounds for $\sigma(X)$, we can get $XX^\top \succ \sigma_{\min}(\Xi)ZZ^\top$ since

$$\begin{aligned}\forall \nu \in \mathbb{R}^T, \quad & \nu^\top XX^\top \nu - \nu^\top \sigma_{\min}(\Xi)ZZ^\top \nu \\ &= \|(\nu^\top \otimes I_p)\operatorname{vec}(X^\top)\|_2^2 - \sigma_{\min}(\Xi)\|(\nu^\top \otimes I_p)\operatorname{vec}(Z^\top)\|_2^2 \\ &= \|(\nu^\top \otimes I_p)\Xi^{1/2}\operatorname{vec}(Z^\top)\|_2^2 - \sigma_{\min}(\Xi)\|(\nu^\top \otimes I_p)\operatorname{vec}(Z^\top)\|_2^2 \\ &\geq 0.\end{aligned}$$

We can similarly show that $XX^\top \prec \sigma_{\max}(\Xi)ZZ^\top$. Given two positive definite matrices A and B, if $A \succ B$ then $B^{-1} \succ A^{-1}$. We thus have that

$$(a\sigma_{\max}(\Xi)ZZ^\top + I_T)^{-1} \prec (aXX^\top + I_T)^{-1} \prec (a\sigma_{\min}(\Xi)ZZ^\top + I_T)^{-1}. \tag{B.77}$$

We then show that for a positive definite matrix $A \in \mathbb{R}^{T\times T}$, $\sigma_{\min}(\Xi)Z^\top AZ \prec X^\top AX \prec$

$\sigma_{\max}(\Xi)Z^\top AZ$: for the upper bound of $X^\top AX$, we have that

$$\begin{aligned}\forall \nu \in \mathbb{R}^T, \quad & \nu^\top \sigma_{\max}(\Xi) Z^\top AZ\nu - \nu^\top X^\top AX\nu \\ & = \sigma_{\max}(\Xi)\|A^{1/2}(\nu^\top \otimes I_T)\operatorname{vec}(Z)\|_2^2 - \|A^{1/2}(\nu^\top \otimes I_T)\Xi^{1/2}\operatorname{vec}(Z)\|_2^2 \\ & \geq 0.\end{aligned}$$

We can follow similar arguments to obtain the lower bound for $X^\top AX$.

Therefore, combining the bounds $\sigma_{\min}(\Xi)Z^\top AZ \prec X^\top AX \prec \sigma_{\max}(\Xi)Z^\top AZ$ and the bounds in (B.77), we obtain

$$\begin{aligned}X^\top(aXX^\top + I)^{-1}X &\succ \sigma_{\min}(\Xi)Z^\top(a\sigma_{\max}(\Xi)ZZ^\top + I)^{-1}Z \\ X^\top(aXX^\top + I)^{-1}X &\prec \sigma_{\max}(\Xi)Z^\top(a\sigma_{\min}(\Xi)ZZ^\top + I)^{-1}Z.\end{aligned}$$

□