Variance Components Estimation under High-dimensional Linear Models

By

XIAOHAN HU
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Applied Mathematics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Xiaodong Li, Chair

Thomas Strohmer

Debashis Paul

Committee in Charge

2025

To my family.

Contents

Abstract

Estimation of signal-to-noise ratios (SNR) and residual variances in high-dimensional linear models has numerous important applications, including selecting tuning parameters for predictive models and estimating heritability in genomics. This dissertation investigates the consistency and asymptotic distribution of two widely used SNR estimators under various model assumptions.

Chapter 2 presents my work [18], supervised by Professor Xiaodong Li, in which we study the restricted maximum likelihood (REML) estimator based on the likelihood formulation of the random effects model. Although the true model assumes an i.i.d. Gaussian priors for both the regression coefficients and the noise variables, we establish consistency and asymptotic normality of the REML estimator under model misspecification, where the true coefficient vector is fixed and noise components may be heterogeneous. In particular, the resulting asymptotic variance has a tractable form, allowing standard error estimation via a measure of noise heterogeneity.

Chapter 3 discusses my joint work [28] with Zhentao Li and Professor Xiaodong Li. While the method-of-moments estimator is commonly used in SNR estimation in single-response settings, we extend this framework to multivariate linear models under both fixed and random effects formulations. In this study, we establish and compare the asymptotic distributions of the proposed estimators. Furthermore, we extend our approach to accommodate cases with residual heteroskedasticity and derive asymptotic inference procedures based on standard error estimation.

In both Chapter 2 and Chapter 3, we validate our theoretical results through extensive numerical simulations. Further discussions and directions for future work are provided in Chapter 4.

Acknowledgments

First and foremost, I would like to express my sincere gratitude to my advisor, Professor Xiaodong Li, for his continuous help and support throughout my PhD program. I am thankful for his extensive knowledge, excellent intuition, and patience. I would also like to thank Professor Thomas Strohmer and Professor Debashis Paul for serving on my dissertation committee, and Professor Qinlan Xia and Professor Krishnakumar Balasubramanian for attending my qualifying exam.

For my first project [18], I would like to acknowledge the support from the NSF via the CAREER Award DMS-1848575 and Professor Debashis Paul for inspiring discussions. For my second project [28], I would like to thank my collaborator Zhentao Li; it was truly a rewarding experience to work with him.

I would also like to thank my department for their dedicated and responsible work attitude. I am grateful to all the senior PhD students I met during my program, whose selfless help meant a great deal to me.

Finally, I owe the most to my family, especially my parents, Xuewen Hu and Fang Zhao. Their love and encouragement have supported me throughout my entire life.

CHAPTER 1

Introduction

1.1. Overview

Estimation and inference for signal-to-noise ratios (SNR) and residual variances in high-dimensional linear models are fundamental problems in statistics, with wide-ranging applications. A prominent application is heritability estimation in genome-wide association studies (GWAS), where the goal is to quantify the proportion of phenotypic variance that can be attributed to genetic variation [11]. Another important application is regarding how to select tuning parameters in regularized regression such as Lasso and Ridge regression ([7, 8, 10, 20, 37]).

Significant theoretical and methodological developments have been made for SNR estimation under both fixed effects models [7, 8, 18, 20, 35, 41] and random effects models [9, 13, 24, 29, 31, 43]. Among the various approaches proposed for SNR estimation and inference, two of the most widely used are the method-of-moments (MM) estimators [7, 13, 31, 41] and likelihood-based estimators under random effects models [6, 15, 29, 36, 43, 45]. In addition to these, other methods have been developed, including the EigenPrism procedure [20], and approaches based on sparsity and penalized regression such as the Lasso [3, 12, 37].

This dissertation consists of two main contributions. First, we investigate the asymptotic behavior of restricted maximum likelihood (REML) estimators under model misspecification. Second, motivated by the fact that most existing work on SNR estimation focuses on univariate responses, we propose a definition of multivariate SNR and show how to perform inference for multivariate SNR using method-of-moments estimators under both fixed effects and random effects models.

1.2. Maximum Likelihood Estimation: Analysis under High-Dimensional Linear Fixed-effects Model

Asymptotic analysis of REML under linear mixed effects models is a well-established topic in the statistical literature; see, for example, [16], [21], [32], and [22]. In recent years, a growing body of work has focused on the behavior of random effects likelihood estimators under model misspecification—specifically, when the true coefficient vector does not follow the commonly assumed i.i.d. Gaussian distribution. As early as [21], it was shown that Gaussian random effects likelihood estimators can be consistent and asymptotically normal even when the true coefficients are i.i.d. but non-Gaussian.

A particularly notable contribution is [24], which demonstrates that such estimators retain consistency and asymptotic normality even when the true model is under sparse random effects. The analysis of REML under model misspecification has also been extended to the case where the coefficient vector is fixed but arbitrary, provided the design matrix has i.i.d. Gaussian entries [8]. That line of analysis relies critically on the rotational invariance of the Gaussian design and utilizes normal approximation tools developed in [9].

In this work, we further investigate the consistency and asymptotic distribution of the SNR estimator under significant model misspecification. Specifically, we consider high-dimensional linear models with heteroscedastic and correlated noise, where the true coefficient vector $\boldsymbol{\beta}$ is fixed and deviates substantially from the assumed i.i.d. Gaussian prior.

Our main results show that the asymptotic variance of $\sqrt{n}\hat{\gamma}$ depends only on the aspect ratio $1/\tau$, the true SNR γ_0, and a parameter κ that captures both the heterogeneity and correlation structure of the noise. The resulting expression for the asymptotic variance is sufficiently tractable to allow estimation of the standard error via consistent estimation of the noise heterogeneity.

In the final section of this chapter, we outline an approach for extending the REML estimator to the group-wise setting, motivated by the problem of partitioning heritability as discussed in [29]. The task of estimating group-specific SNRs is also closely related to group-regularized ridge regression, as explored in [19]. However, due to technical challenges in extending the normal approximation results from [4], a complete asymptotic analysis of the group SNR estimator $\hat{\gamma}$ is left for future work.

The main context of this chapter is adapted from my joint work with Professor Xiaodong Li and the preprint was posted on ArXiv [18].

1.3. Method of Moments Estimation: Analysis under Multivariate High-Dimensional Linear Models

The goal of this section is to develop inference procedures for the signal-to-noise ratio (SNR) under both multivariate fixed effects and random effects models. While SNR estimation and inference have been extensively studied in the univariate setting, the multivariate case remains comparatively underexplored. Existing works in the multivariate domain are often application-driven. For instance, [13] proposed a definition of multivariate heritability based on a multivariate random effects model, developed a method-of-moments (MM) estimator for it, and applied their approach to estimate the heritability of brain shape using MRI data. Our work builds on the framework introduced in [13], but with substantially broader scope. We not only extend the analysis to fixed effects models—an area that, to our knowledge, has not been studied in the high-dimensional multivariate SNR estimation—but also consider general noise structures in the random effects case, where asymptotic analysis is significantly more challenging. We note that alternative definitions of multivariate SNR have been proposed in the literature (e.g., [47]); however, our formulation for the multivariate fixed effects model is novel in the literature to our knowledge.

The MM estimator we develop under the fixed effects model can be viewed as a extension in multivariate case of the estimator in univariate proposed by [7]. For the random effects model, our approach is most related to that of [13], but their analysis is limited to standard Gaussian noise. We emphasize that these extensions are far from straightforward. In this work, we rigorously establish the asymptotic distributions of the proposed SNR estimators and demonstrate how they can be used to perform valid statistical inference in both model settings.

This chapter is adapted from my joint work with Zhentao Li and Professor Xiaodong Li. The corresponding preprint was posted on ArXiv [28].

CHAPTER 2

Maximum Likelihood Estimation: Analysis under High-Dimensional Linear Fixed-effects Model

2.1. Problem Statement and Method

We focus on the following high-dimensional linear model in this chapter:

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.1}$$

where $\boldsymbol{Z}$ is an $n \times p$ design matrix with p being allowed to be greater than n, $\boldsymbol{\beta}$ is the vector of regression coefficients, and $\boldsymbol{y}$ is the response vector. For the noise vector $\boldsymbol{\varepsilon}$, we assume it satisfies $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$, where $\boldsymbol{\Sigma}_\varepsilon$ is a positive definite matrix. This implies that we allow for correlated and heteroscedastic noise in the linear model. In particular, we denote the diagonal entries of $\boldsymbol{\Sigma}_\varepsilon$ as $\sigma_1^2, \ldots, \sigma_n^2$. Also denote the average noise level as $\sigma_0^2 = \frac{1}{n}\sum_{i=1}^n \sigma_i^2$. Our goal is to make inference about the signal-to-noise ratio (SNR) parameter

$$\gamma_0 \coloneqq \|\boldsymbol{\beta}\|^2/\sigma_0^2.$$

2.1.1. REML Based on Homogeneous and Gaussian Random Effects. In this chapter, we are interested in the SNR estimator based on the likelihood of the Gaussian random effects model, in which the coefficient vector is modeled as $p^{-1/2}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is assumed to consist of i.i.d. $\mathcal{N}(0, \sigma_\alpha^2)$ variables. In addition, the noise terms are assumed to be independent and follow the same distribution $\mathcal{N}(0, \sigma_\varepsilon^2)$. Comparing the true model and the postulated model, it is clear that σ_0^2 corresponds to σ_ε^2, $\|\boldsymbol{\beta}\|^2$ corresponds to σ_α^2, and $\gamma_0 = \|\boldsymbol{\beta}\|^2/\sigma_0^2$ corresponds to $\gamma \coloneqq \sigma_\alpha^2/\sigma_\varepsilon^2$. Based on this postulated homogeneous and Gaussian random effects model, REML estimation, i.e. maximum likelihood estimation, can be derived for the variance components σ_α^2 and σ_ε^2 [22]. In fact, under the above Gaussian random effects model, there holds $\boldsymbol{y} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Omega})$, where

$$\boldsymbol{\Omega} = \boldsymbol{\Omega}(\sigma_\varepsilon^2, \sigma_\alpha^2) \coloneqq \sigma_\varepsilon^2 \boldsymbol{I}_n + \frac{\sigma_\alpha^2}{p} \boldsymbol{Z}\boldsymbol{Z}^\top \coloneqq \sigma_\varepsilon^2 \boldsymbol{V}_\gamma,$$

and

$$\boldsymbol{V}_\gamma = \boldsymbol{I}_n + \frac{\gamma}{p} \boldsymbol{Z}\boldsymbol{Z}^\top. \tag{2.2}$$

Then, the log-likelihood function for $(\sigma_\varepsilon^2, \sigma_\alpha^2)$ is given as below:

$$l(\sigma_\varepsilon^2, \sigma_\alpha^2) = c - \frac{1}{2} \log \det(\boldsymbol{\Omega}) - \frac{1}{2} \boldsymbol{y}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{y},$$

where c is a constant. By taking the partial derivatives of the log-likelihood with respect to σ_ε^2 and σ_α^2 to obtain the score functions, we got the following likelihood equations:

$$\begin{cases} S_{\sigma_\varepsilon^2}(\sigma_\varepsilon^2, \sigma_\alpha^2) \coloneqq \frac{1}{2} \boldsymbol{y}^\top \boldsymbol{\Omega}^{-2} \boldsymbol{y} - \frac{1}{2} \operatorname{trace}(\boldsymbol{\Omega}^{-1}) = 0 \\ S_{\sigma_\alpha^2}(\sigma_\varepsilon^2, \sigma_\alpha^2) \coloneqq \frac{1}{2} \boldsymbol{y}^\top \boldsymbol{\Omega}^{-1} \frac{1}{p} \boldsymbol{Z}\boldsymbol{Z}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{y} - \frac{1}{2} \operatorname{trace}\left(\boldsymbol{\Omega}^{-1} \frac{1}{p} \boldsymbol{Z}\boldsymbol{Z}^\top\right) = 0. \end{cases}$$

By the fact that $\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top = \frac{1}{\gamma}(\boldsymbol{V}_\gamma - \boldsymbol{I}_n)$, the above set of equations can yield a single equation about the SNR $\gamma = \sigma_\alpha^2/\sigma_\varepsilon^2$:

$$\Delta(\gamma) \coloneqq \boldsymbol{y}^\top \boldsymbol{B}_\gamma \boldsymbol{y} = 0. \tag{2.3}$$

where

$$\boldsymbol{B}_\gamma = \frac{\boldsymbol{V}_\gamma^{-1}}{n} - \frac{\boldsymbol{V}_\gamma^{-2}}{\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})}. \tag{2.4}$$

Let $\hat{\gamma}$ be a solution to (2.3), which is referred to the (misspecified) REML estimator of the true SNR $\gamma_0 = \|\boldsymbol{\beta}\|^2/\sigma_0^2$.

2.1.2. Misspecification Analysis of REML. We study the consistency and asymptotic distribution of $\hat{\gamma}$ when the Gaussian random-effects model is significantly misspecified.

In particular, the actual coefficient vector $\boldsymbol{\beta}$ is a general fixed one, and the noise $\boldsymbol{\varepsilon}$ is both heteroskedastic and correlated. Naturally, there is a trade-off between the misspecification of $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$, and the assumptions placed on the design matrix $\mathbf{Z}$.

Our main results, presented in the next section, show that the consistency and asymptotic distribution of $\hat{\gamma}$ can be rigorously established as long as the entries of $\mathbf{Z}$ are independent, symmetric, and sub-Gaussian standardized random variables. The skew-free assumption is imposed primarily for technical reasons, and we will employ numerical simulations indicating that it may be relaxed. All analyses are conducted under the asymptotically proportional setting $n, p \to \infty$ with $n/p \to \tau > 0$, where $1/\tau$ is commonly referred to as the limiting aspect ratio.

In our main results in Section 2.2, we will show that the asymptotic variance of $\sqrt{n}\hat{\gamma}$ depends only on the aspect ratio $1/\tau$, the true SNR γ_0, and a parameter κ that characterizes both the heterogeneity and correlation of the noise terms. To estimate this variance and thereby make inferences about the true SNR γ_0, we also need to estimate the average noise level σ_0^2 and the parameter κ. In fact, using the SNR estimate $\hat{\gamma}$ from the postulated Gaussian random-effects model, we can then estimate σ_0^2 through

$$\hat{\sigma}^2 = \frac{1}{n}\boldsymbol{y}^\top \boldsymbol{V}_{\hat{\gamma}}^{-1}\boldsymbol{y}. \tag{2.5}$$

One intuition of this estimator is the following identity based on the postulated (and misspecified) Gaussian and homogeneous random effects model

$$\mathbb{E}[\boldsymbol{y}^\top \boldsymbol{V}_\gamma^{-1}\boldsymbol{y}] = \mathbb{E}[\boldsymbol{V}_\gamma^{-1}\boldsymbol{y}\boldsymbol{y}^\top] = \mathbb{E}[\boldsymbol{V}_\gamma^{-1}\boldsymbol{\Omega}] = n\sigma_\varepsilon^2.$$

In heteroscedastic case the estimation of κ is in general difficult under the case of correlated noise. However, when the noise is heterogeneous but uncorrelated, there is a natural approach to estimating κ. In this case, κ can be simply referred to as the *heterogeneity parameter*,

since

$$\kappa = \frac{1}{n\sigma_0^4}\|\mathbf{\Sigma}_\varepsilon\|_F^2 = \frac{1}{n\sigma_0^4}\sum_{i=1}^{n}\sigma_i^4. \tag{2.6}$$

Under our assumptions on the design matrix, it is easy to get

$$\begin{aligned}\mathbb{E}[y_i^4] &= \sum_{j=1}^{p}\left(\mathbb{E}[z_{ij}^4]-3\right)\beta_j^4 + 3\|\boldsymbol{\beta}\|_2^4 + 6\|\boldsymbol{\beta}\|_2^2\sigma_i^2 + 3\sigma_i^4 \\ &\approx 3\|\boldsymbol{\beta}\|_2^4 + 6\|\boldsymbol{\beta}\|_2^2\sigma_i^2 + 3\sigma_i^4,\end{aligned}$$

which implies $(1/n)\sum_{i=1}^{n}\mathbb{E}[y_i^4] \approx 3\|\boldsymbol{\beta}\|_2^4 + 6\|\boldsymbol{\beta}\|_2^2\sigma_0^2 + 3\kappa\sigma_0^4$. By this heuristic, we give the following estimate for the heterogeneity parameter

$$\hat{\kappa} := \frac{1}{3n\hat{\sigma}^4}\sum_{i=1}^{n}y_i^4 - (\hat{\gamma}_n^2 + 2\hat{\gamma}_n). \tag{2.7}$$

We also show the consistency of $\hat{\sigma}^2$ and $\hat{\kappa}$ Section 2.2.

2.2. Main Results

We first introduce our result on the consistency of $\hat{\gamma}$ and $\hat{\sigma}^2$:

THEOREM 2.2.1. *Consider the linear model (2.1) with the asymptotic setting $n, p \to \infty$ such that $\sqrt{n}\left|\frac{n}{p}-\tau\right| \to 0$, where $\tau > 0$ is a fixed constant. Assume that the entries of the design matrix $\boldsymbol{Z}$ are independent, symmetric, sub-Gaussian, and unit-variance random variables, and their maximum sub-Gaussian norm is uniformly upper bounded by some numerical constant C_0. Let $\boldsymbol{\varepsilon}$ be the vector of correlated and heteroscedastic Gaussian noise: $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{\Sigma}_\varepsilon)$ with variances (diagonal entries) $\sigma_1^2, \ldots, \sigma_n^2$, so that*

(1) $\max_{i\in[n]}\sigma_i^2$ *is uniformly bounded by* C_0;

(2) $\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2 = \sigma_0^2$, *where* σ_0^2 *is set to be fixed for all* n;

(3) $\|\mathbf{\Sigma}_\epsilon\|_F = o(n)$.

Let $\boldsymbol{\beta}$ be the coefficient vector with fixed two-norm $\|\boldsymbol{\beta}\|^2 > 0$ for all n, which implies the SNR $\gamma_0 := \|\boldsymbol{\beta}\|^2/\sigma_0^2$ is fixed for all n.

Under the above conditions, there is a sequence of estimates $\hat{\gamma}_n$ *as solutions to* (2.3) *satisfying* $\hat{\gamma}_n \xrightarrow{P} \gamma_0$ *as* $n \to \infty$. *Moreover, the corresponding sequence of noise variance estimate in* (2.5) *satisfies* $\hat{\sigma}^2 \xrightarrow{P} \sigma_0^2$.

Before we state our next result regarding the asymptotic distribution of $\hat{\gamma}$, we need to introduce the following probability density function of the Marčenko-Pastur law with the parameter $\tau > 0$:

$$f_\tau(x) = \frac{1}{2\pi\tau x}\sqrt{(b_+(\tau) - x)(x - b_-(\tau))}1_{\{b_-(\tau)\leq x\leq b_+(\tau)\}},$$

where $b_\pm(\tau) = (1 \pm \sqrt{\tau})^2$. Note that the Marčenko-Pastur law also has a point mass $1 - \tau^{-1}$ at the origin when $\tau > 1$. With $f_\tau(x)$, and any $\tau, \gamma > 0$, we define the following quantities based on the Marčenko-Pastur law for any positive integer k:

$$h_k(\gamma, \tau) = \int_{b_-(\tau)}^{b_+(\tau)} \frac{1}{(1+\gamma x)^k} f_\tau(x) + \left(1 - \frac{1}{\tau}\right) 1_{\{\tau>1\}}. \tag{2.8}$$

With these quantities determined as integrals based on the Marčenko-Pastur law, we are able to obtain the following result on the asymptotic distribution of $\hat{\gamma}$ by imposing additional assumptions on the infinity norm of $\boldsymbol{\beta}$, and both the Frobenius and operator norms of $\boldsymbol{\Sigma}_\varepsilon$:

THEOREM 2.2.2. *In addition to the assumptions in Theorem 2.2.1, we further assume* $\|\boldsymbol{\beta}\|_\infty = o(p^{-1/4})$. *For the noise* $\boldsymbol{\varepsilon}$, *we make the following additional assumptions on its covariance matrix:*

(1) $\|\boldsymbol{\Sigma}_\varepsilon\|$ *is uniformly bounded;*

(2) $\kappa = \frac{1}{n\sigma_0^4}\|\boldsymbol{\Sigma}_\varepsilon\|_F^2$ *is fixed for all* n.

Then, with $h_k(\gamma_0, \tau)$ *as in* (2.8), *as* $n \to \infty$,

$$\sqrt{n}\,(\hat{\gamma} - \gamma_0) \Longrightarrow \mathcal{N}\left(0, 2\gamma_0^2\left(\frac{1}{h_2(\gamma_0,\tau) - h_1^2(\gamma_0,\tau)} + \kappa - \tau - 1\right)\right). \tag{2.9}$$

Note that the asymptotic variance of $\hat{\gamma}$ given in Theorem 2.2.2 relies solely on the limiting aspect ratio $1/\tau$, true SNR γ_0, and the parameter κ that is determined by the correlation and heterogeneity of the noise $\boldsymbol{\varepsilon}$. γ_0 can be consistently estimated by $\hat{\gamma}$ and the following result gurantees the consistency of $\hat{\kappa}$:

PROPOSITION 2.2.1. *Under the assumptions in Theorem 2.2.2, if $\boldsymbol{\varepsilon}$ consists of independent heteroscedastic variables, the estimate of the heterogeneity parameter given in (2.7) satisfies $\hat{\kappa} \xrightarrow{P} \kappa$.*

REMARK 2.2.1. *(Asymptotic variance) It is worth emphasizing that when the noise variables are independent and homogeneous, which implies that $\kappa = 1$, the asymptotic distribution given in (2.9) is consistent with the result derived from i.i.d. Gaussian design in [8]. In fact, an explicit formula can be derived for the asymptotic variance based essentially on the Stieltjes transform of the Marčenko-Pastur distribution, see e.g. Lemma 3.11 in [1]. Define*

$$m_\tau(z) = \int_{b_-(\tau)}^{b_+(\tau)} \frac{1}{x+z} f_\tau(x)\mathrm{d}x + \frac{1}{z}\left(1 - \frac{1}{\tau}\right) 1_{\{\tau>1\}} = \frac{(\tau - z - 1) + \sqrt{(\tau - z - 1)^2 + 4z\tau}}{2z\tau}.$$

Then we can obtain

$$h_1(\gamma, \tau) = \frac{1}{\gamma} m_\tau\left(\frac{1}{\gamma}\right) = \frac{(\tau\gamma - 1 - \gamma) + \sqrt{(\tau\gamma - 1 - \gamma)^2 + 4\tau\gamma}}{2\tau\gamma},$$

and

$$h_2(\gamma, \tau) = -\frac{1}{\gamma^2} m_\tau'\left(\frac{1}{\gamma}\right) = -\frac{(\tau\gamma - \tau + \gamma + 1)\left(-\gamma - 1 + \sqrt{(\tau\gamma - 1 - \gamma)^2 + 4\tau\gamma}\right)}{2\gamma^2\tau^2\sqrt{(\tau\gamma - 1 - \gamma)^2 + 4\tau\gamma}}.$$

We illustrate the asymptotic variance in Figure 2.1 with $\kappa = 1$ and $n = 100$. From this figure, fixing the aspect ratio $1/\tau$, the variance of $\hat{\gamma}$ increases in the true SNR γ_0; while fixing γ_0, the variance of $\hat{\gamma}$ first decreases and then increases in the aspect ratio $1/\tau$.

2.3. Simulations

In this section, we conduct numerical experiments. Throughout our numerical experiments, we use the Minorization-Maximization (MM) algorithm given in [46] to maximize (2.1.1) and hence obtain the random effects likelihood estimate $\hat{\gamma}$ and $\hat{\sigma}^2$.

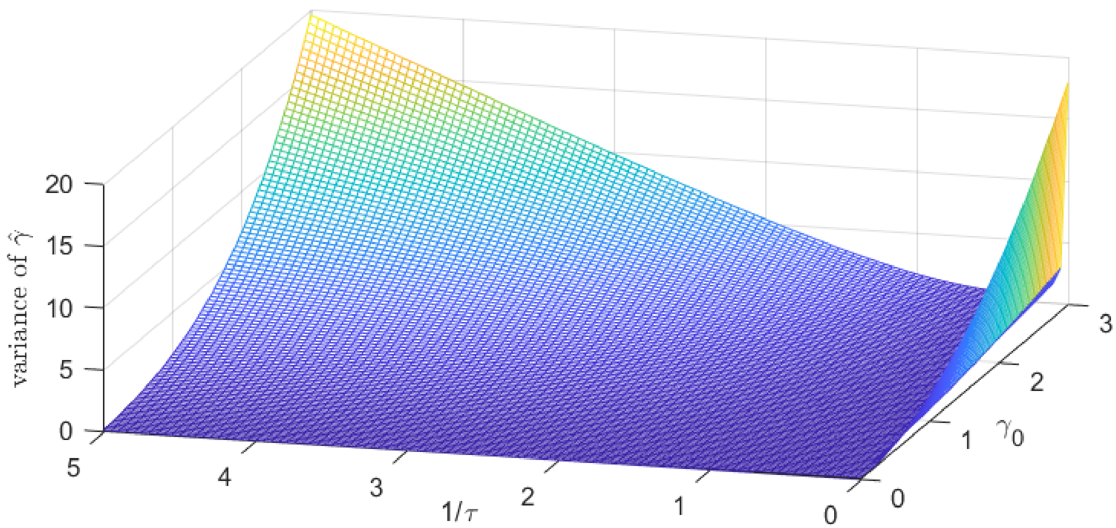


FIGURE 2.1. Asymptotic variance of $\hat{\gamma}$ with $\kappa = 1$ and $n = 100$.

2.3.1. Consistency of $\hat{\gamma}$, $\hat{\sigma}^2$ and $\hat{\kappa}$. In this subsection, we consider the linear model with heterogeneous but uncorrelated noise, and then demonstrate the consistency of REML $\hat{\gamma}$ and $\hat{\sigma}^2$, as well as $\hat{\kappa}$ defined in (2.7). Here we only consider uncorrelated noise since we need to show the behavior of $\hat{\kappa}$. For the case of correlated noise, we will illustrate the sampling distribution of $\hat{\gamma}$ in the next subsection.

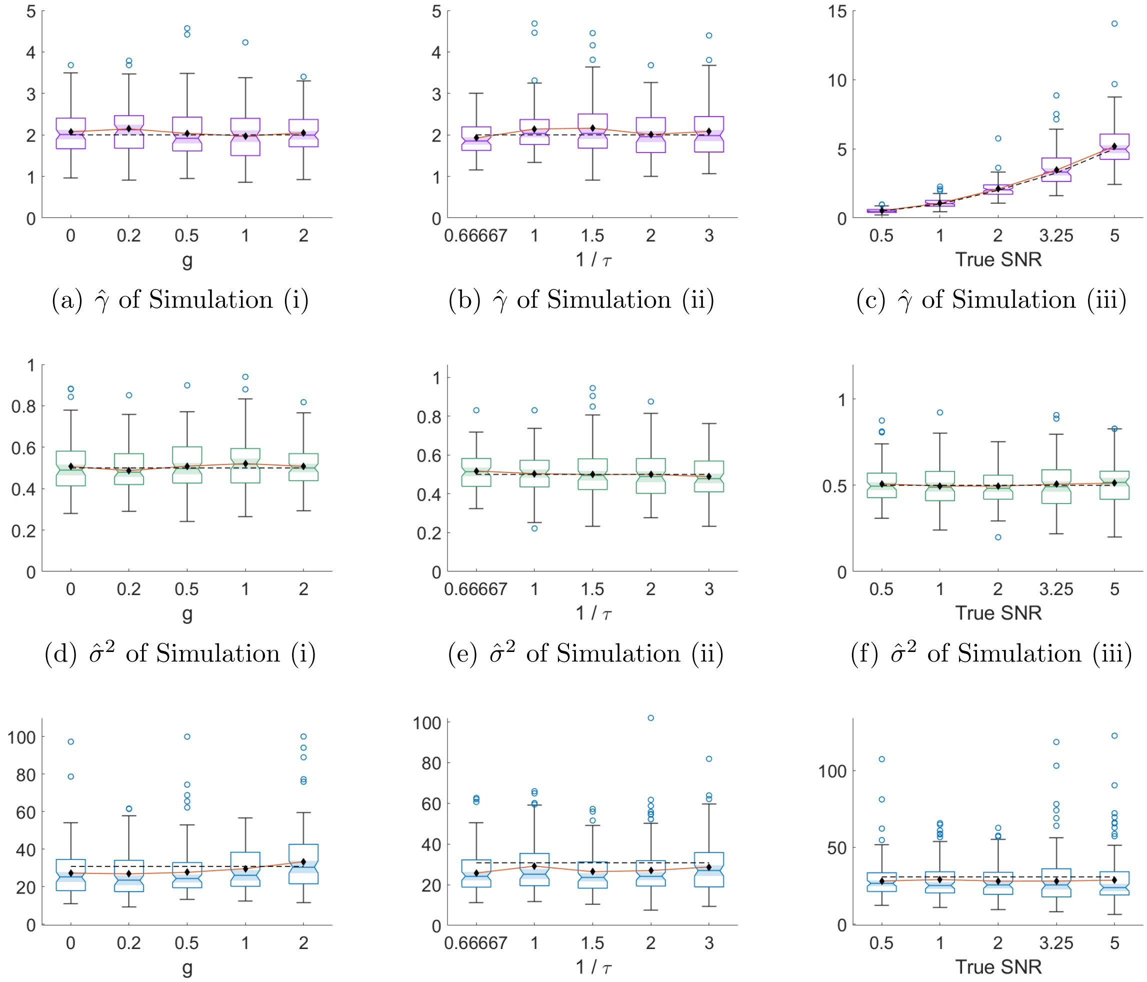
We assume that the coefficient vector $\boldsymbol{\beta}$ is generated in the form of

$$\boldsymbol{\beta} \propto \left(1, 2^{-g}, 3^{-g}, \cdots, p^{-g}\right)^\top, \tag{2.10}$$

where $g \geq 0$ determines the rate of decay for the coefficients, and the norm of $\boldsymbol{\beta}$ is determined by σ_0^2 and the SNR γ_0 by $\|\boldsymbol{\beta}\|_2^2 = \gamma_0 \sigma_0^2$.

For heteroscedastic independent noise, we generate σ_i^2 by the geometric sequence by first generating $(\sigma_{1'}^2, \sigma_{2'}^2 \cdots, \sigma_{n'}^2)$ in the form of

$$(\sigma_{1'}^2, \sigma_{2'}^2 \cdots, \sigma_{n'}^2) \propto \left(1, q, q^2, \cdots, q^n\right), \tag{2.11}$$



(a) $\hat{\gamma}$ of Simulation (i) (b) $\hat{\gamma}$ of Simulation (ii) (c) $\hat{\gamma}$ of Simulation (iii)

(d) $\hat{\sigma}^2$ of Simulation (i) (e) $\hat{\sigma}^2$ of Simulation (ii) (f) $\hat{\sigma}^2$ of Simulation (iii)

(g) $\hat{\kappa}$ of Simulation (i) (h) $\hat{\kappa}$ of Simulation (ii) (i) $\hat{\kappa}$ of Simulation (iii)

FIGURE 2.2. Estimates of SNR and noise level for simulations (i)(ii)(iii) under the t5 design. Each simulation is conducted over 100 independent Monte Carlo samples. The true SNR γ_0, σ_0^2 and κ_0 are marked in dash line. The black diamonds represent average estimates by $\hat{\gamma}, \hat{\sigma}_0^2$ and $\hat{\kappa}$

where $q > 0$ and $\sum_{i=1}^n \sigma_{i'}^2 = n\sigma_0^2$. Next, $(\sigma_{1'}^2, \sigma_{2'}^2 \cdots, \sigma_{n'}^2)$ is shuffled randomly to generate $(\sigma_1^2, \sigma_2^2 \cdots, \sigma_n^2)$. Throughout this section, we chose $q = 0.95$ and $\sigma_0^2 = 0.5$, which also gives $\kappa = 30.7692$ by fixing $n = 1200$.

We consider the following settings on the key parameters to investigate and illustrate how the performance of $\hat{\gamma}$ relies on the magnitude decay in $\boldsymbol{\beta}$, the aspect ratio p/n, and the SNR γ_0.

(i) (Varying magnitude decay in $\boldsymbol{\beta}$): Fix $n = 1200$, $p = 2000, \gamma_0 = 2$. Let g be varied from 0 to 2.

(ii) (Varying aspect ratio): Fix $n = 1200$, $g = 0.5$, $\gamma_0 = 2$. Let the aspect ratio $1/\tau = p/n$ be varied from $2/3$ to 3.

(iii) (Varying SNR): Fix $n = 1200$, $p = 2000$, $g = 0.5$. Let γ_0 be varied from 0.5 to 5.

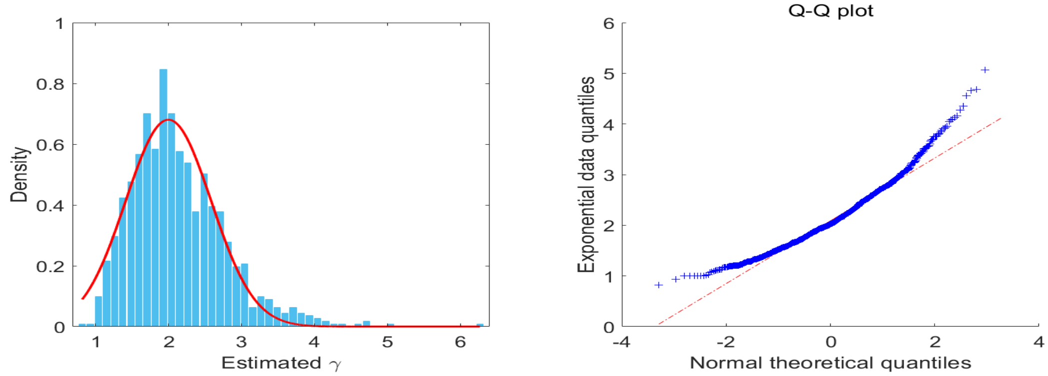
Each simulation consists of 100 independent Monte Carlo samples. The performances of $\hat{\gamma}$, $\hat{\sigma}^2$ and $\hat{\kappa}$ under simulation settings (i)(ii)(iii) are shown in Figure 2.2 for design matrices with i.i.d. t_5 entries. All of these estimators appear to be consistent under various circumstances. In particular, we can see that the variance of estimators $\hat{\gamma}$ keeps more or less the same over different magnitude decays in $\boldsymbol{\beta}$, while increases with the aspect ratio $1/\tau \in [2/3, 3]$, and also increases with the true SNR $\gamma_0 \in [1/2, 5]$. These observations are in line with the asymptotic variance presented in (2.9), which has also been illustrated in Figure 2.1.

2.3.2. Distribution of $\hat{\gamma}$. Now let's study the sampling distribution of $\hat{\gamma}$ empirically for heterogeneous and correlated noise. Here we consider the setting $n = 1200$, $p = 2000$, $\sigma_0^2 = 0.5$, and $\gamma_0 = 2$. For the coefficient vector, assume $\boldsymbol{\beta}_0$ generated from (2.10) with $g = 0.5$. For the heterogeneous and correlated noise, in addition to the variances generated according to (2.11) with $q = 0.95$, we impose the pairwise covariances as $\Sigma_{ij} = \rho^{|i-j|} \cdot \sigma_i \sigma_j$ with $\rho = 0.1$. The resulting κ defined in Theorem 2.2.2 is $\kappa = \|\boldsymbol{\Sigma}\|_F^2/(n\sigma_0^4) = 30.8188$. We conduct Monte Carlo simulations with 1000 independent samples under the following settings of design matrices:

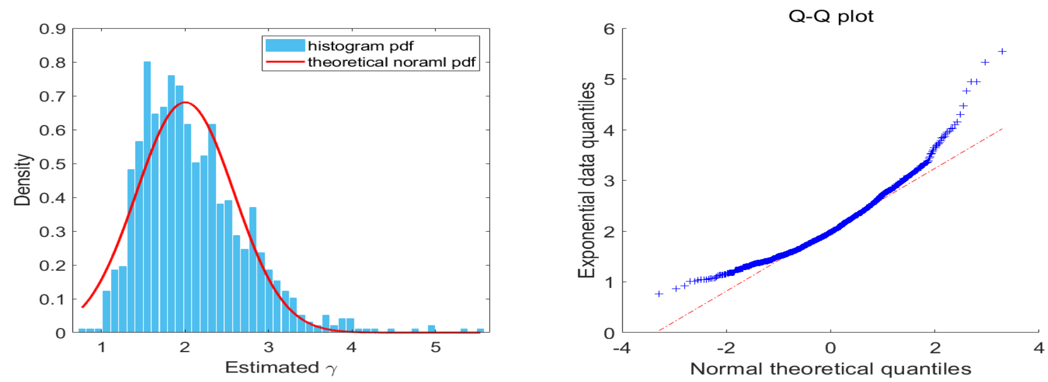
(i) The entries of $\boldsymbol{Z}$ are i.i.d. Rademacher random variables.

(ii) The entries of $\boldsymbol{Z}$ are i.i.d. standardized t_5 random variables.

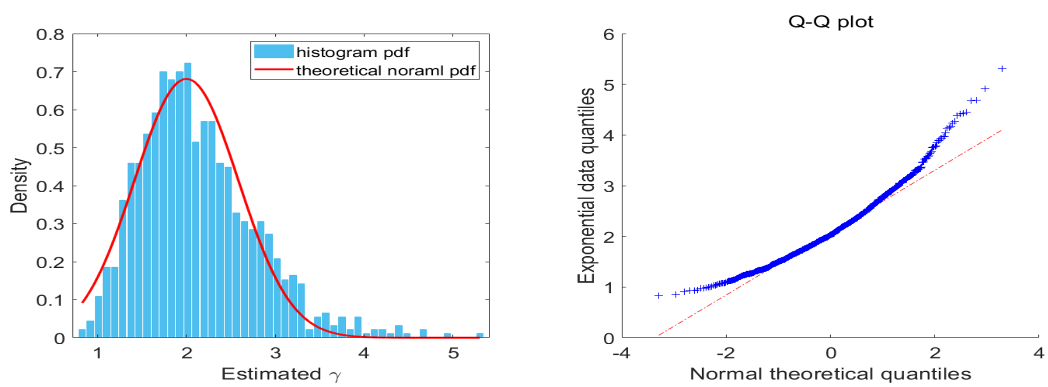
(iii) The standardized genotype model proposed in [24]: First, let the allele frequencies for SNPs be generated from $f_i \sim \text{Unif}[0.05, 0.5]$ for $i = 1, \ldots, p$. Next, generate the entries of the genotype matrix $\boldsymbol{U}$ by following a discrete distribution over $\{0, 1, 2\}$ with assigned probabilities $(1 - f_j)^2$, $2f_j(1 - f_j)$, and f_j^2, respectively. Finally,



(a) Rademacher design

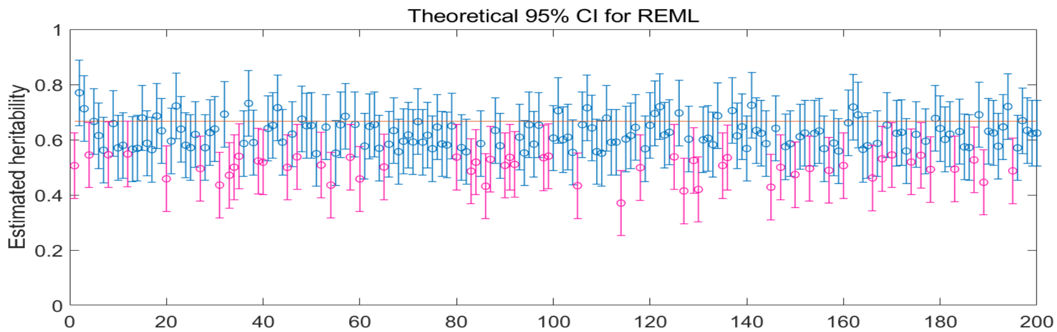


(b) Standardized t_5 design

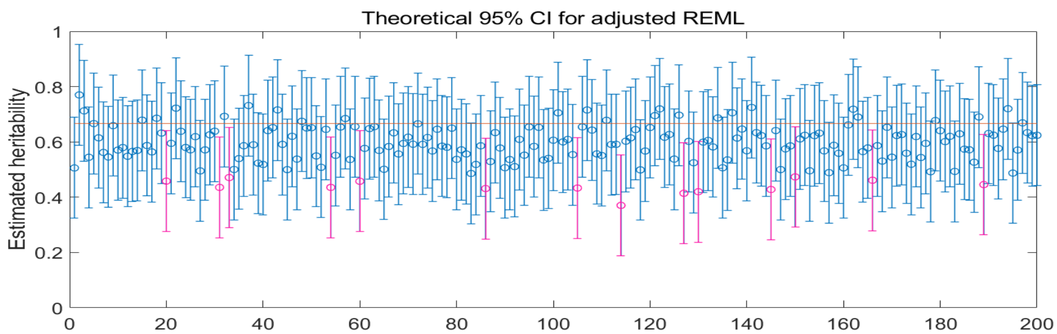


(c) Standardized genotype design

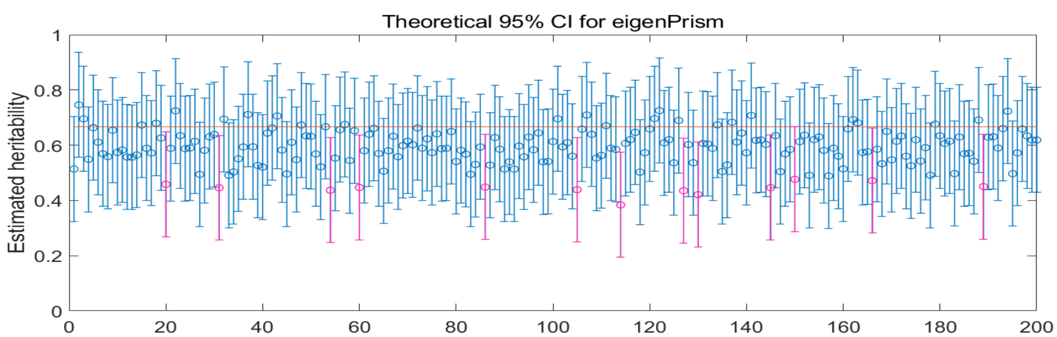
FIGURE 2.3. Probability density of the estimated SNR $\hat{\gamma}$ and the normal Q–Q plot of corresponding $\hat{\gamma}$ sets. In the probability density graph, the purple curve shows the pdf of normal distribution with sample mean and sample variance and the red curve shows the pdf of our theoretical normal distribution when the features are independent.



(a) REML



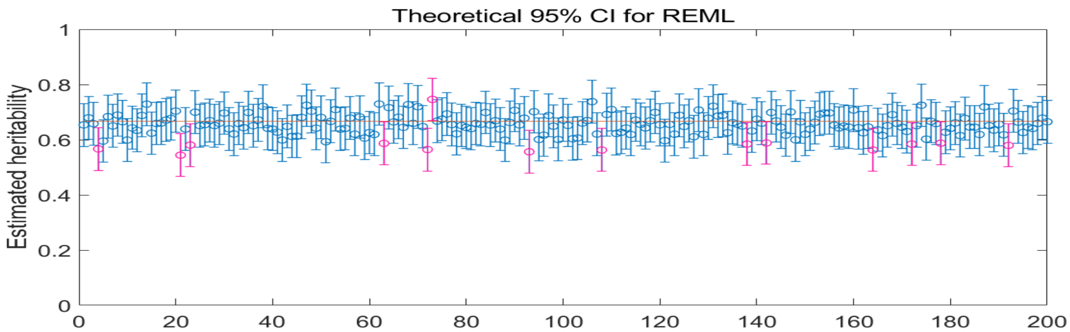
(b) Adjusted REML



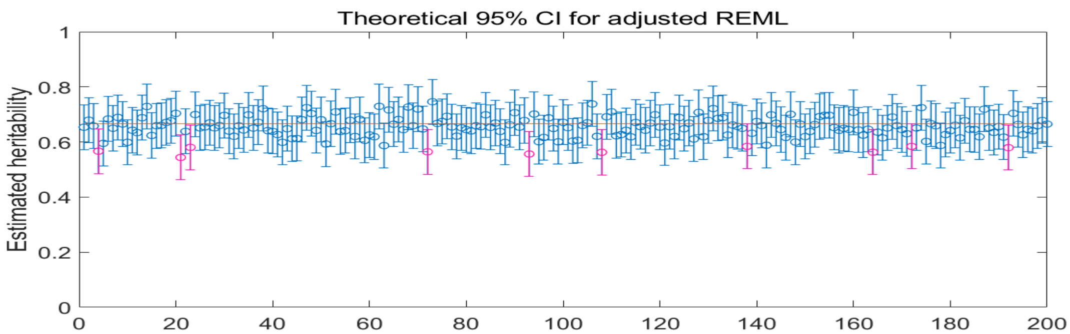
(c) EigenPrism

FIGURE 2.4. Plug-in 95% CI for 200 independent datasets for REML, adjusted REML (with heterogeneous) and eigenPrism when $\kappa = 30.77$. The estimates $\hat{h}$ are marked as circles and the true SNRs h_0 are marked by the red line. The purple bars indicate the cases when the 95% CI does not cover h_0.

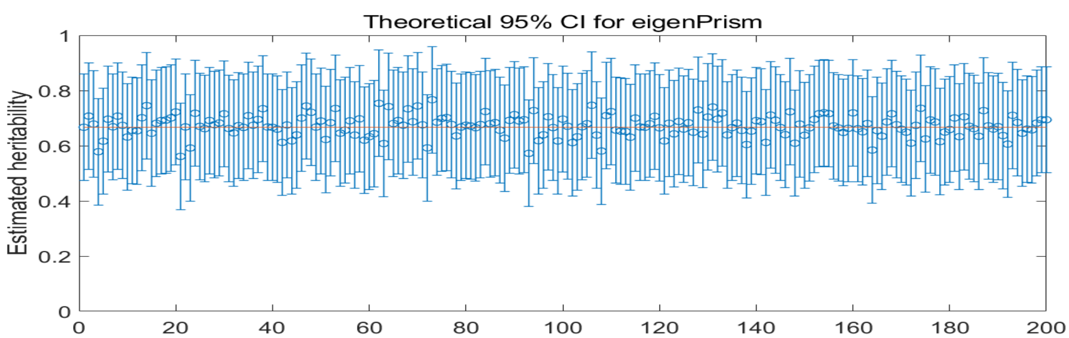
standardize each column of $\boldsymbol{U}$ to have zero mean and unit variance to obtain the design matrix $\boldsymbol{Z}$.



(a) REML



(b) Adjusted REML



(c) EigenPrism

FIGURE 2.5. Plug-in 95% CI for 200 independent datasets for REML, adjusted REML (with heterogeneous) and eigenPrism when $\kappa = 2.54$. The estimates $\hat{h}$ are marked as circles and the true SNRs h_0 are marked by the red line. The purple bars indicate the cases when the 95% CI does not cover h_0.

2.3.3. Compare REML, adjusted REML and eigenPrism γ_0. The heritability h is

$$h = \frac{1}{\gamma + 1},$$

then we define the estimator of heritability to be

$$\hat{h} = \frac{1}{\hat{\gamma} + 1}.$$

To compare REML(without considering κ), adjusted REML and eigenPrism, we consider the signal setting (2.10), and set the parameters $n = 1200$, $p = 2000$, $\gamma_0 = 2$, $\sigma_0^2 = 0.5$ and $g \sim \text{Unif}[0, 2]$. The entries of design matrix $\boldsymbol{Z}$ are i.i.d standard normal. We just change the noise variance setting (2.11) with parameters $q = 0.95$ in case 1 and $q = 0.995$ in case 2, the correspondingly are $\kappa_0 = 30.77$ and $\kappa_0 = 2.54$. Results are shown in Figure 2.4 and 2.5.

2.4. Proof of Main Results

In this section, we give proof for Theorems in Section 2.2. In Section 2.4.1, we will present some useful supporting lemmas; in Section 2.4.2, we present a new representation based on rademacher sequences which will be useful in our proof; in Section 2.4.3, we present a proof for our main result Theorem 3.2.4; in Section 2.4.4, we present a proof for our main result Theorem 2.2.2; in Section 2.4.5 we present a proof for 2.2.1; we leave proof of lemmas used in former subsections to the appendix.

2.4.1. Supporting Lemmas.

Lemma 2.4.0.1. *Under the assumptions of Theorem 3.2.4, we have*

$$\max_{i \in [n]} \varepsilon_i^2 = O_P(\log n) \tag{2.12}$$

and

$$\left| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 - \sigma_0^2 \right| = o_P(1). \tag{2.13}$$

Moreover, under the assumptions of Theorem 2.2.2, there holds

$$\frac{1}{\sqrt{n}} \left(\left(\sum_{i=1}^{n} \varepsilon_i^2 \right) - n\sigma_0^2 \right) \Longrightarrow \mathcal{N}(0, 2\kappa\sigma_0^4). \tag{2.14}$$

LEMMA 2.4.0.2 (Theorem 9.10 of [1]). *Under the assumptions of Theorem 3.2.4, for* $\boldsymbol{V}_\gamma$ *defined in* (2.2) *and any integer* $k > 0$, *it is obvious that* $\|\boldsymbol{V}_\gamma^{-k}\| \le 1$. *Moreover, we have*

$$\left|\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-k}\right) - h_k(\gamma, \tau)\right| = O_P\left(\frac{1}{n}\right),$$

where $h_k(\gamma, \tau)$ *is defined in* (2.8).

A key technique in proving Theorem 3.2.4 is the "leave-k-out" argument developed in [24]. Here we list some useful notations.

DEFINITION 2.4.1. *Denote* $\boldsymbol{Z} = [\boldsymbol{z}_1, \ldots, \boldsymbol{z}_p]$ *as a concatenation of column vectors. For any subset* $C \subset \{1, \ldots, p\}$, *denote* $\boldsymbol{V}_{\gamma,-C} \coloneqq \boldsymbol{V}_\gamma - \frac{\gamma}{p}\sum_{k\in C}\boldsymbol{z}_k\boldsymbol{z}_k^\top$. *For example, for any* $i \neq j$,

$$\boldsymbol{V}_{\gamma,-ij} \coloneqq \boldsymbol{V}_{\gamma,-\{ij\}} = \boldsymbol{V}_\gamma - \frac{\gamma}{p}\left(\boldsymbol{z}_i\boldsymbol{z}_i^\top + \boldsymbol{z}_j\boldsymbol{z}_j^\top\right).$$

Furthermore, for $1 \le i, j \le p$, *define*

$$\eta_{ij,C}^{(l)} \coloneqq \boldsymbol{z}_i^\top \boldsymbol{V}_{\gamma,-C}^{-l}\boldsymbol{z}_j. \tag{2.15}$$

Finally, in the case $C = \emptyset$, *simply denote*

$$\eta_{ij}^{(l)} \coloneqq \boldsymbol{z}_i^\top \boldsymbol{V}_\gamma^{-l}\boldsymbol{z}_j. \tag{2.16}$$

The proofs of the following five results, Lemma 2.4.1.1 to Lemma 2.4.1.5, essentially follow the arguments or ideas in [24], though there might be some small differences. For completeness, we provide self-contained proofs for these results in the appendix except for Lemma 2.4.1.4, which has been explicitly given in the supplement of [24] (See Proposition S.1 therein).

LEMMA 2.4.1.1. *Under the conditions of Theorem 3.2.4, we have*

$$\max_{k\in[p]}\left|\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-l}\right) - \operatorname{trace}\left(\boldsymbol{V}_{\gamma,-k}^{-l}\right)\right| \le 2^l - 1, \quad l = 1, 2, 3, 4, \tag{2.17}$$

and for $\eta^{(l)}_{kk,k}$ *defined in* (2.15),

$$\max_{k\in[p]}\left|\frac{1}{n}\eta^{(l)}_{kk,k}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-l}\right)\right|=O_P\left(\sqrt{\frac{\log n}{n}}\right),\quad l=1,2.$$

LEMMA 2.4.1.2. *Under the conditions of Theorem 3.2.4, for fixed* $\gamma>0$, *we have*

$$\max_{1\le k\le p}\left|\frac{1}{n}\boldsymbol{z}_k^\top\boldsymbol{V}_\gamma^{-1}\boldsymbol{z}_k-\frac{1}{n}\frac{\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})}{1+\frac{\gamma}{p}\operatorname{trace}(V_\gamma^{-1})}\right|=O_P\left(\sqrt{\frac{\log n}{n}}\right), \tag{2.18}$$

$$\max_{1\le k\le p}\left|\frac{1}{n}\boldsymbol{z}_k^\top\boldsymbol{V}_\gamma^{-2}\boldsymbol{z}_k-\frac{1}{n}\frac{\operatorname{trace}(\boldsymbol{V}_\gamma^{-2})}{\left(1+\frac{\gamma}{p}\operatorname{trace}(V_\gamma^{-1})\right)^2}\right|=O_P\left(\sqrt{\frac{\log n}{n}}\right), \tag{2.19}$$

$$\max_{1\le k\le p}\left|\frac{1}{n}\boldsymbol{z}_k^\top\boldsymbol{V}_\gamma^{-l}\boldsymbol{z}_k-\frac{1}{np}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-l}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right|=O_P\left(\sqrt{\frac{\log n}{n}}\right),\quad l=1,2, \tag{2.20}$$

$$\max_{1\le k\le p}\left|\left(\boldsymbol{z}_k^\top\boldsymbol{B}_\gamma\boldsymbol{z}_k\right)^l-\left(\frac{1}{p}\operatorname{trace}\left(\boldsymbol{B}_\gamma\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right)^l\right|=O_P\left(\sqrt{\frac{\log n}{n}}\right),\quad l=1,2, \tag{2.21}$$

$$\left|\frac{1}{np}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\boldsymbol{Z}\boldsymbol{Z}^\top\right)-\frac{1}{n}\frac{\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})}{1+\frac{\gamma}{p}\operatorname{trace}(V_\gamma^{-1})}\right|=O_P\left(\frac{1}{n}\right), \tag{2.22}$$

and

$$\left|\frac{1}{np}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\boldsymbol{Z}\boldsymbol{Z}^\top\right)-\frac{1}{n}\frac{\operatorname{trace}(\boldsymbol{V}_\gamma^{-2})}{\left(1+\frac{\gamma}{p}\operatorname{trace}(V_\gamma^{-1})\right)^2}\right|=O_P\left(\frac{1}{n}\right). \tag{2.23}$$

LEMMA 2.4.1.3. *Under the conditions of Theorem 3.2.4, for fixed* $\gamma>0$, *we have*

$$
\begin{aligned}
&\max_{1\le k\le p}\left|\frac{1}{n}\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-1}\boldsymbol{z}_k - \frac{1}{np}\,\mathrm{trace}\left(\boldsymbol{V}_\gamma^{-1}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right.\\
&\left.-\frac{1}{\left(1+\frac{\gamma}{p}\,\mathrm{trace}(V_\gamma^{-1})\right)^2}\left(\frac{1}{n}\eta_{kk,k}^{(1)} - \frac{1}{n}\,\mathrm{trace}(V_\gamma^{-1})\right)\right| = O_P\left(\frac{\log n}{n}\right),
\end{aligned}
\tag{2.24}
$$

and

$$
\begin{aligned}
&\max_{1\le k\le p}\left|\frac{1}{n}\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-2}\boldsymbol{z}_k - \frac{1}{np}\,\mathrm{trace}\left(\boldsymbol{V}_\gamma^{-2}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right.\\
&+\frac{\mathrm{trace}(\boldsymbol{V}_\gamma^{-2})}{\left(1+\frac{\gamma}{p}\,\mathrm{trace}(V_\gamma^{-1})\right)^3}\frac{2\gamma}{p}\left(\frac{1}{n}\eta_{kk,k}^{(1)} - \frac{1}{n}\,\mathrm{trace}(V_\gamma^{-1})\right)\\
&\left.-\frac{1}{\left(1+\frac{\gamma}{p}\,\mathrm{trace}(V_\gamma^{-1})\right)^2}\left(\frac{1}{n}\eta_{kk,k}^{(2)} - \frac{1}{n}\,\mathrm{trace}(V_\gamma^{-2})\right)\right| = O_P\left(\frac{\log n}{n}\right).
\end{aligned}
$$

Lemma 2.4.1.4. *Under the conditions of Theorem 3.2.4, for fixed $\gamma > 0$ and $l = 1, 2$, we have*

$$
\max_{1\le k\le p}\mathbb{E}\left[\left(\frac{1}{n}\eta_{kk,k}^{(l)} - \frac{1}{n}\,\mathrm{trace}(V_{\gamma,-k}^{-l})\right)^2\right] \le \frac{C}{n},
$$

and

$$
\max_{1\le i<j\le p}\left|\mathbb{E}\left[\left(\frac{1}{n}\eta_{ii,i}^{(l)} - \frac{1}{n}\,\mathrm{trace}(V_{\gamma,-i}^{-l})\right)\left(\frac{1}{n}\eta_{jj,j}^{(l)} - \frac{1}{n}\,\mathrm{trace}(V_{\gamma,-j}^{-l})\right)\right]\right| \le \frac{C}{n^2},
$$

where C is a constant independent of n.

Lemma 2.4.1.5. *Under the conditions of Theorem 3.2.4, for fixed $\gamma > 0$ and $l = 1, 2$, we have*

$$
\max_{k\ne j}|\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-l}\boldsymbol{z}_j|^2 = O_P(n\log n) \quad \textit{and} \quad \max_{k\ne j}|\boldsymbol{z}_k^\top \boldsymbol{B}_\gamma\boldsymbol{z}_j|^2 = O_P\left(\frac{\log n}{n}\right).
\tag{2.25}
$$

Further, under the assumptions of Theorem 2.2.2, we have

$$(2.26) \qquad \begin{cases} \dfrac{1}{p(p-1)} \displaystyle\sum_{i \neq j} n \left(\boldsymbol{z}_i^\top \boldsymbol{B}_\gamma \boldsymbol{z}_j \right)^2 & = \overline{\theta}_1(\gamma, \tau) + o_P(1) \\ \displaystyle\sum_{i \neq j} \beta_i^2 \beta_j^2 n \left(\boldsymbol{z}_i^\top \boldsymbol{B}_\gamma \boldsymbol{z}_j \right)^2 & = \|\boldsymbol{\beta}\|^4 \overline{\theta}_1(\gamma, \tau) + o_P(1), \end{cases}$$

where $\overline{\theta}_1(\gamma, \tau) > 0$ is a constant only depending on γ and τ.

The above lemmas rely crucially on the "leave-k-column-out" argument in [**24**]. Given we are dealing with heteroscedastic and correlated noises, we also need the following results, which rely on a similar "leave-k-row-out" argument.

Lemma 2.4.1.6. *For any fixed $\gamma > 0$, under the assumptions of Theorem 3.2.4, we have*

$$(2.27) \qquad \max_{i \in [n]} \left| \left(\boldsymbol{V}_\gamma^{-l} \right)_{ii} - \frac{1}{n} \operatorname{trace} \left(\boldsymbol{V}_\gamma^{-l} \right) \right| = O_P \left(\sqrt{\frac{\log n}{n}} \right), \quad l = 1, 2, 3, 4,$$

which implies

$$(2.28) \qquad \max_{i \in [n]} \left| (\boldsymbol{B}_\gamma)_{ii} - \frac{1}{n} \operatorname{trace} (\boldsymbol{B}_\gamma) \right| = O_P \left(\sqrt{\frac{\log n}{n^3}} \right)$$

and

$$(2.29) \qquad \max_{i \in [n]} \left| (\boldsymbol{B}_\gamma)_{ii}^2 - \left(\frac{1}{n} \operatorname{trace} (\boldsymbol{B}_\gamma) \right)^2 \right| = O_P \left(\sqrt{\frac{\log n}{n^5}} \right).$$

Lemma 2.4.1.7. *For any fixed $\gamma > 0$, under the assumptions of Theorem 3.2.4, we have*

$$(2.30) \qquad \max_{1 \leq i < j \leq n} |(\boldsymbol{B}_\gamma)_{ij}| = O_P \left(\sqrt{\frac{\log n}{n^3}} \right)$$

and under the assumptions of Theorem 2.2.2

$$
\begin{cases} n\sum_{i\neq j}(\boldsymbol{B}_\gamma)_{ij}^2 = \overline{\theta}_2(\gamma,\tau)+o_P(1) \\ n\sum_{i\neq j}\varepsilon_i^2\varepsilon_j^2(\boldsymbol{B}_\gamma)_{ij}^2 = \overline{\theta}_2(\gamma,\tau)\sigma_0^4+o_P(1), \end{cases} \tag{2.31}
$$

where $\overline{\theta}_2(\gamma,\tau)>0$ is a constant only depending on γ and τ.

LEMMA 2.4.1.8. *For any fixed $\gamma>0$, under the assumptions of Theorem 3.2.4, we have*

$$
\max_{k\in[p]}\max_{i\in[n]}\left|\left(\boldsymbol{V}_{\gamma,-k}^{-l}\right)_{ii}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-l}\right)\right| = O_P\left(\sqrt{\frac{\log n}{n}}\right),\quad l=1,2,3,4. \tag{2.32}
$$

2.4.2. New Representation based on Rademacher Sequences. Since the entries of $\boldsymbol{Z}$ are independent and symmetric, we can replace the original design matrix $\boldsymbol{Z}$ with $\tilde{\tilde{\boldsymbol{Z}}}=\boldsymbol{\Lambda}_\zeta\boldsymbol{Z}\boldsymbol{\Lambda}_\xi$ with the diagonal matrices

$$
\boldsymbol{\Lambda}_\zeta=\operatorname{diag}(\zeta_1,\dots,\zeta_n),\quad \boldsymbol{\Lambda}_\xi=\operatorname{diag}(\xi_1,\dots,\xi_p),
$$

with ζ_i's and ξ_j's are i.i.d. Rademacher random variables that are also independent of $\boldsymbol{Z}$, since $\boldsymbol{Z}$ and $\tilde{\tilde{\boldsymbol{Z}}}$ have the same distribution. We also denote

$$
\boldsymbol{\xi}=(\xi_1,\dots,\xi_n)^\top \quad\text{and}\quad \boldsymbol{\zeta}=(\zeta_1,\dots,\zeta_n)^\top.
$$

Under this new representation of the design matrix, the linear model (2.1) becomes

$$
\boldsymbol{y}=\boldsymbol{\Lambda}_\zeta\boldsymbol{Z}\boldsymbol{\Lambda}_\xi\boldsymbol{\beta}+\boldsymbol{\varepsilon}. \tag{2.33}
$$

We want to emphasize that under this new representation, we still define $\boldsymbol{V}_\gamma$ and $\boldsymbol{B}_\gamma$ as before:

$$
\boldsymbol{V}_\gamma=\boldsymbol{I}_n+\frac{\gamma}{p}\boldsymbol{Z}\boldsymbol{Z}^\top,\quad\text{and}\quad \boldsymbol{B}_\gamma=\frac{\boldsymbol{V}_\gamma^{-1}}{n}-\frac{\boldsymbol{V}_\gamma^{-2}}{\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})}.
$$

However, the representation of the estimating equation (2.3) should be changed. In fact, the original $\boldsymbol{Z}\boldsymbol{Z}^\top$ is replaced with $\boldsymbol{\Lambda}_\zeta\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{\Lambda}_\zeta$. Therefore, the original $\boldsymbol{V}_\gamma$ defined in (2.2) should be replaced with

$$\tilde{\tilde{\boldsymbol{V}}}_\gamma = \boldsymbol{I}_n + \frac{\gamma}{p}\boldsymbol{\Lambda}_\zeta \boldsymbol{Z}\boldsymbol{Z}^\top \boldsymbol{\Lambda}_\zeta = \boldsymbol{\Lambda}_\zeta \left(\boldsymbol{I}_n + \frac{\gamma}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\boldsymbol{\Lambda}_\zeta = \boldsymbol{\Lambda}_\zeta \boldsymbol{V}_\gamma \boldsymbol{\Lambda}_\zeta.$$

Also, it is easy to see that the original $\boldsymbol{B}_\gamma$ should be replaced with $\tilde{\tilde{\boldsymbol{B}}}_\gamma = \boldsymbol{\Lambda}_\zeta \boldsymbol{B}_\gamma \boldsymbol{\Lambda}_\zeta$. Therefore, the estimating equation (2.3) should be rewritten as

$$\begin{aligned}
\Delta(\gamma) &:= \boldsymbol{y}^\top \tilde{\tilde{\boldsymbol{B}}}_\gamma \boldsymbol{y} \\
&= \left(\boldsymbol{\Lambda}_\zeta \boldsymbol{Z}\boldsymbol{\Lambda}_\xi \boldsymbol{\beta} + \boldsymbol{\varepsilon}\right)^\top \boldsymbol{\Lambda}_\zeta \boldsymbol{B}_\gamma \boldsymbol{\Lambda}_\zeta \left(\boldsymbol{\Lambda}_\zeta \boldsymbol{Z}\boldsymbol{\Lambda}_\xi \boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) \\
&= \boldsymbol{\xi}^\top \boldsymbol{\Lambda}_{\boldsymbol{\beta}} \boldsymbol{Z}^\top \boldsymbol{B}_\gamma \boldsymbol{Z}\boldsymbol{\Lambda}_{\boldsymbol{\beta}} \boldsymbol{\xi} + 2\boldsymbol{\xi}^\top \boldsymbol{\Lambda}_{\boldsymbol{\beta}} \boldsymbol{Z}^\top \boldsymbol{B}_\gamma \boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}} \boldsymbol{\zeta} + \boldsymbol{\zeta}^\top \boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}} \boldsymbol{B}_\gamma \boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}} \boldsymbol{\zeta} \\
&= [\boldsymbol{\xi}^\top, \boldsymbol{\zeta}^\top] \begin{bmatrix} \boldsymbol{\Lambda}_\beta \boldsymbol{Z}^\top \boldsymbol{B}_\gamma \boldsymbol{Z}\boldsymbol{\Lambda}_\beta & \boldsymbol{\Lambda}_\beta \boldsymbol{Z}^\top \boldsymbol{B}_\gamma \boldsymbol{\Lambda}_\epsilon \\ \boldsymbol{\Lambda}_\epsilon \boldsymbol{B}_\gamma \boldsymbol{Z}\boldsymbol{\Lambda}_\beta & \boldsymbol{\Lambda}_\epsilon \boldsymbol{B}_\gamma \boldsymbol{\Lambda}_\epsilon \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\zeta} \end{bmatrix},
\end{aligned} \tag{2.34}$$

where

$$\boldsymbol{\Lambda}_{\boldsymbol{\beta}} = \mathrm{diag}(\beta_1, \ldots, \beta_n) \quad \text{and} \quad \boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}} = \mathrm{diag}(\varepsilon_1, \ldots, \varepsilon_n).$$

Note that now $\Delta(\gamma)$ is a random variable about $\boldsymbol{Z}$, $\boldsymbol{\varepsilon}$, $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$. Straightforward calculation gives the conditional mean of $\Delta(\gamma)$ on $\boldsymbol{Z}$ and $\boldsymbol{\varepsilon}$:

$$\widetilde{\Delta}_*(\gamma) := \mathbb{E}\left[\Delta(\gamma)\middle|\boldsymbol{Z}, \boldsymbol{\varepsilon}\right] = \sum_{k=1}^p \beta_k^2 \boldsymbol{z}_k^\top \boldsymbol{B}_\gamma \boldsymbol{z}_k + \mathrm{trace}\left(\boldsymbol{\Lambda}_\varepsilon^2 \boldsymbol{B}_\gamma\right). \tag{2.35}$$

Furthermore, the conditional variance of $\sqrt{n}(\Delta(\gamma))$ on $\boldsymbol{Z}$ and $\boldsymbol{\varepsilon}$ can also derived as in the following lemma, the proof of which is deferred to the appendix.

Lemma 2.4.1.9. *The conditional variance of $\Delta(\gamma)$ given $\boldsymbol{Z}$ and $\boldsymbol{\varepsilon}$ has the formula*

$$\begin{aligned}
&Var\left[\sqrt{n}(\Delta(\gamma))|\boldsymbol{Z}, \boldsymbol{\varepsilon}\right] \\
&= \underbrace{2n \sum_{1\le k\ne j\le p} \beta_k^2 \beta_j^2 \left(\boldsymbol{z}_k^\top \boldsymbol{B}_\gamma \boldsymbol{z}_j\right)^2}_{V1} + \underbrace{4n \sum_{k=1}^p \beta_k^2 \boldsymbol{z}_k^\top \boldsymbol{B}_\gamma \boldsymbol{\Lambda}_\varepsilon^2 \boldsymbol{B}_\gamma \boldsymbol{z}_k}_{V_2} + \underbrace{2n \sum_{1\le k\ne j\le n} \varepsilon_k^2 \varepsilon_j^2 (\boldsymbol{B}_\gamma)_{kj}^2}_{V_3}.
\end{aligned} \tag{2.36}$$

2.4.3. Proof of Theorem 3.2.4. With $\widetilde{\Delta}_*(\gamma)$ defined in (2.35) and for any fixed $\gamma > 0$, we first aim at showing

$$\Delta(\gamma) - \widetilde{\Delta}_*(\gamma) \xrightarrow[n\to\infty]{P} 0. \tag{2.37}$$

First, by (2.25) in Lemma 2.4.1.5, we have

$$\sum_{k\neq j} \beta_k^2 \beta_j^2 \left(\boldsymbol{z}_k^\top \boldsymbol{B}_\gamma \boldsymbol{z}_j\right)^2 \leq \left(\max_{k\neq j} |\boldsymbol{z}_k^\top \boldsymbol{B}_\gamma \boldsymbol{z}_j|^2\right) \|\boldsymbol{\beta}\|_2^4 = O_P\left(\frac{\log n}{n}\right).$$

Second, since $\boldsymbol{z}_k$'s are sub-Gaussian vectors, it is obvious that

$$\max_{1\leq k\leq p} \|\boldsymbol{z}_k\|^2 = O_P(n).$$

Also, a simple consequence of Lemma 2.4.0.2 gives $\|\boldsymbol{B}_\gamma\| = O_P(1/n)$, and Lemma 2.4.0.1 implies $\|\boldsymbol{\Lambda}_\varepsilon^2\| \leq O(\log n)$. Therefore,

$$\sum_{k=1}^p \beta_k^2 \boldsymbol{z}_k^\top \boldsymbol{B}_\gamma \boldsymbol{\Lambda}_\varepsilon^2 \boldsymbol{B}_\gamma \boldsymbol{z}_k \leq \|\boldsymbol{\beta}\|_2^2 \|\boldsymbol{B}_\gamma\|^2 \|\boldsymbol{\Lambda}_\varepsilon^2\| \left(\max_{1\leq k\leq p} \|\boldsymbol{z}_k\|^2\right) = O_P\left(\frac{\log n}{n}\right).$$

Third, by (2.12) in Lemma 2.4.0.1 and (2.30) in Lemma 2.4.1.7,

$$\sum_{1\leq k\neq j\leq n} \varepsilon_k^2 \varepsilon_j^2 (\boldsymbol{B}_\gamma)_{kj}^2 = O_P\left(\frac{\log^3 n}{n}\right).$$

Plug the above bounds to (2.36), for any $\delta > 0$, by the conditional Chebyshev's inequality, we have

$$\mathbb{P}\left\{\left|\Delta(\gamma) - \widetilde{\Delta}_*(\gamma)\right| > \delta \middle| Z, \boldsymbol{\varepsilon}\right\} \leq \frac{\operatorname{Var}\left[\Delta(\gamma) | \boldsymbol{Z}, \boldsymbol{\varepsilon}\right]}{\delta^2} \xrightarrow[n\to\infty]{P} 0.$$

Then, by the dominated convergence theorem, we have proved (2.37).

Now, define

$$\Delta_{**}(\gamma) = \sigma_0^2 \operatorname{trace}\left(\boldsymbol{B}_\gamma \boldsymbol{V}_{\gamma_0}\right) = \sigma_0^2 \operatorname{trace}\left(\boldsymbol{B}_\gamma \left(\boldsymbol{I}_n + \frac{\gamma_0}{p} \boldsymbol{Z}\boldsymbol{Z}^\top\right)\right). \tag{2.38}$$

By (2.21) in Lemma 2.4.1.2, we can easily obtain

$$\left|\sum_{k=1}^{p} \beta_k^2 \left(\boldsymbol{z}_k^\top \boldsymbol{B}_\gamma \boldsymbol{z}_k\right) - \frac{\|\boldsymbol{\beta}\|^2}{p} \operatorname{trace}(\boldsymbol{B}_\gamma \boldsymbol{Z}\boldsymbol{Z}^\top)\right| = O_P\left(\sqrt{\frac{\log n}{n}}\right). \tag{2.39}$$

On the other hand, by Lemma 2.4.0.1 and (2.28) in Lemma 2.4.1.6, we have

$$\left|\operatorname{trace}\left(\boldsymbol{\Lambda}_\varepsilon^2 \boldsymbol{B}_\gamma\right) - \frac{1}{n}\operatorname{trace}\left(\boldsymbol{\Lambda}_\varepsilon^2\right)\operatorname{trace}\left(\boldsymbol{B}_\gamma\right)\right| = O_P\left(\sqrt{\frac{\log^3 n}{n}}\right).$$

Furthermore, by Lemmas 2.4.0.1 and 2.4.0.2, we have

$$\left|\frac{1}{n}\operatorname{trace}\left(\boldsymbol{\Lambda}_\varepsilon^2\right)\operatorname{trace}\left(\boldsymbol{B}_\gamma\right) - \sigma_0^2 \operatorname{trace}\left(\boldsymbol{B}_\gamma\right)\right| = o_P(1).$$

Combine the above two inequalities,

$$\left|\operatorname{trace}\left(\boldsymbol{\Lambda}_\varepsilon^2 \boldsymbol{B}_\gamma\right) - \sigma_0^2 \operatorname{trace}\left(\boldsymbol{B}_\gamma\right)\right| = o_P(1). \tag{2.40}$$

Then, by (2.35), (2.38), (2.39), and (2.40), we have

$$\left|\widetilde{\Delta}_*(\gamma) - \Delta_{**}(\gamma)\right| = o_P(1).$$

Combined with (2.37), we have

$$\Delta(\gamma) - \Delta_{**}(\gamma) \xrightarrow[n\to\infty]{P} 0.$$

Finally, we have the following result that characterizes the limit of $\Delta_{**}(\gamma)$ for any $\gamma > 0$.

Lemma 2.4.1.10 ([24]). *Under the assumption of Theorem 3.2.4, we have*

$$\Delta_{**}(\gamma) \xrightarrow{a.s.} c_\gamma,$$

where $c_\gamma > 0$ for $\gamma < \gamma_0$, $c_{\gamma_0} = 0$, and $c_\gamma < 0$ for $\gamma > \gamma_0$.

This result is basically given in [24], and we give a detailed proof in the appendix for self-containedness.

Then, for any $\gamma > 0$, there holds $\Delta(\gamma) \xrightarrow{P} c_\gamma$, which is positive, zero, or negative, depending on whether γ is smaller than, equal to, or greater than γ_0. Then by the argument of Theorem 3.7 in [26], with probability tending to one, the equation $\Delta(\gamma) = 0$ has a root $\hat{\gamma}_n$ such that it converges to γ_0 in probability.

Consistency of $\hat{\sigma}^2$. Let's turn to show $\hat{\sigma}^2_\varepsilon \xrightarrow{P} \sigma_0^2$, where the noise variance estimate is defined in (2.5). Let $s_n(\gamma) = \frac{1}{n}\boldsymbol{y}^\top \boldsymbol{V}_\gamma^{-1}\boldsymbol{y}$. The noise variance estimate is then $\hat{\sigma}^2 = s_n(\hat{\gamma})$. From the previous sections, we know $s_n(\gamma)$ converges to a continuous function $\overline{s}(\gamma)$ in probability. For example, if $\tau < 1$, we have

$$\overline{s}(\gamma) = \sigma_0^2 \int_{b_-(\tau)}^{b_+(\tau)} \left(\frac{1+\gamma_0 x}{1+\gamma x}\right) f_\tau(x)\mathrm{d}x,$$

which gives $\overline{s}(\gamma_0) = \sigma_0^2$.

An important observation is that $s_n(\gamma)$ is decreasing. For any small $\delta > 0$ and $\epsilon > 0$, we know

$$s_n(\gamma - \delta) \leq \overline{s}(\gamma - \delta) + \epsilon \quad \text{and} \quad s_n(\gamma + \delta) \geq \overline{s}(\gamma + \delta) - \epsilon$$

with probability tending to 1. On the other hand, $\hat{\gamma}_n \to \gamma_0$ in probability implies that $\gamma_0 - \delta < \hat{\gamma}_n < \gamma_0 + \delta$ with probability tending to 1. Therefore, we have

$$\overline{s}(\gamma_0 - \delta) + \epsilon \geq s_n(\gamma_0 - \delta) \geq s_n(\hat{\gamma}_n) \geq s_n(\gamma_0 + \delta) \geq \overline{s}(\gamma_0 + \delta) - \epsilon$$

with probability tending to 1. Since δ an ϵ can be arbitrarily small, we have

$$\hat{\sigma}^2 = s_n(\hat{\gamma}) \xrightarrow{P} \overline{s}(\gamma_0) = \sigma_0^2.$$

2.4.4. Proof of Theorem 2.2.2. Through the analysis of asymptotic distribution, we use the shorthand $h_k = h_k(\gamma_0, \tau)$ for $k = 1, 2, 3, 4$, where $h_k(\gamma_0, \tau)$ is defined in (2.8).

2.4.4.1. *Decomposition of* $\Delta(\gamma_0)$. The following lemma essentially given in [24] (without a detailed proof) reduces the asymptotic distribution of $\hat{\gamma}$ to that of $\Delta(\gamma_0)$. For the sake of completeness, we give a detailed proof for it in Appendix:

LEMMA 2.4.1.11. *Under the conditions of Theorem 3.2.4, assume* $\hat{\gamma}_n$ *is a sequence of roots of* $\Delta(\gamma) = 0$, *which converges to* γ_0 *in probability. Then*

$$\sqrt{n}\,(\hat{\gamma} - \gamma_0) = -\frac{\sqrt{n}\Delta(\gamma_0)}{\Delta'_\infty(\gamma_0)} + o_P(1), \tag{2.41}$$

where $\Delta'_\infty(\gamma_0)$ *is the limit of* $\Delta'(\gamma)$ *as* $\gamma = \gamma_0$ *and has the formula*

$$\Delta'_\infty(\gamma_0) = \frac{\sigma_0^2}{\gamma_0}\frac{h_1^2 - h_2}{h_1}.$$

To investigate the asymptotic distribution of $\sqrt{n}\Delta(\gamma_0)$, consider the following orthogonal decomposition:

$$\Delta(\gamma_0) = (\Delta(\gamma_0) - \widetilde{\Delta}_*(\gamma_0)) + \widetilde{\Delta}_*(\gamma_0).$$

In other words, the expectation is taken with respect to Rademacher random variables ξ_i's and ζ_i's. We aim to derive the asymptotic joint distribution of

$$\left(\sqrt{n}(\Delta(\gamma_0) - \widetilde{\Delta}_*(\gamma_0)), \sqrt{n}\widetilde{\Delta}_*(\gamma_0)\right).$$

2.4.4.2. *Conditional Variance of* $\Delta(\gamma_0) - \widetilde{\Delta}_*(\gamma)$. In order to derive the asymptotic joint distribution of $\left(\sqrt{n}(\Delta(\gamma_0) - \widetilde{\Delta}_*(\gamma)), \sqrt{n}\widetilde{\Delta}_*(\gamma)\right)$, we first need to study the conditional distribution of $\Delta(\gamma_0) - \widetilde{\Delta}_*(\gamma)$ given $\boldsymbol{Z}$ and $\boldsymbol{\varepsilon}$. This consists of two steps: conditional variance and conditional normality. Let's first study its conditional variance. Note we have

$$\mathrm{Var}\left[\sqrt{n}(\Delta(\gamma_0) - \widetilde{\Delta}_*(\gamma))\middle|\boldsymbol{Z}, \boldsymbol{\varepsilon}\right] = \mathrm{Var}\left[\sqrt{n}(\Delta(\gamma_0))\middle|\boldsymbol{Z}, \boldsymbol{\varepsilon}\right].$$

LEMMA 2.4.1.12. *Under the condition of Theorem 2.2.2, with V_1, V_2 and V_3 defined in (2.36), we have*

$$V_1 \xrightarrow{P} 2\sigma_0^4 \left((1 - 2h_1 + h_2) + \frac{4h_2 - 2h_3}{h_1} + \frac{h_2 - 2h_3 + h_4}{h_1^2} - \tau \left(h_1 - \frac{h_2}{h_1} \right)^2 \right), \tag{2.42}$$

$$V_2 \xrightarrow{P} 4\sigma_0^4 \left((h_1 - h_2) - 2\frac{h_2 - h_3}{h_1} + \frac{h_3 - h_4}{h_1^2} \right), \tag{2.43}$$

$$V_3 \xrightarrow{P} 2\sigma_0^4 \left(h_2 - \frac{2h_3}{h_1} + \frac{h_4}{h_1^2} - \left(h_1 - \frac{h_2}{h_1} \right)^2 \right). \tag{2.44}$$

Consequently,

$$Var\left[\sqrt{n}(\Delta(\gamma_0))|\boldsymbol{Z}, \boldsymbol{\varepsilon}\right] \xrightarrow{P} 2\sigma_0^4 \left(\frac{h_2 - h_1^2}{h_1^2} - (\tau + 1) \left(h_1 - \frac{h_2}{h_1} \right)^2 \right).$$

Limit of V_1. Since $\sum_{j \neq i} \beta_i^2 \beta_j^2 = \|\boldsymbol{\beta}\|^4 - \sum_{k=1}^p \beta_k^4$ and $\sum_{k=1}^p \beta_k^4 = o(1)$ (by the assumption $\|\boldsymbol{\beta}\|_\infty = o(p^{-1/4})$), by (2.26) in Lemma 2.4.1.5, we have

$$\sum_{j \neq i} \beta_i^2 \beta_j^2 \left| \frac{1}{p(p-1)} \sum_{j \neq i} n \left(\boldsymbol{z}_i^\top \boldsymbol{B}_{\gamma_0} \boldsymbol{z}_j \right)^2 - \theta_3 \right| = o_P(1),$$

and

$$\left| \sum_{j \neq i} \beta_i^2 \beta_j^2 n \left(\boldsymbol{z}_i^\top \boldsymbol{B}_{\gamma_0} \boldsymbol{z}_j \right)^2 - \|\boldsymbol{\beta}\|^4 \theta_3 \right| = o_P(1).$$

These bounds imply that

$$\left| 2n \sum_{j \neq i} \beta_i^2 \beta_j^2 \left(\boldsymbol{z}_i^\top \boldsymbol{B}_{\gamma_0} \boldsymbol{z}_j \right)^2 - 2n \left(\|\boldsymbol{\beta}\|^4 - \sum_{k=1}^p \beta_k^4 \right) \frac{1}{p(p-1)} \sum_{j \neq i} \left(\boldsymbol{z}_i^\top \boldsymbol{B}_{\gamma_0} \boldsymbol{z}_j \right)^2 \right| = o_P(1). \tag{2.45}$$

Then for $\sum_{j \neq i} \left(\boldsymbol{z}_i^\top \boldsymbol{B}_{\gamma_0} \boldsymbol{z}_j \right)^2$, since

$$\text{trace} \left(\left(\boldsymbol{B}_\gamma \frac{1}{p} \boldsymbol{Z} \boldsymbol{Z}^\top \right)^2 \right) = \frac{1}{p^2} \left(\sum_{i \neq j} \left(\boldsymbol{z}_i^\top \boldsymbol{B}_\gamma \boldsymbol{z}_j \right)^2 + \sum_{k=1}^p \left(\boldsymbol{z}_k^\top \boldsymbol{B}_\gamma \boldsymbol{z}_k \right)^2 \right),$$

by (2.21) in Lemma 2.4.1.2, we can have that

$$\left| \left(\frac{p}{p-1} \operatorname{trace} \left(\left(\boldsymbol{B}_{\gamma_0} \frac{1}{p} \boldsymbol{Z}\boldsymbol{Z}^\top \right)^2 \right) - \frac{1}{p-1} \left(\operatorname{trace} \left(\boldsymbol{B}_{\gamma_0} \frac{1}{p} \boldsymbol{Z}\boldsymbol{Z}^\top \right) \right)^2 \right) \right.$$
$$\left. - \frac{1}{p(p-1)} \sum_{j \neq i} \left(\boldsymbol{z}_i^\top \boldsymbol{B}_{\gamma_0} \boldsymbol{z}_j \right)^2 \right|$$
$$= \frac{1}{p-1} \left| \frac{1}{p} \sum_{k=1}^{p} \left(\boldsymbol{z}_k^\top \boldsymbol{B}_\gamma \boldsymbol{z}_k \right)^2 - \left(\operatorname{trace} \left(\boldsymbol{B}_{\gamma_0} \frac{1}{p} \boldsymbol{Z}\boldsymbol{Z}^\top \right) \right)^2 \right|$$
$$\leq \frac{1}{p-1} \max_{1 \leq k \leq p} \left| \left(\boldsymbol{z}_k^\top \boldsymbol{B}_\gamma \boldsymbol{z}_k \right)^2 - \left(\frac{1}{p} \operatorname{trace} \left(\boldsymbol{B}_\gamma \boldsymbol{Z}\boldsymbol{Z}^\top \right) \right)^2 \right| = o_P\left(n^{-1} \right). \tag{2.46}$$

Finally, combining (2.45) and (2.46), we can have

$$\left| 2n \left(\|\boldsymbol{\beta}\|^4 - \sum_{k=1}^{p} \beta_k^4 \right) \left(\frac{p}{p-1} \operatorname{trace} \left(\left(\boldsymbol{B}_{\gamma_0} \frac{1}{p} \boldsymbol{Z}\boldsymbol{Z}^\top \right)^2 \right) - \frac{1}{p-1} \left(\operatorname{trace} \left(\boldsymbol{B}_{\gamma_0} \frac{1}{p} \boldsymbol{Z}\boldsymbol{Z}^\top \right) \right)^2 \right) \right.$$
$$\left. - 2n \sum_{j \neq i} \beta_i^2 \beta_j^2 \left(\boldsymbol{z}_i^\top \boldsymbol{B}_{\gamma_0} \boldsymbol{z}_j \right)^2 \right| = o_P(1). \tag{2.47}$$

Further, Lemma 2.4.0.2 implies

$$\begin{cases} n \operatorname{trace} \left(\left(\boldsymbol{B}_{\gamma_0} \frac{1}{p} \boldsymbol{Z}\boldsymbol{Z}^\top \right)^2 \right) & \xrightarrow{P} \frac{1}{\gamma_0^2} \left((-1 - 2h_1 + h_2) - 2 \frac{h_1 - 2h_2 + h_3}{h_1} + \frac{h_2 - 2h_3 + h_4}{h_1^2} \right), \\ \left(\operatorname{trace} \left(\boldsymbol{B}_{\gamma_0} \frac{1}{p} \boldsymbol{Z}\boldsymbol{Z}^\top \right) \right)^2 & \xrightarrow{P} \frac{1}{\gamma_0^2} \left(\frac{h_1^2 - h_2}{h_1} \right)^2. \end{cases}$$

Combine these limits, the bound in (2.47), the fact $\gamma_0 = \|\boldsymbol{\beta}\|^2 / \sigma_0^2$, and the fact $\sum_{k=1}^{p} \beta_k^4 = o(1)$, we finish the proof of (2.42).

Limit of V_2. For any fixed $\gamma > 0$, straightforward calculation gives

$$n \boldsymbol{z}_k^\top \boldsymbol{B}_\gamma \boldsymbol{\Lambda}_\varepsilon^2 \boldsymbol{B}_\gamma \boldsymbol{z}_k = \frac{1}{n} \boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-1} \boldsymbol{\Lambda}_\varepsilon^2 \boldsymbol{V}_\gamma^{-1} \boldsymbol{z}_k - 2 \frac{\frac{1}{n} \boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-1} \boldsymbol{\Lambda}_\varepsilon^2 \boldsymbol{V}_\gamma^{-2} \boldsymbol{z}_k}{\frac{1}{n} \operatorname{trace}(\boldsymbol{V}_{\gamma_0}^{-1})} + \frac{\frac{1}{n} \boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-2} \boldsymbol{\Lambda}_\varepsilon^2 \boldsymbol{V}_\gamma^{-2} \boldsymbol{z}_k}{\left(\frac{1}{n} \operatorname{trace}(\boldsymbol{V}_{\gamma_0}^{-1}) \right)^2}. \tag{2.48}$$

Then using Sherman-Morrison-Woodbury formula (Theorem A.1.3), we can have

$$\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-1}\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{V}_\gamma^{-1}\boldsymbol{z}_k = \frac{\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_k}{\left(1+\frac{\gamma}{p}\eta_{kk,k}^{(1)}\right)^2},$$

$$\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-1}\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{V}_\gamma^{-2}\boldsymbol{z}_k = \frac{\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{V}_{\gamma,-k}^{-2}\boldsymbol{z}_k}{\left(1+\frac{\gamma}{p}\eta_{kk,k}^{(1)}\right)^2} - \frac{\frac{\gamma}{p}\eta_{kk,k}^{(2)}\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_k}{\left(1+\frac{\gamma}{p}\eta_{kk,k}^{(1)}\right)^3},$$

$$\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-2}\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{V}_\gamma^{-2}\boldsymbol{z}_k = \frac{\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-2}\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{V}_{\gamma,-k}^{-2}\boldsymbol{z}_k}{\left(1+\frac{\gamma}{p}\eta_{kk,k}^{(1)}\right)^2} - \frac{2\frac{\gamma}{p}\eta_{kk,k}^{(2)}\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{V}_{\gamma,-k}^{-2}\boldsymbol{z}_k}{\left(1+\frac{\gamma}{p}\eta_{kk,k}^{(1)}\right)^3}$$

$$+\frac{\left(\frac{\gamma}{p}\eta_{kk,k}^{(2)}\right)^2\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_k}{\left(1+\frac{\gamma}{p}\eta_{kk,k}^{(1)}\right)^4}. \tag{2.49}$$

By Lemma 2.4.0.1, we have

$$\|\boldsymbol{V}_{\gamma,-k}^{-l}\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{V}_{\gamma,-k}^{-m}\| \le \|\boldsymbol{\Lambda}_\varepsilon^2\| = O_P\left(\log n\right)$$

and hence

$$\|\boldsymbol{V}_{\gamma,-k}^{-l}\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{V}_{\gamma,-k}^{-m}\|_F = O_P\left(\sqrt{n}\log n\right).$$

Then, by Hanson-Wright inequality and taking the uniform bound, we can easily get

$$\max_{k\in[p]}\left|\frac{1}{n}\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-l}\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{V}_{\gamma,-k}^{-m}\boldsymbol{z}_k - \frac{1}{n}\operatorname{trace}\left(\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{V}_{\gamma,-k}^{-(l+m)}\right)\right| = o_P(1). \tag{2.50}$$

By Lemma 2.4.1.8 and Lemma 2.4.0.1, we have

$$\max_{k\in[p]}\left|\frac{1}{n}\operatorname{trace}\left(\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{V}_{\gamma,-k}^{-(l+m)}\right) - \frac{1}{n^2}\operatorname{trace}\left(\boldsymbol{\Lambda}_\varepsilon^2\right)\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-(l+m)}\right)\right| = o_P(1). \tag{2.51}$$

By Lemma 2.4.0.1 again, we have

$$\max_{k\in[p]}\left|\frac{1}{n^2}\operatorname{trace}\left(\boldsymbol{\Lambda}_\varepsilon^2\right)\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-(l+m)}\right) - \sigma_0^2\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-(l+m)}\right)\right| = o_P(1). \tag{2.52}$$

Combining (2.50), (2.51) and (2.52) gives

$$\max_{k\in[p]}\left|\frac{1}{n}\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-l}\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{V}_{\gamma,-k}^{-m}\boldsymbol{z}_k-\sigma_0^2\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-(l+m)}\right)\right|=o_P\left(1\right). \tag{2.53}$$

Combining (2.53), (2.49), (2.48), Lemma 2.4.1.1 and Lemma 2.4.0.2, there exists some constant $C(\gamma,\tau)$, such that

$$\max_{k\in[p]}\left|n\boldsymbol{z}_k^\top \boldsymbol{B}_\gamma\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{B}_\gamma\boldsymbol{z}_k-C(\gamma,\tau)\right|=o_P(1).$$

This further implies both

$$\begin{aligned}&\left|C(\gamma,\tau)-n\operatorname{trace}\left(\boldsymbol{B}_\gamma\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{B}_\gamma\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right|\\&\le\left|C(\gamma,\tau)-\frac{n}{p}\sum_{k=1}^p\boldsymbol{z}_k^\top\boldsymbol{B}_\gamma\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{B}_\gamma\boldsymbol{z}_k\right|\\&\le\max_{k\in[p]}\left|C(\gamma,\tau)-n\boldsymbol{z}_k^\top\boldsymbol{B}_\gamma\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{B}_\gamma\boldsymbol{z}_k\right|=o_P(1),\end{aligned}$$

and

$$\begin{aligned}&\left|n\sum_{k=1}^p\beta_k^2\boldsymbol{z}_k^\top\boldsymbol{B}_\gamma\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{B}_\gamma\boldsymbol{z}_k-\|\beta\|^2C(\gamma,\tau)\right|\\&\le\|\beta\|^2\max_{k\in[p]}\left|C(\gamma,\tau)-n\boldsymbol{z}_k^\top\boldsymbol{B}_\gamma\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{B}_\gamma\boldsymbol{z}_k\right|=o_P(1).\end{aligned}$$

Combining the above inequalities, there holds

$$\left|n\sum_{k=1}^p\beta_k^2\boldsymbol{z}_k^\top\boldsymbol{B}_\gamma\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{B}_\gamma\boldsymbol{z}_k-n\|\boldsymbol{\beta}\|^2\operatorname{trace}\left(\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{B}_\gamma\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\boldsymbol{B}_\gamma\right)\right|=o_P(1). \tag{2.54}$$

Finally, note that

$$\boldsymbol{B}_\gamma \frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top \boldsymbol{B}_\gamma = \frac{1}{\gamma}\left(\frac{\boldsymbol{V}_\gamma^{-1}}{n^2} - \left(\frac{2}{n\,\mathrm{trace}(\boldsymbol{V}_\gamma^{-1})} + \frac{1}{n^2}\right)\boldsymbol{V}_\gamma^{-2} + \left(\frac{2}{n\,\mathrm{trace}(\boldsymbol{V}_\gamma^{-1})} + \frac{1}{\left(\mathrm{trace}(\boldsymbol{V}_\gamma^{-1})\right)^2}\right)\boldsymbol{V}_\gamma^{-3} - \frac{1}{\left(\mathrm{trace}(\boldsymbol{V}_\gamma^{-1})\right)^2}\boldsymbol{V}_\gamma^{-4}\right).$$

By (2.27) in Lemma 2.4.1.6, there holds

$$\max_{i\in[n]}\left| n\left(\boldsymbol{B}_\gamma \frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top \boldsymbol{B}_\gamma\right)_{ii} - \mathrm{trace}\left(\boldsymbol{B}_\gamma^2 \frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right| = O_P\left(\sqrt{\frac{\log n}{n^3}}\right).$$

Similar to (2.40), we have

$$\left| n\,\mathrm{trace}\left(\boldsymbol{\Lambda}_\varepsilon^2 \boldsymbol{B}_\gamma \frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top \boldsymbol{B}_\gamma\right) - n\sigma_0^2\,\mathrm{trace}\left(\boldsymbol{B}_\gamma^2 \frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right| = o_P(1). \tag{2.55}$$

Combining (2.54) and (2.55) and letting $\gamma = \gamma_0$, we have

$$\left| 4n\sum_{k=1}^p \beta_k^2 \boldsymbol{z}_k^\top \boldsymbol{B}_{\gamma_0}\boldsymbol{\Lambda}_\varepsilon^2 \boldsymbol{B}_{\gamma_0}\boldsymbol{z}_k - 4\sigma_0^2\|\beta\|^2 n\,\mathrm{trace}\left(\boldsymbol{B}_{\gamma_0}^2 \frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right| = o_P(1). \tag{2.56}$$

Notice that

$$\begin{aligned}
&4\sigma_0^2\|\beta\|^2 n\,\mathrm{trace}\left(\boldsymbol{B}_{\gamma_0}^2 \frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right) \\
&= \frac{4\sigma_0^2\|\beta\|^2}{\gamma_0} n\,\mathrm{trace}\left(\boldsymbol{B}_{\gamma_0}^2\left(\boldsymbol{V}_{\gamma_0} - \boldsymbol{I}_n\right)\right) \\
&= 4\sigma_0^4 n\,\mathrm{trace}\left(\frac{\boldsymbol{V}_{\gamma_0}^{-1} - \boldsymbol{V}_{\gamma_0}^{-2}}{n^2} - 2\frac{\boldsymbol{V}_{\gamma_0}^{-2} - \boldsymbol{V}_{\gamma_0}^{-3}}{n\,\mathrm{trace}(\boldsymbol{V}_{\gamma_0}^{-1})} + \frac{\boldsymbol{V}_{\gamma_0}^{-3} - \boldsymbol{V}_{\gamma_0}^{-4}}{(\mathrm{trace}(\boldsymbol{V}_{\gamma_0}^{-1}))^2}\right).
\end{aligned}$$

Therefore, we got (2.43) by Lemma 2.4.0.2.

Limit of V_3 By (2.31) in Lemma 2.4.1.7, (2.29) in Lemma 2.4.1.6, and Lemma 2.4.0.2, we have

$$\begin{aligned}
V_3 &= (2n)\sum_{i\neq j}\varepsilon_i^2\varepsilon_j^2(\boldsymbol{B}_{\gamma_0})_{ij}^2 \\
&= 2\overline{\theta}_2(\gamma,\tau)\sigma_0^4 + o_P(1) \\
&= \sigma_0^4\left(2n\sum_{i\neq j}(\boldsymbol{B}_{\gamma_0})_{ij}^2\right) + o_P(1) \\
&= \sigma_0^4\left(2n\,\mathrm{trace}\left(\boldsymbol{B}_{\gamma_0}^2\right) - 2n\sum_{i=1}^{n}(\boldsymbol{B}_{\gamma_0})_{ii}^2\right) + o_P(1) \\
&= \sigma_0^4\left(2n\,\mathrm{trace}\left(\boldsymbol{B}_{\gamma_0}^2\right) - 2\left(\mathrm{trace}\left(\boldsymbol{B}_{\gamma_0}\right)\right)^2\right) + o_P(1) \\
&\xrightarrow{P} 2\sigma_0^4\left(\left(h_2 - \frac{2h_3}{h_1} + \frac{h_4}{h_1^2}\right) - \left(h_1 - \frac{h_2}{h_1}\right)^2\right).
\end{aligned}$$

2.4.4.3. *Conditional Distribution of* $\Delta(\gamma_0) - \widetilde{\Delta}_*(\gamma_0)$. Recall from (2.34) that

$$\sqrt{n}\Delta(\gamma_0) = \begin{pmatrix}\boldsymbol{\xi}^\top & \boldsymbol{\zeta}^\top\end{pmatrix}\underbrace{\sqrt{n}\begin{bmatrix}\boldsymbol{\Lambda}_\beta\boldsymbol{Z}^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{Z}\boldsymbol{\Lambda}_\beta & \boldsymbol{\Lambda}_\beta\boldsymbol{Z}^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{\Lambda}_\epsilon \\ \boldsymbol{\Lambda}_\epsilon\boldsymbol{B}_{\gamma_0}\boldsymbol{Z}\boldsymbol{\Lambda}_\beta & \boldsymbol{\Lambda}_\epsilon\boldsymbol{B}_{\gamma_0}\boldsymbol{\Lambda}_\epsilon\end{bmatrix}}_{\boldsymbol{Q}}\begin{pmatrix}\boldsymbol{\xi} \\ \boldsymbol{\zeta}.\end{pmatrix}$$

Note that we can represent $\boldsymbol{Q}$ as

$$\boldsymbol{Q} = \begin{pmatrix}\boldsymbol{\Lambda_\beta} & 0 \\ 0 & \frac{1}{\sqrt{n}}\boldsymbol{\Lambda_\varepsilon}\end{pmatrix}\begin{pmatrix}\boldsymbol{Z}^\top \\ \sqrt{n}\boldsymbol{I}_n\end{pmatrix}\sqrt{n}\boldsymbol{B}_{\gamma_0}\begin{pmatrix}\boldsymbol{Z} & \sqrt{n}\boldsymbol{I}_n\end{pmatrix}\begin{pmatrix}\boldsymbol{\Lambda_\beta} & 0 \\ 0 & \frac{1}{\sqrt{n}}\boldsymbol{\Lambda_\varepsilon}\end{pmatrix}.$$

Since $\widetilde{\Delta}_*(\gamma_0) = \mathbb{E}[\Delta(\gamma_0)|\boldsymbol{Z},\boldsymbol{\varepsilon}]$, we have

$$\sqrt{n}(\Delta(\gamma_0) - \widetilde{\Delta}_*(\gamma_0)) = \begin{pmatrix}\boldsymbol{\xi}^\top & \boldsymbol{\zeta}^\top\end{pmatrix}(\boldsymbol{Q} - \breve{\boldsymbol{Q}})\begin{pmatrix}\boldsymbol{\xi} \\ \boldsymbol{\zeta}\end{pmatrix}.$$

where $\breve{\boldsymbol{Q}}$ is a diagonal matrix that maintains the diagonal part of $\boldsymbol{Q}$. In other words, conditional on $\boldsymbol{Z}$ and $\boldsymbol{\varepsilon}$, $\sqrt{n}(\Delta(\gamma_0) - \widetilde{\Delta}_*(\gamma_0))$ is a quadratic form about $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$.

Here we aim to use Theorem A.1.4 in the appendix to establish a normal approximation of the conditional distribution of $\sqrt{n}(\Delta(\gamma_0) - \widetilde{\Delta}_*(\gamma_0))$ given $\boldsymbol{Z}$ and $\boldsymbol{\varepsilon}$. In other words, we intend to show

$$\frac{\|\boldsymbol{Q} - \breve{\boldsymbol{Q}}\|}{\|\boldsymbol{Q} - \breve{\boldsymbol{Q}}\|_F} = o_P(1). \tag{2.57}$$

To establish a lower bound $\|\boldsymbol{Q} - \breve{\boldsymbol{Q}}\|_F$, consider the block $\sqrt{n}\boldsymbol{\Lambda}_\beta \boldsymbol{Z}^\top \boldsymbol{B}_{\gamma_0} \boldsymbol{Z} \boldsymbol{\Lambda}_\beta$ of $\boldsymbol{Q}$, there holds

$$\|\boldsymbol{Q} - \breve{\boldsymbol{Q}}\|_F \geq \sqrt{n \sum_{i \neq j} \beta_i^2 \beta_j^2 (\boldsymbol{z}_j^\top \boldsymbol{B}_{\gamma_0} \boldsymbol{z}_i)^2},$$

the right-hand side of which converges to some nonvanishing limit based on (2.42).

Let's now establish an upper bound of $\|\boldsymbol{Q} - \breve{\boldsymbol{Q}}\|$. First, we have

$$\|\boldsymbol{Q} - \breve{\boldsymbol{Q}}\| \leq \|\boldsymbol{Q}\| + \|\breve{\boldsymbol{Q}}\| \leq 2\|\boldsymbol{Q}\|,$$

where the last inequality is due to the fact that all diagonal entries are bounded by the operator norm in magnitude. On the other hand,

$$\begin{aligned} \|\boldsymbol{Q}\| &\leq \left\| \begin{pmatrix} \boldsymbol{\Lambda}_{\boldsymbol{\beta}}^2 & 0 \\ 0 & \frac{1}{n}\boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}}^2 \end{pmatrix} \right\| \left\| \begin{pmatrix} \boldsymbol{Z}^\top \\ \sqrt{n}\boldsymbol{I}_n \end{pmatrix} \right\|^2 \|\sqrt{n}\boldsymbol{B}_{\gamma_0}\| \\ &\leq \max_{k \in [p],\ i \in [n]} \left\{ \beta_k^2, \frac{1}{n}\varepsilon_i^2 \right\} \cdot \left(\|\boldsymbol{Z}\| + \sqrt{n} \right)^2 \cdot \left\| \sqrt{n}\boldsymbol{B}_{\gamma_0} \right\|. \end{aligned}$$

By the assumption $\|\beta\|_\infty^2 = o_P(n^{-1/2})$ as well as the bound given in (2.12), we have

$$\max_{k \in [p],\ i \in [n]} \left\{ \beta_k^2, \frac{1}{n}\varepsilon_i^2 \right\} = o_P(n^{-1/2}).$$

Note that we have $\|\boldsymbol{Z}\| = O_P(\sqrt{n})$ by Theorem A.1.6. Also, Lemma 2.4.0.2 implies $\|\sqrt{n}\boldsymbol{B}_{\gamma_0}\| = O_P(n^{-1/2})$. Combining the above, we have $\|\boldsymbol{Q}\| = o_P(1)$. This completes the proof of (2.57).

Therefore, by Theorem A.1.4, we have

$$\mathbb{P}\left\{\frac{\sqrt{n}(\Delta(\gamma_0)-\widetilde{\Delta}_*(\gamma_0))}{\sqrt{\mathrm{Var}\left[\sqrt{n}\Delta(\gamma_0)|\boldsymbol{Z},\boldsymbol{\varepsilon}\right]}}\leq t\middle|\boldsymbol{Z},\boldsymbol{\varepsilon}\right\}\xrightarrow{P}\Phi(t), \tag{2.58}$$

where $\Phi(t)$ is the c.d.f of standard normal distribution.

2.4.4.4. *Asymptotic Distribution of* $\sqrt{n}\widetilde{\Delta}_*(\gamma_0)$. This subsection is intended to show the following result that characterizes the asymptotic distribution of $\widetilde{\Delta}_*(\gamma_0)$ defined in (2.35). Note that by the definition of $\boldsymbol{B}_\gamma$, we always have $\mathrm{trace}(\boldsymbol{B}_\gamma\boldsymbol{V}_\gamma)=0$. Therefore,

$$\|\boldsymbol{\beta}\|^2\,\mathrm{trace}\left(\boldsymbol{B}_{\gamma_0}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)+\sigma_0^2\,\mathrm{trace}\left(\boldsymbol{B}_{\gamma_0}\right)=\sigma_0^2\,\mathrm{trace}\left(\boldsymbol{B}_{\gamma_0}\boldsymbol{V}_{\gamma_0}\right)=0.$$

Then, we can represent $\widetilde{\Delta}_*(\gamma_0)$ as

$$\begin{aligned}\widetilde{\Delta}_*(\gamma_0)&=\sum_{k=1}^p\beta_k^2\boldsymbol{z}_k^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{z}_k-\|\boldsymbol{\beta}\|^2\,\mathrm{trace}\left(\boldsymbol{B}_{\gamma_0}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\\&\quad+\mathrm{trace}\left(\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{B}_{\gamma_0}\right)-\sigma_0^2\,\mathrm{trace}\left(\boldsymbol{B}_{\gamma_0}\right)\\&=\mathrm{trace}\left((\boldsymbol{\Lambda}_\varepsilon^2-\sigma_0^2\boldsymbol{I}_n)\boldsymbol{B}_{\gamma_0}\right)\\&\quad+\sum_{k=1}^p\beta_k^2\frac{1}{n}\left(\boldsymbol{z}_k^\top\boldsymbol{V}_{\gamma_0}^{-1}\boldsymbol{z}_k-\mathrm{trace}\left(\boldsymbol{V}_{\gamma_0}^{-1}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right)\\&\quad+\frac{n}{\mathrm{trace}(\boldsymbol{V}_{\gamma_0}^{-1})}\sum_{k=1}^p\beta_k^2\frac{1}{n}\left(\boldsymbol{z}_k^\top\boldsymbol{V}_{\gamma_0}^{-2}\boldsymbol{z}_k-\mathrm{trace}\left(\boldsymbol{V}_{\gamma_0}^{-2}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right).\end{aligned} \tag{2.59}$$

By Lemma 2.4.1.3 and Lemma 2.4.1.1, we have

$$\begin{aligned}&\sum_{k=1}^p\beta_k^2\frac{1}{n}\left(\boldsymbol{z}_k^\top\boldsymbol{V}_{\gamma_0}^{-1}\boldsymbol{z}_k-\mathrm{trace}\left(\boldsymbol{V}_{\gamma_0}^{-1}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right)\\&=\frac{1}{\left(1+\frac{\gamma}{p}\,\mathrm{trace}(\boldsymbol{V}_\gamma^{-1})\right)^2}\sum_{k=1}^p\beta_k^2\left(\frac{1}{n}\eta_{kk,k}^{(1)}-\frac{1}{n}\,\mathrm{trace}\left(\boldsymbol{V}_{\gamma,-k}^{-1}\right)\right)+O_P\left(\frac{\log n}{n}\right)\end{aligned} \tag{2.60}$$

and

$$\sum_{k=1}^{p}\beta_k^2\frac{1}{n}\left(\boldsymbol{z}_k^\top\boldsymbol{V}_{\gamma_0}^{-2}\boldsymbol{z}_k-\operatorname{trace}\left(\boldsymbol{V}_{\gamma_0}^{-2}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right)$$

$$=-\frac{2\gamma\operatorname{trace}(\boldsymbol{V}_\gamma^{-2})}{p\left(1+\frac{\gamma}{p}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})\right)^3}\sum_{k=1}^{p}\beta_k^2\left(\frac{1}{n}\eta_{kk,k}^{(1)}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_{\gamma,-k}^{-1}\right)\right)$$

$$+\frac{1}{\left(1+\frac{\gamma}{p}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})\right)^2}\sum_{k=1}^{p}\beta_k^2\left(\frac{1}{n}\eta_{kk,k}^{(2)}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_{\gamma,-k}^{-2}\right)\right)+O_P\left(\frac{\log n}{n}\right). \tag{2.61}$$

Then, for $l=1,2$,

$$\mathbb{E}\left(\sum_{k=1}^{p}\beta_k^2\left(\frac{1}{n}\eta_{kk,k}^{(l)}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_{\gamma,-k}^{-l}\right)\right)\right)^2$$

$$=\sum_{k=1}^{p}\beta_k^4\,\mathbb{E}\left[\left(\frac{1}{n}\eta_{kk,k}^{(l)}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_{\gamma,-k}^{-l}\right)\right)^2\right]$$

$$+\sum_{i\neq j}\beta_i^2\beta_j^2\,\mathbb{E}\left[\left(\frac{1}{n}\eta_{ii,i}^{(l)}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_{\gamma,-i}^{-l}\right)\right)\left(\frac{1}{n}\eta_{jj,j}^{(l)}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_{\gamma,-j}^{-l}\right)\right)\right]$$

$$\leq\frac{C}{n}\|\boldsymbol{\beta}\|_4^4+\frac{C}{n^2}\|\boldsymbol{\beta}\|_2^4=o\left(\frac{1}{n}\right), \tag{2.62}$$

for which we have used Lemma 2.4.1.4 as well as the fact $\|\boldsymbol{\beta}\|_4=o(1)$ (by the assumption $\|\boldsymbol{\beta}\|_\infty=o(p^{-1/4})$). Then by (2.60), (2.61), and (2.62), in connection with Lemma 2.4.0.2, there holds

$$\sum_{k=1}^{p}\beta_k^2\frac{1}{n}\left(\boldsymbol{z}_k^\top\boldsymbol{V}_{\gamma_0}^{-l}\boldsymbol{z}_k-\operatorname{trace}\left(\boldsymbol{V}_{\gamma_0}^{-l}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right)=o_P\left(\frac{1}{\sqrt{n}}\right),\quad l=1,2. \tag{2.63}$$

Then, equation (2.59) implies

$$\widetilde{\Delta}_*(\gamma_0)=\operatorname{trace}\left((\boldsymbol{\Lambda}_\varepsilon^2-\sigma_0^2\boldsymbol{I}_n)\boldsymbol{B}_{\gamma_0}\right)+o_P\left(\frac{1}{\sqrt{n}}\right). \tag{2.64}$$

Before deriving the asymptotic distribution of $\widetilde{\Delta}_*(\gamma_0)$, we first introduce a lemma, which is essentially an analogy to (2.63):

LEMMA 2.4.1.13. *Under the conditions of Theorem 2.2.2, for any fixed $\gamma > 0$, there holds*

$$\sum_{i=1}^{n}(\varepsilon_i^2 - \sigma_0^2)\left((\boldsymbol{B}_\gamma)_{ii} - \frac{1}{n}\operatorname{trace}(\boldsymbol{B}_\gamma)\right) = o_P\left(\frac{1}{\sqrt{n}}\right). \tag{2.65}$$

With this lemma, equation (2.64) gives

$$\begin{aligned}\widetilde{\Delta}_*(\gamma_0) &= \sum_{i=1}^{n}(\varepsilon_i^2 - \sigma_0^2)(\boldsymbol{B}_{\gamma_0})_{ii} + o_P\left(\frac{1}{\sqrt{n}}\right) \\ &= \frac{1}{n}\left(\sum_{i=1}^{n}(\varepsilon_i^2 - \sigma_0^2)\right)\operatorname{trace}(\boldsymbol{B}_{\gamma_0}) + o_P\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Recall that Lemma 2.4.0.2 implies $\operatorname{trace}(\boldsymbol{B}_{\gamma_0}) \xrightarrow{P} h_1 - \frac{h_2}{h_1}$. Moreover, (2.14) in Lemma 2.4.0.1 gives

$$\frac{1}{\sqrt{n}}\left(\left(\sum_{i=1}^{n}\varepsilon_i^2\right) - n\sigma_0^2\right) \Longrightarrow \mathcal{N}(0, 2\kappa\sigma_0^4).$$

Then, by the Slutsky's theorem, we get

$$\sqrt{n}\widetilde{\Delta}_*(\gamma_0) = \sqrt{n}\,\mathbb{E}[\Delta(\gamma_0)|\boldsymbol{Z}, \boldsymbol{\varepsilon}] \Longrightarrow \mathcal{N}\left(0, 2\kappa\sigma_0^4\left(\frac{h_2}{h_1} - h_1\right)^2\right). \tag{2.66}$$

2.4.4.5. *Asymptotic Distribution of $\hat{\gamma}$.* Denote

$$\begin{cases}\nu_1 = 2\kappa\sigma_0^4\left(\frac{h_2}{h_1} - h_1\right)^2 \\ \nu_2 = 2\sigma_0^4\left(\frac{h_2 - h_1^2}{h_1^2} - (\tau + 1)\left(h_1 - \frac{h_2}{h_1}\right)^2\right),\end{cases} \tag{2.67}$$

which are the asymptotic variances given in (2.66) and Lemma 2.4.1.12.

To establish the asymptotic distribution of $\hat{\gamma}$, we only need to find that of $\sqrt{n}\Delta(\gamma_0)$ by Lemma 2.4.1.11. Furthermore, it suffices to find the asymptotic joint distribution of

$$\left(\sqrt{n}(\Delta(\gamma_0) - \widetilde{\Delta}_*(\gamma)), \sqrt{n}\widetilde{\Delta}_*(\gamma)\right).$$

For any $(t, s) \in \mathbb{R}^2$, we have

$$
\begin{aligned}
&\mathbb{P}\left\{\frac{\sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\nu_1}} \leq t, \frac{\sqrt{n}\Delta(\gamma_0) - \sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\mathrm{Var}\left[\sqrt{n}\Delta(\gamma_0)|\boldsymbol{Z}, \boldsymbol{\varepsilon}\right]}} \leq s\right\} \\
&= \mathbb{E}\left[\mathbb{P}\left\{\frac{\sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\nu_1}} \leq t, \frac{\sqrt{n}\Delta(\gamma_0) - \sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\mathrm{Var}\left[\sqrt{n}\Delta(\gamma_0)|\boldsymbol{Z}, \boldsymbol{\varepsilon}\right]}} \leq s \middle| \boldsymbol{Z}, \boldsymbol{\varepsilon}\right\}\right] \\
&= \mathbb{E}\left[\mathbb{1}_{\{\sqrt{n}\widetilde{\Delta}_*(\gamma_0)/\sqrt{\nu_1} \leq t\}} \mathbb{P}\left\{\frac{\sqrt{n}\Delta(\gamma_0) - \sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\mathrm{Var}\left[\sqrt{n}\Delta(\gamma_0)|\boldsymbol{Z}, \boldsymbol{\varepsilon}\right]}} \leq s \middle| \boldsymbol{Z}, \boldsymbol{\varepsilon}\right\}\right].
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\left|\mathbb{E}\left[\mathbb{1}_{\{\sqrt{n}\widetilde{\Delta}_*(\gamma_0)/\sqrt{\nu_1} \leq t\}} \mathbb{P}\left\{\frac{\sqrt{n}\Delta(\gamma_0) - \sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\mathrm{Var}\left[\sqrt{n}\Delta(\gamma_0)|\boldsymbol{Z}, \boldsymbol{\varepsilon}\right]}} \leq s \middle| \boldsymbol{Z}, \boldsymbol{\varepsilon}\right\}\right] - \mathbb{E}\left[\mathbb{1}_{\{\sqrt{n}\widetilde{\Delta}_*(\gamma_0)/\sqrt{\nu_1} \leq t\}} \Phi(t)\right]\right| \\
&\leq \mathbb{E}\left[\mathbb{1}_{\{\sqrt{n}\widetilde{\Delta}_*(\gamma_0)/\sqrt{\nu_1} \leq t\}} \left|\mathbb{P}\left\{\frac{\sqrt{n}\Delta(\gamma_0) - \sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\mathrm{Var}\left[\sqrt{n}\Delta(\gamma_0)|\boldsymbol{Z}, \boldsymbol{\varepsilon}\right]}} \leq s \middle| \boldsymbol{Z}, \boldsymbol{\varepsilon}\right\} - \Phi(s)\right|\right] \\
&\leq \mathbb{E}\left[\left|\mathbb{P}\left\{\frac{\sqrt{n}\Delta(\gamma_0) - \sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\mathrm{Var}\left[\sqrt{n}\Delta(\gamma_0)|\boldsymbol{Z}, \boldsymbol{\varepsilon}\right]}} \leq s \middle| \boldsymbol{Z}, \boldsymbol{\varepsilon}\right\} - \Phi(s)\right|\right] \to 0,
\end{aligned}
$$

where the last inequality is due to (2.58). By (2.66), we have

$$
\mathbb{E}\left[\mathbb{1}_{\{\sqrt{n}\widetilde{\Delta}_*(\gamma_0)/\sqrt{\nu_1} \leq t\}} \Phi(s)\right] = \mathbb{P}\left\{\frac{\sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\nu_1}} \leq t\right\} \Phi(s) \to \Phi(t)\Phi(s).
$$

Thus we can have

$$
\mathbb{P}\left\{\frac{\sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\nu_1}} \leq t,\ \frac{\sqrt{n}\Delta(\gamma_0) - \sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\mathrm{Var}\left[\sqrt{n}\Delta(\gamma_0)|\boldsymbol{Z}, \boldsymbol{\varepsilon}\right]}} \leq s\right\} \to \Phi(t)\Phi(s),
$$

which implies that

$$
\left(\frac{\sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\nu_1}}, \frac{\sqrt{n}\Delta(\gamma_0) - \sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\mathrm{Var}\left[\sqrt{n}\Delta(\gamma_0)|\boldsymbol{Z},\ \boldsymbol{\varepsilon}\right]}}\right) \Longrightarrow (X_1, X_2),
$$

where $[X_1, X_2] \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{I}_2)$. By Lemma 2.4.1.12, we have

$$
\mathrm{Var}\left[\sqrt{n}\Delta(\gamma_0)|\boldsymbol{Z}, \boldsymbol{\varepsilon}\right] \xrightarrow{P} \nu_2.
$$

Then, the Slutsky's theorem implies

$$\left(\frac{\sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\nu_1}}, \frac{\sqrt{n}\Delta(\gamma_0)-\sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\mathrm{Var}\left[\sqrt{n}\Delta(\gamma_0)|\boldsymbol{Z},\boldsymbol{\varepsilon}\right]}}, \sqrt{\mathrm{Var}\left[\sqrt{n}\Delta(\gamma_0)|\boldsymbol{Z},\boldsymbol{\varepsilon}\right]}\right) \Longrightarrow \left(X_1, X_2, \sqrt{\nu_2}\right).$$

Letting $g(x,y,z)=\sqrt{\nu_1}x+yz$, by the continuous mapping theorem, we can have

$$\begin{aligned}\sqrt{n}\Delta(\gamma_0) &= \sqrt{\nu_1}\frac{\sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\nu_1}} + \frac{\sqrt{n}\Delta(\gamma_0)-\sqrt{n}\widetilde{\Delta}_*(\gamma_0)}{\sqrt{\mathrm{Var}\left[\sqrt{n}\Delta(\gamma_0)|\boldsymbol{Z},\boldsymbol{\varepsilon}\right]}} \cdot \sqrt{\mathrm{Var}\left[\sqrt{n}\Delta(\gamma_0)|\boldsymbol{Z},\boldsymbol{\varepsilon}\right]} \\ &\Longrightarrow \sqrt{\nu_1}X_1+\sqrt{\nu_2}X_2.\end{aligned}$$

Finally, by Lemma 2.4.1.11 and the Slutsky's theorem, we have

$$\sqrt{n}\left(\hat{\gamma}-\gamma_0\right) \Longrightarrow \mathcal{N}\left(0, \frac{\nu_1+\nu_2}{(\Delta_\infty'(\gamma_0))^2}\right)$$

By $\Delta_\infty'(\gamma_0)=\frac{\sigma_0^2}{\gamma_0}\frac{h_1^2-h_2}{h_1}$ and the expressions in (2.67), simplifying the formula, we have

$$\sqrt{n}\left(\hat{\gamma}-\gamma_0\right) \Longrightarrow \mathcal{N}\left(0, 2\gamma_0^2\left(\frac{1}{h_2-h_1^2}+\kappa-\tau-1\right)\right).$$

2.4.5. Proof of Propostion 2.2.1. Straightforward calculation gives

$$\mathbb{E}[y_i^4] = \sum_{j=1}^{p}\left(\mathbb{E}[z_{ij}^4]-3\right)\beta_j^4 + 3\|\boldsymbol{\beta}\|_2^4 + 6\|\boldsymbol{\beta}\|_2^2\sigma_i^2 + 3\sigma_i^4. \tag{2.68}$$

Then, when the noise is uncorrelated,

$$\kappa = \frac{1}{n\sigma_0^4}\sum_{i=1}^{n}\sigma_i^4 = \frac{1}{3n\sigma_0^4}\sum_{i=1}^{n}\mathbb{E}[y_i^4] - (\gamma_0^2+2\gamma_0) + \frac{1}{3n\sigma_0^4}\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\mathbb{E}[z_{ij}^4]-3\right)\beta_j^4.$$

By the assumption $\|\boldsymbol{\beta}\|_\infty = o(p^{-1/4})$ and z_{ij} is sub-Gaussian, it is obvious that

$$\frac{1}{3n\sigma_0^4}\sum_{i=1}^{n}\sum_{j=1}^{p}\left(\mathbb{E}[z_{ij}^4]-3\right)\beta_j^4 \le \max_{i\in[n];j\in[p]}\left|\mathbb{E}[z_{ij}^4]-3\right|\frac{1}{3\sigma_0^4}\sum_{j=1}^{p}\beta_j^4 = o(1).$$

Furthermore, similar to (2.68), straightforward calculation implies $\max_{1\le i\le n} \mathbb{E}[y_i^8] = O(1)$, which implies

$$\mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} y_i^4\right] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[y_i^4] \le \frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}[y_i^8] = O\left(\frac{1}{n}\right),$$

Then we can have $\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[y_i^4] = \frac{1}{n}\sum_{i=1}^{n} y_i^4 + O_P(n^{-1/2})$. Combining the above, we have

$$\frac{1}{3n\sigma_0^4}\sum_{i=1}^{n} y_i^4 - (\gamma_0^2 + 2\gamma_0) \xrightarrow{P} \kappa.$$

In Theorem 3.2.4 we have already shown that $\hat{\sigma}^2 \xrightarrow{P} \sigma_0^2$ and $\hat{\gamma}_n \xrightarrow{P} \gamma_0$. By the Slutsky's theorem, we obtain $\hat{\kappa} \coloneqq \frac{1}{3n\hat{\sigma}^4}\sum_{i=1}^{n} y_i^4 - (\hat{\gamma}_n^2 + 2\hat{\gamma}_n) \xrightarrow{P} \kappa$.

2.5. Extention to Group SNR Estimation

Let's consider an extension of the standard linear model (2.1) to the case in which the design matrix is partitioned according to several groups of features. In other works, let's assume that the linear model can be represented as

$$\boldsymbol{y} = \sum_{i=1}^{s} \boldsymbol{Z}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon} \tag{2.69}$$

where $\boldsymbol{Z}_i \in \mathbb{R}^{n\times p_i}$ is the design matrix corresponding to the i-th feature group, and $\boldsymbol{\beta}_i \in \mathbb{R}^{p_i}$ is the corresponding vector of regression coefficients, so the design matrix $\boldsymbol{Z} = [\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_s]$ is partitioned into s feature groups. In this case, besides estimating σ_0^2, we are interested in estimating the group SNRs $\gamma_{0i} \coloneqq \|\boldsymbol{\beta}_i\|^2/\sigma_0^2$ for $i = 1, \ldots, s$. This model is motivated by the problem of partitioning heritability discussed in [29]. Also, the problem of estimating group SNR is closely connected to group regularized ridge regression [19].

As with the standard case, we consider a linear random effects model corresponding to (2.69). Assume the i.i.d. noise follows $\mathcal{N}(0, \sigma_\varepsilon^2)$, and replace $\boldsymbol{\beta}_i$ with $p_i^{-\frac{1}{2}}\boldsymbol{\alpha}_i$, where $\boldsymbol{\alpha}_i$ consists of i.i.d. Gaussian random variables with distribution $N(0, \sigma_{\alpha_i}^2)$ for $i = 1, \ldots, s$. Then, the linear model with feature groups (2.69) corresponds to a random effects model

$$\boldsymbol{y} = \sum_{i=1}^{s} p_i^{-\frac{1}{2}} \boldsymbol{Z}_i \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}.$$

As with the standard case, the true values of the parameters σ_ε^2, $\sigma_{\alpha_i}^2$ and $\gamma_i = \sigma_{\alpha_i}^2/\sigma_\varepsilon^2$ are σ_0^2, $\|\boldsymbol{\beta}_i\|^2$ and $\gamma_{0i} = \|\boldsymbol{\beta}_i\|^2/\sigma_0^2$ respectively.

Linear random effects models with feature groups have been well studied in the literature; see also [21]. The log-likelihood function is

$$l(\sigma_\varepsilon^2, \sigma_{\alpha_1}^2, \ldots, \sigma_{\alpha_s}^2) = c - \frac{1}{2} \log\det(\boldsymbol{\Omega}) - \frac{1}{2} \boldsymbol{y}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{y}, \tag{2.70}$$

where

$$\boldsymbol{\Omega} = \boldsymbol{\Omega}(\sigma_\varepsilon^2, \sigma_{\alpha_1}^2, \ldots, \sigma_{\alpha_s}^2) := \sigma_\varepsilon^2 \boldsymbol{I}_n + \sum_{i=1}^{s} \frac{\sigma_{\alpha_i}^2}{p_i} \boldsymbol{Z}_i \boldsymbol{Z}_i^\top := \sigma_\varepsilon^2 \boldsymbol{V}_\gamma, \tag{2.71}$$

and

$$\boldsymbol{V}_\gamma = \boldsymbol{I}_n + \sum_{i=1}^{s} \frac{\gamma_i}{p_i} \boldsymbol{Z}_i \boldsymbol{Z}_i^\top. \tag{2.72}$$

Taking the partial derivatives with respect to the variance parameters, we obtain the score functions and the likelihood equations:

$$\begin{cases} S_{\sigma_\varepsilon^2}(\sigma_\varepsilon^2, \sigma_{\alpha_1}^2, \ldots, \sigma_{\alpha_s}^2) := \frac{1}{2} \boldsymbol{y}^\top \boldsymbol{\Omega}^{-2} \boldsymbol{y} - \frac{1}{2} \operatorname{trace}(\boldsymbol{\Omega}^{-1}) = 0 \\ S_{\sigma_\alpha^2}(\sigma_\varepsilon^2, \sigma_{\alpha_1}^2, \ldots, \sigma_{\alpha_s}^2) := \frac{1}{2} \boldsymbol{y}^\top \boldsymbol{\Omega}^{-1} \left(\frac{1}{p_i} \boldsymbol{Z}_i \boldsymbol{Z}_i^\top \right) \boldsymbol{\Omega}^{-1} \boldsymbol{y} - \frac{1}{2} \operatorname{trace}\left(\boldsymbol{\Omega}^{-1} \frac{1}{p_i} \boldsymbol{Z}_i \boldsymbol{Z}_i^\top \right) = 0, \quad 1 \le i \le s. \end{cases}$$

Similar to the estimating equation of SNR estimation (2.3), the above set of equations lead to the following set of likelihood equations for the vector of group SNRs $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_s)^\top$:

$$\Delta^{(i)}(\boldsymbol{\gamma}) = \frac{\boldsymbol{y}^\top \boldsymbol{V}_\gamma^{-1} \left(\frac{1}{p_i} \boldsymbol{Z}_i \boldsymbol{Z}_i^\top \right) \boldsymbol{V}_\gamma^{-1} \boldsymbol{y}}{\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1} \frac{1}{p_i} \boldsymbol{Z}_i \boldsymbol{Z}_i^\top \right)} - \frac{\boldsymbol{y}^\top \boldsymbol{V}_\gamma^{-2} \boldsymbol{y}}{\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1} \right)} = 0, \quad 1 \le i \le s. \tag{2.73}$$

Our analysis of $\hat{\gamma}$ as a solution of (2.3) in the standard case cannot be extended to (2.73) due to several technical difficulties. For example, the calculation of the asymptotic variance of $\hat{\gamma}$ in the standard case relies on the leave-two-out analysis in random matrix theory that has been derived in [24], but this is difficult to be extended to analyzing the solution to (2.73). Thus, we here only present a preliminary result: we show that the true vector of group SNRs $\boldsymbol{\gamma}_0 = (\gamma_{01}, \ldots, \gamma_{0s})^\top$ is *asymptotically* a root of the likelihood functions defined in (2.73) under certain conditions, i.e., $\Delta^{(i)}(\boldsymbol{\gamma}_0) \xrightarrow{P} 0$ for $i = 1, \ldots, s$.

THEOREM 2.5.1. *Consider the linear fixed effects model with feature groups* (2.69)*, where* $\boldsymbol{Z} = [\boldsymbol{Z}_1, \boldsymbol{Z}_2, \cdots, \boldsymbol{Z}_s]$ *is an* $n \times (p_1 + p_2 + \cdots + p_s)$ *design matrix whose entries are independent, symmetric, sub-Gaussian, and variance-one random variables, and their maximum sub-Gaussian norm is uniformly upper bounded by some numerical constant* C_0*. Let* $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^\top, \ldots, \boldsymbol{\beta}_s^\top]^\top$ *be the* $(p_1 + \cdots + p_s) \times 1$ *vector of regression parameters, and let* $\boldsymbol{\varepsilon}$ *be the* $n \times 1$ *vector of independent noise with mean zero and variance* σ_0^2*. For* $i = 1, \ldots, s$*, denote the* i*-th group SNR as* $\gamma_{0i} \coloneqq \|\boldsymbol{\beta}_i\|^2/\sigma_0^2$ *for* $i = 1, \ldots, s$*.*
Consider the asymptotic setting $n, p_1, p_2, \cdots, p_s \to \infty$ *such that* $n/p_i \to \tau_i > 0$ *for* $i = 1, \ldots, s$*. Also, assume that* $\sigma_0^2 > 0$ *and* $\gamma_{01}, \ldots, \gamma_{0s} > 0$ *are fixed constants for all* n*. Then the likelihood functions of the group SNRs defined in* (2.73) *satisfy* $\Delta^{(i)}(\boldsymbol{\gamma}_0) \xrightarrow{P} 0$*, for* $i = 1, \ldots, s$*.*

2.5.0.1. *Simulations.* In this part, we present some preliminary empirical investigations on the properties of the random effects likelihood estimators $\hat{\boldsymbol{\gamma}}$ discussed in Section 2.5. We will not provide extensive simulation results on its asymptotic distribution, since that has not been addressed in Theorem 2.5.1. Instead, we focus on demonstrating the consistency of $\hat{\boldsymbol{\gamma}}$. For simplicity, we only consider the linear model (2.69) with two feature groups

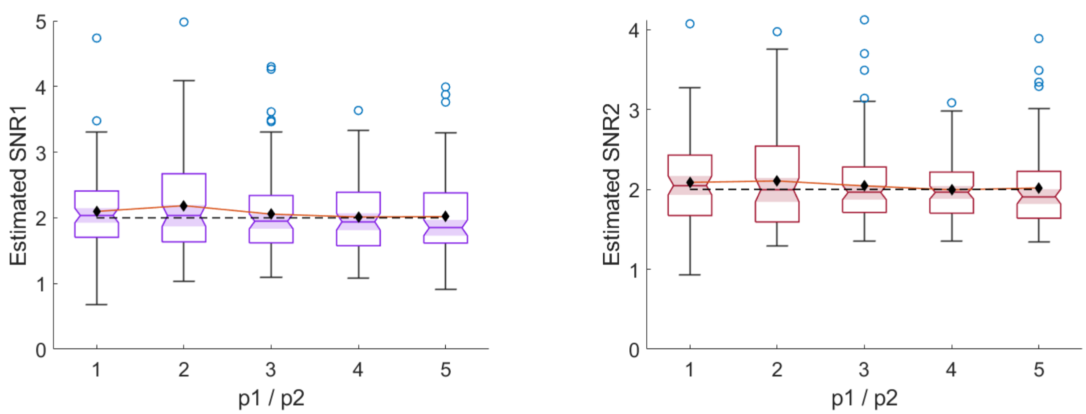
$$y = \boldsymbol{Z}_1\boldsymbol{\beta}_1 + \boldsymbol{Z}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}.$$

We apply the MM algorithm proposed in [46] to estimate the variance parameters. A detailed summary of this algorithm is given in Algorithm 1.

Algorithm 1 MM algorithm for the random effects MLE of the fixed effects model (2.69).

Require: The design matrix $\boldsymbol{Z}_i (i = 1, \ldots, s)$ and the vector of responses $\boldsymbol{y}$;
Ensure: Maximum likelihood estimates $\hat{\gamma}_1, \cdots \hat{\gamma}_s$ and $\hat{\sigma}^2_\varepsilon$

1: Initialize $\sigma_i^{(0)} > 0, i = 0, 1, \cdots, s$;
2: **repeat**
3:
$$\boldsymbol{\Omega}_{(t)} \leftarrow \sigma^2_{0,(t)} + \sum_{i=1}^{s} (\sigma^2_{i,(t)}/p_i) \boldsymbol{Z}_i \boldsymbol{Z}_i^\top;$$
4:
$$\sigma^2_{0,(t+1)} \leftarrow \sigma^2_{0,(t)} \sqrt{\frac{\boldsymbol{y}^\top \boldsymbol{\Omega}^{-2}_{(t)} \boldsymbol{y}}{\text{trace}\left(\boldsymbol{\Omega}^{-1}_{(t)}\right)}};$$
5:
$$\sigma^2_{i,(t+1)} \leftarrow \sigma^2_{i,(t)} \sqrt{\frac{\boldsymbol{y}^\top \boldsymbol{\Omega}^{-1}_{(t)} p_i^{-1} \boldsymbol{Z}_i \boldsymbol{Z}_i^\top \boldsymbol{\Omega}^{-1}_{(t)} \boldsymbol{y}}{\text{trace}\left(\boldsymbol{\Omega}^{-1}_{(t)} p_i^{-1} \boldsymbol{Z}_i \boldsymbol{Z}_i^\top\right)}}, \quad i = 1, \cdots, s;$$
6: **until** the log-likelihood function satisfies
$$\left| l(\sigma^2_{0,(t+1)}, \sigma^2_{1,(t+1)}, \cdots, \sigma^2_{s,(t+1)}) - l(\sigma^2_{0,(t)}, \sigma^2_{1,(t)}, \cdots, \sigma^2_{s,(t)}) \right| < 10^{-4};$$
7: Set the final maximum likelihood estimates as
$$\hat{\sigma}^2_\varepsilon \leftarrow \sigma^2_{0,(t)}, \quad \hat{\gamma}_i \leftarrow \sigma^2_{i,(t)}/\sigma^2_{0,(t)}, \quad i = 1, \ldots, s.$$



(a) $\hat{\gamma}_1$ (b) $\hat{\gamma}_2$

Figure 2.6. Estimates of the SNR γ_1, γ_2 from 100 independent datasets. Structured Z_1 and Z_2 simulated from i.i.d. Rademacher distribution. The true SNR γ_{01}, γ_{02} are marked in dash line and the black diamonds represent the averages.

We fix $n = 1000$, $p = p_1 + p_2 = 2100$, $\gamma_{01} = \gamma_{02} = 2$, $\sigma_0^2 = 0.5$. Both $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are generated according to the magnitude-decay model (2.10) with $g_1 = g_2 = 0.5$. The design matrices $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$ are assumed to follow the i.i.d. Rademacher model.

We are particularly interested in understanding the empirical properties of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ with varying balancedness between the two groups, which is characterized by the ratio p_1/p_2. Simulation results shown in Figure 2.6 suggests that the balancedness between the groups may not have significant impact on the performance of $\hat{\gamma}_1$ and $\hat{\gamma}_2$.

2.5.0.2. *Proof of Theorem 2.5.1.* The proof is technically more involved but has a similar roadmap with the standard case. In particular, for $i = 1, \ldots, s$, we define

$$\Delta_{**}^{(i)}(\boldsymbol{\gamma}) = \sigma_0^2\left(1 - \frac{\gamma_{0i}}{\gamma_i}\right)\left(\frac{\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\frac{1}{p_i}\boldsymbol{Z}_i\boldsymbol{Z}_i^\top\right)}{\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\frac{1}{p_i}\boldsymbol{Z}_i\boldsymbol{Z}_i^\top\right)} - \frac{\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\right)}{\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)}\right)$$
$$+ \sigma_0^2\sum_{r\neq i}\gamma_r\left(\frac{\gamma_{0r}}{\gamma_r} - \frac{\gamma_{0i}}{\gamma_i}\right)\left(\frac{\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\frac{1}{p_r}\boldsymbol{Z}_r\boldsymbol{Z}_r^\top\boldsymbol{V}_\gamma^{-1}\frac{1}{p_i}\boldsymbol{Z}_i\boldsymbol{Z}_i^\top\right)}{\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\frac{1}{p_i}\boldsymbol{Z}_i\boldsymbol{Z}_i^\top\right)} - \frac{\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\frac{1}{p_r}\boldsymbol{Z}_r\boldsymbol{Z}_r^\top\right)}{\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)}\right).$$

It is obvious that $\Delta_{**}^{(i)}(\boldsymbol{\gamma}_0) = 0$. Then, it suffices to show $\Delta^{(i)}(\boldsymbol{\gamma}_0) - \Delta_{**}^{(i)}(\boldsymbol{\gamma}_0) \xrightarrow{P} 0$. Again, we can introduce $\Delta_{*}^{(i)}(\boldsymbol{\gamma}_0)$, which is the mean of $\Delta^{(i)}(\boldsymbol{\gamma}_0)$ conditional on $\boldsymbol{Z}$, as a intermediate step to establish $\Delta^{(i)}(\boldsymbol{\gamma}_0) \approx \Delta_{**}^{(i)}(\boldsymbol{\gamma}_0)$.

For the linear model with a partitioned design (2.69), the likelihood functions with respect to the vector of SNR $\boldsymbol{\gamma}$ have been given in (2.73), i.e.,

$$\Delta^{(i)}(\boldsymbol{\gamma}) = \frac{\frac{1}{n}\boldsymbol{y}^\top\boldsymbol{V}_\gamma^{-1}\left(\frac{1}{p_i}\boldsymbol{Z}_i\boldsymbol{Z}_i^\top\right)\boldsymbol{V}_\gamma^{-1}\boldsymbol{y}}{\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\frac{1}{p_i}\boldsymbol{Z}_i\boldsymbol{Z}_i^\top\right)} - \frac{\frac{1}{n}\boldsymbol{y}^\top\boldsymbol{V}_\gamma^{-2}\boldsymbol{y}}{\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)}. \tag{2.74}$$

Our goal is to show that the asymptotic limit of $\Delta^{(i)}(\boldsymbol{\gamma}_0)$ is 0. Following the proof ideas for Theorem 3.2.4, we also use the trick of Rademacher sequences, which means the response vector can be represented as

$$\boldsymbol{y} = \sum_{i=1}^{s}\left(\sum_{k=1}^{p_i}\beta_{ik}\xi_{ik}\boldsymbol{z}_{ik}\right) + \boldsymbol{\varepsilon}.$$

For $i = 1, \ldots, s$, let

$$A^{(i)} \coloneqq \left(\sum_{r=1}^{s}\left(\sum_{k=1}^{p_r}\beta_{rk}\xi_{rk}\boldsymbol{z}_{rk}\right)\right)^{\top}\boldsymbol{V}_{\gamma}^{-1}(\frac{1}{p_i}\boldsymbol{Z}_i\boldsymbol{Z}_i^{\top})\boldsymbol{V}_{\gamma}^{-1}\left(\sum_{r=1}^{s}\left(\sum_{k=1}^{p_r}\beta_{rk}\xi_{rk}\boldsymbol{z}_{rk}\right)\right),$$

$$B^{(i)} \coloneqq \left(\sum_{r=1}^{s}\left(\sum_{k=1}^{p_r}\beta_{rk}\xi_{rk}\boldsymbol{z}_{rk}\right)\right)^{\top}\boldsymbol{V}_{\gamma}^{-1}(\frac{1}{p_i}\boldsymbol{Z}_i\boldsymbol{Z}_i^{\top})\boldsymbol{V}_{\gamma}^{-1}\boldsymbol{\varepsilon},$$

$$C^{(i)} \coloneqq \boldsymbol{\varepsilon}^{\top}\boldsymbol{V}_{\gamma}^{-1}(\frac{1}{p_i}\boldsymbol{Z}_i\boldsymbol{Z}_i^{\top})\boldsymbol{V}_{\gamma}^{-1}\boldsymbol{\varepsilon},$$

and

$$A^{(0)} \coloneqq \left(\sum_{r=1}^{s}\left(\sum_{k=1}^{p_r}\beta_{rk}\xi_{rk}\boldsymbol{z}_{rk}\right)\right)^{\top}\boldsymbol{V}_{\gamma}^{-2}\left(\sum_{r=1}^{s}\left(\sum_{k=1}^{p_r}\beta_{rk}\xi_{rk}\boldsymbol{z}_{rk}\right)\right),$$

$$B^{(0)} \coloneqq \left(\sum_{r=1}^{s}\left(\sum_{k=1}^{p_r}\beta_{rk}\xi_{rk}\boldsymbol{z}_{rk}\right)\right)^{\top}\boldsymbol{V}_{\gamma}^{-2}\boldsymbol{\varepsilon},$$

$$C^{(0)} \coloneqq \boldsymbol{\varepsilon}^{\top}\boldsymbol{V}_{\gamma}^{-2}\boldsymbol{\varepsilon}.$$

The the conditional expectation of the likelihood functions can be written as

$$\Delta_{*}^{(i)}(\gamma) \coloneqq \mathbb{E}[\Delta^{(i)}(\gamma)|\boldsymbol{Z}]$$

(2.75)

$$= \frac{\frac{1}{n}\left(\mathbb{E}[A^{(i)}|\boldsymbol{Z}] + 2\,\mathbb{E}[B^{(i)}|\boldsymbol{Z}] + \mathbb{E}[C^{(i)}|\boldsymbol{Z}]\right)}{\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_{\gamma}^{-1}\frac{1}{p_i}\boldsymbol{Z}_i\boldsymbol{Z}_i^{\top}\right)} - \frac{\frac{1}{n}\left(\mathbb{E}[A^{(0)}|\boldsymbol{Z}] + 2\,\mathbb{E}[B^{(0)}|\boldsymbol{Z}] + \mathbb{E}[C^{(0)}|\boldsymbol{Z}]\right)}{\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_{\gamma}^{-1}\right)}.$$

Recall that we have also defined

$$\Delta_{**}^{(i)}(\boldsymbol{\gamma}) = \sigma_0^2\left(1 - \frac{\gamma_{0i}}{\gamma_i}\right)\left(\frac{\operatorname{trace}\left(\boldsymbol{V}_{\gamma}^{-2}\frac{1}{p_i}\boldsymbol{Z}_i\boldsymbol{Z}_i^{\top}\right)}{\operatorname{trace}\left(\boldsymbol{V}_{\gamma}^{-1}\frac{1}{p_i}\boldsymbol{Z}_i\boldsymbol{Z}_i^{\top}\right)} - \frac{\operatorname{trace}\left(\boldsymbol{V}_{\gamma}^{-2}\right)}{\operatorname{trace}\left(\boldsymbol{V}_{\gamma}^{-1}\right)}\right)$$
$$+ \sigma_0^2\sum_{r\neq i}\gamma_r\left(\frac{\gamma_{0r}}{\gamma_r} - \frac{\gamma_{0i}}{\gamma_i}\right)\left(\frac{\operatorname{trace}\left(\boldsymbol{V}_{\gamma}^{-1}\frac{1}{p_r}\boldsymbol{Z}_r\boldsymbol{Z}_r^{\top}\boldsymbol{V}_{\gamma}^{-1}\frac{1}{p_i}\boldsymbol{Z}_i\boldsymbol{Z}_i^{\top}\right)}{\operatorname{trace}\left(\boldsymbol{V}_{\gamma}^{-1}\frac{1}{p_i}\boldsymbol{Z}_i\boldsymbol{Z}_i^{\top}\right)} - \frac{\operatorname{trace}\left(\boldsymbol{V}_{\gamma}^{-2}\frac{1}{p_r}\boldsymbol{Z}_r\boldsymbol{Z}_r^{\top}\right)}{\operatorname{trace}\left(\boldsymbol{V}_{\gamma}^{-1}\right)}\right).$$

Then we can follow a similar argument for the standard linear model in Section 2.4 to show $\Delta^{(i)}(\gamma) - \Delta_{*}^{(i)}(\gamma) \xrightarrow{P} 0$ and $\Delta_{*}^{(i)}(\gamma) - \Delta_{**}^{(i)}(\gamma) \xrightarrow{P} 0$.

CHAPTER 3

Method of Moments Estimation: Analysis under Multivariate High-dimensional Linear Models

3.1. Problem Statement and Method

In this chapter, we consider the following multiple response high-dimensional linear model:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{E}, \tag{3.1}$$

where $\boldsymbol{Y} = [\boldsymbol{y}_1, \dots, \boldsymbol{y}_n]^\top \in \mathbb{R}^{n\times q}$ is the response matrix, $\boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n]^\top \in \mathbb{R}^{n\times p}$ is the design matrix, $\boldsymbol{E} = [\boldsymbol{e}_1, \dots, \boldsymbol{e}_n]^\top \in \mathbb{R}^{n\times q}$ is the noise matrix, and $\mathbf{B} = [\boldsymbol{b}_1, \dots, \boldsymbol{b}_p]^\top \in \mathbb{R}^{p\times q}$ is the coefficient matrix. We assume the rows of the noise matrix $\boldsymbol{E}$ satisfies $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n \overset{i.i.d.}{\sim} \mathcal{N}(0, \boldsymbol{\Sigma}_e)$, which is also represented as $vec(\boldsymbol{E}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_e \otimes \boldsymbol{I}_n)$.

We consider both the fixed and random effects models in this chapter.

- (Fixed effects model) Assume $\boldsymbol{B}$ corresponds to fixed effects. In this case, we assume $\boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n]^\top$ is a random design whose rows are independently drawn from $\mathcal{N}(\mathbf{0}, \boldsymbol{I}_p)$. This assumption is common in univariate high-dimensional statistics fixed effects models, see, e.g. [7, 8, 18, 20].
- (Random effects model) Assume $\boldsymbol{B}$ corresponds to random effects. More specifically, assume the rows of $\boldsymbol{B}$ satisfies $\boldsymbol{b}_1, \dots, \boldsymbol{b}_p \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \frac{1}{p}\boldsymbol{\Sigma}_b)$, or equivalently $vec(\boldsymbol{B}) \sim \mathcal{N}(\mathbf{0}, \frac{1}{p}\boldsymbol{\Sigma}_b \otimes \boldsymbol{I}_p)$. Correspondingly, we assume the rows of $\boldsymbol{X}$, i.e. $\boldsymbol{x}_i$'s, are independently drawn from a population with mean zero and covariance $\boldsymbol{\Sigma}$. In comparison to the fixed effects models, we allow the population covariance of the predictors to be correlated, and do not require the normality.

The goal of this paper is to make inferences about the signal-to-noise ratio (SNR) under the two models, which are defined as below.

- (Fixed Effects Models) Denote $\rho^2 = \frac{1}{q}\operatorname{tr}(\boldsymbol{B}^\top \boldsymbol{B})$.
- (Random Effects Models) Denote $\rho^2 = \frac{1}{q}\operatorname{tr}(\boldsymbol{\Sigma}_b)$.

For both models, denote $\sigma^2 = \frac{1}{q}\operatorname{tr}(\boldsymbol{\Sigma}_e)$, and define the signal-to-noise ratio (SNR) as $r^2 = \rho^2/(\rho^2+\sigma^2)$. Our definition of SNR under the multivariate random effects model is similar to that defined in [13], while that under the multivariate fixed effects model is novel in the literature to our knowledge.

3.1.1. Method-of-Moments Estimators. we introduce method-of-moments estimators of the SNR r^2 under both the fixed and random effects models. Under the fixed effects model, our estimator can be viewed as an extension of [7] to the multivariate case; under the random effects model, our estimator is similar to [13].

Fixed Effects Model. Denote $\boldsymbol{W}_b = \boldsymbol{B}^\top \boldsymbol{B}$. By the Wishart moments results summarized in [7], it is straightforward to obtain

$$\mathbb{E}\left[\frac{1}{n}\boldsymbol{Y}^\top \boldsymbol{Y}\right] = \boldsymbol{W}_b + \boldsymbol{\Sigma}_e$$
$$\text{and} \quad \mathbb{E}\left[\frac{1}{n^2}\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}\right] = \frac{p+n+1}{n}\boldsymbol{W}_b + \frac{p}{n}\boldsymbol{\Sigma}_e$$

which result in the following unbiased method-of-moments estimators of $\boldsymbol{W}_b$ and $\boldsymbol{\Sigma}_e$

$$\widehat{\boldsymbol{W}}_b := -\frac{p}{n(n+1)}\boldsymbol{Y}^\top \boldsymbol{Y} + \frac{1}{n(n+1)}\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y} \tag{3.2}$$

and

$$\widehat{\boldsymbol{\Sigma}}_e := \frac{p+n+1}{n(n+1)}\boldsymbol{Y}^\top \boldsymbol{Y} - \frac{1}{n(n+1)}\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}. \tag{3.3}$$

As a consequence, our method-of-moments estimators of ρ^2 and σ^2 are

$$\hat{\rho}^2 = \frac{1}{q}\operatorname{tr}(\widehat{\boldsymbol{W}}_b) \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{q}\operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_e). \tag{3.4}$$

Random Effects Model. Under the random effects model, denote

$$\boldsymbol{S}_n = \frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X} \quad \text{and} \quad \hat{g}_k = \frac{1}{p}\operatorname{tr}(\boldsymbol{S}_n^k) \quad \text{for } k = 1, 2, 3, ...$$

Straightforward calculation under the random effects model gives

$$\mathbb{E}\left[\frac{1}{n}\boldsymbol{Y}^\top \boldsymbol{Y}\middle|\boldsymbol{X}\right] = \boldsymbol{\Sigma}_e + \hat{g}_1\boldsymbol{\Sigma}_b$$
$$\text{and} \quad \mathbb{E}\left[\frac{1}{n^2}\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}\middle|\boldsymbol{X}\right] = \frac{p}{n}\hat{g}_1\boldsymbol{\Sigma}_e - \hat{g}_2\boldsymbol{\Sigma}_b.$$

These equations result in the method-of-moments estimators of $\boldsymbol{\Sigma}_b$ and $\boldsymbol{\Sigma}_e$

$$\widehat{\boldsymbol{\Sigma}}_b := \frac{1}{\hat{g}_2 - \frac{p}{n}\hat{g}_1^2}\left(-\frac{p\hat{g}_1}{n^2}\boldsymbol{Y}^\top \boldsymbol{Y} + \frac{1}{n^2}\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}\right) \tag{3.5}$$

and

$$\widehat{\boldsymbol{\Sigma}}_e := \frac{1}{\hat{g}_2 - \frac{p}{n}\hat{g}_1^2}\left(\frac{\hat{g}_2}{n}\boldsymbol{Y}^\top \boldsymbol{Y} - \frac{\hat{g}_1}{n^2}\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}\right). \tag{3.6}$$

Again, our method-of-moments estimators of ρ^2 and σ^2 are

$$\hat{\rho}^2 = \frac{1}{q}\operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_b) \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{q}\operatorname{tr}(\widehat{\boldsymbol{\Sigma}}_e). \tag{3.7}$$

The corresponding SNR estimator is

$$\hat{r}^2 = \frac{\hat{\rho}^2}{\hat{\rho}^2 + \hat{\sigma}^2}. \tag{3.8}$$

3.2. Main Results

In this section, we introduce our main asymptotic results on the asymptotic distributions of the aforementioned method-of-moments estimators of SNR under both fixed effects model and random effects model.Under random effects model, we consider homoskedastic and heteroskedastic cases. In addition, we will also discuss two scenarios of heteroskedasticity, under which consistent estimators of the variances of these estimators are proposed. Our goal is to make inferences about the SNR $r^2 \coloneqq \rho^2/(\rho^2+\sigma^2)$, which relies on the derivation of the asymptotic distribution of $(\hat{\rho}^2, \hat{\sigma}^2)$.

3.2.1. Fixed Effects Model. First, we specify the following conditions on the distributions of the random design matrix $\boldsymbol{X}$, the coefficient matrix $\boldsymbol{B}$, and the noise matrix $\boldsymbol{E}$.

Assumption 3.2.1 (high-dimensional asymptotics). *The following conditions are assumed to hold:*

- *The sample size $n \to \infty$ while the dimensionality $p(n) \to \infty$ as well, such that the aspect ratio $p(n)/n \to \tau > 0$. The number of responses q is fixed.*
- *The design matrix $\boldsymbol{X}$ is generated with $x_{ij} \sim \mathcal{N}(0,1)$, $1 \le i \le n, 1 \le j \le p$.*

Assumption 3.2.2. *The matrix $\mathbf{B}$ is assumed to be a $p \times q$ deterministic coefficient matrix. Also, $\rho^2 = \frac{1}{q}\operatorname{tr}(\boldsymbol{B}^\top \boldsymbol{B})$ is assumed to be fixed over all instances of n.*

Assumption 3.2.3. *The random noise matrix $\boldsymbol{E}$ is assumed to satisfy $\boldsymbol{E} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}_e)$, where $\boldsymbol{\Sigma}_e$ and thereby $\sigma^2 = \operatorname{tr}(\boldsymbol{\Sigma}_e)/q$ are fixed over all instances of n.*

We are now ready to introduce our asymptotic distribution results for $(\hat{\rho}^2, \hat{\sigma}^2)$ under the above assumptions for the homoskedastic cases.

Theorem 3.2.1. *Under Assumptions 3.2.1, 3.2.2 and 3.2.3, we have*

$$n^{1/2}\boldsymbol{V}^{-1/2}\begin{bmatrix}\hat{\sigma}^2-\sigma^2\\ \hat{\rho}^2-\rho^2\end{bmatrix} \Longrightarrow \mathcal{N}(0, \boldsymbol{I}_2),$$

where the 2×2 symmtric matrix $\boldsymbol{V}$ is defined by

$$V_{11} = \frac{2}{q^2(n+1)^2}\left\{\left(n^2+np\right)\|\boldsymbol{B}^\top\boldsymbol{B}\|_F^2 + 2pn\operatorname{tr}\left(\boldsymbol{\Sigma}_e\boldsymbol{B}^\top\boldsymbol{B}\right) + \left(n^2+np\right)\operatorname{tr}(\boldsymbol{\Sigma}_e^2)\right\}$$

$$V_{22} = \frac{2}{q^2(n+1)^2}\left\{\left(4n^2+np\right)\|\boldsymbol{B}^\top\boldsymbol{B}\|_F^2 + \left(2n^2+2pn\right)\operatorname{tr}\left(\boldsymbol{\Sigma}_e\boldsymbol{B}^\top\boldsymbol{B}\right) + pn\operatorname{tr}(\boldsymbol{\Sigma}_e^2)\right\}$$

and

$$V_{12} = -\frac{2}{q^2(n+1)^2}\left\{\left(2n^2+np\right)\|\boldsymbol{B}^\top\boldsymbol{B}\|_F^2 + 2np\,\mathrm{tr}\left(\boldsymbol{\Sigma}_e\boldsymbol{B}^\top\boldsymbol{B}\right) + pn\,\mathrm{tr}(\boldsymbol{\Sigma}_e^2)\right\}$$

Consequently, we have

$$n^{1/2}(\hat{r}^2 - r^2)/\sigma_r \Rightarrow \mathcal{N}(0,1),$$

where

$$\sigma_r^2 = \rho^4/(\rho^2+\sigma^2)^4 V_{11} + \sigma^4/(\rho^2+\sigma^2)^4 V_{22} - 2\rho^2\sigma^2/(\rho^2+\sigma^2)^4 V_{12}. \tag{3.9}$$

We now highlight several important aspects of Theorem 3.2.1 that underscore our contributions:

REMARK 3.2.1 (Asymptotic distribution). *Our results extend the method-of-moments estimator for the linear fixed effects model developed in [7] to the multivariate setting described in (3.1). When the number of responses $q = 1$, our asymptotic result for $\hat{\gamma}$ exactly recovers the univariate result established in [7].*

REMARK 3.2.2 (Inference). *Theorem 3.2.1 shows that the true SNR r^2 can be consistently estimated by $\hat{r}^2$. To construct confidence intervals for r^2, we note that the asymptotic variance of $\hat{r}^2$, as characterized in Theorem 3.2.1, depends on n, p, q, the variance components (σ^2, ρ^2), the coefficient matrix $\boldsymbol{B}$, and the noise covariance matrix $\boldsymbol{\Sigma}_e$. Although it's hard to estimate $\boldsymbol{B}$ directly, we can estimate $\boldsymbol{B}^\top\boldsymbol{B}$. Then the variance components (σ^2, ρ^2) defined in Section 3.1.1 can be consistently estimated using $(\hat{\sigma}^2, \hat{\rho}^2)$ from (3.4), while $\boldsymbol{B}^\top\boldsymbol{B}$ and $\boldsymbol{\Sigma}_e$ can be estimated via $\widehat{\boldsymbol{W}}_b$ and $\widehat{\boldsymbol{\Sigma}}_e$, as given in (3.2) and (3.3), respectively.*

REMARK 3.2.3 (Gaussian design matrix). *Although assumption 3.2.1 requires that the design matrix $\boldsymbol{X}$ has i.i.d. standard Gaussian entries, which is in agreement with the setting of [7]. However, simulation results in Tables 3.1 and 3.2 indicate that the estimator also performs well when the entries of $\boldsymbol{X}$ are i.i.d. sub-Gaussian (e.g., drawn from a SNP-like distribution).*

Extending the condition of Gaussian entries in the above result to sub-Gaussian ones would be an interesting direction for future work.

Remark 3.2.4 (Assumptions on the coefficient matrix). *Our results impose no sparsity assumptions on the coefficient matrix* $\boldsymbol{B}$*. Simulation results in Tables 3.1 and 3.2 demonstrate that the estimator remains effective in both sparse and dense regimes, highlighting the robustness of our approach.*

3.2.2. Random Effects Model.

Assumption 3.2.4 (high-dimensional asymptotics). *The following conditions are assumed to hold:*

- *The sample size* $n \to \infty$ *while the dimensionality* $p(n) \to \infty$ *as well, such that the aspect ratio* $p(n)/n \to \tau > 0$*. Also,* q *is allowed to diverge.*
- *The design matrix* $\boldsymbol{X}$ *is generated as* $\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{\Sigma}^{1/2}$ *for an* $n \times p$ *matrix* $\boldsymbol{Z}$ *with i.i.d. sub-Gaussian entries satisfying* $\mathbb{E}[Z_{ij}] = 0$, $\operatorname{Var}[Z_{ij}] = 1$, *and* $\|Z_{ij}\|_{\psi_2} \le C_0$ *for all* $1 \le i \le n$ *and* $1 \le j \le p$.
- *The eigenvalues of the* $p \times p$ *positive semidefinite covariance matrix* $\boldsymbol{\Sigma}$ *are assumed to have uniformly bounded eigenvalues:* $0 < C' \le \lambda_j(\boldsymbol{\Sigma}) \le C$ *for* $1 \le j \le p$*, where* C' *and* C *are uniform over all instances of* n.
- *The spectral distribution* $F_{\boldsymbol{\Sigma}}$ *of* $\boldsymbol{\Sigma}$ *converges to a limit probability distribution* H *supported on* $[0, \infty)$*, which is referred to as the population spectral distribution (PSD).*

Assumption 3.2.5. *The random coefficient matrix* $\mathbf{B}$ *is assumed to satisfy* $\boldsymbol{B} \sim \mathcal{N}(\mathbf{0}, \frac{1}{p}\boldsymbol{I}_p \otimes \boldsymbol{\Sigma}_b)$*, where* $\|\boldsymbol{\Sigma}_b\|$ *is uniformly bounded, and* $\rho^2 = \operatorname{tr}(\boldsymbol{\Sigma}_b)/q$ *is fixed over all instances of* n.

Assumption 3.2.6. *The random noise matrix* $\boldsymbol{E}$ *is assumed to satisfy* $\boldsymbol{E} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}_e)$*, where* $\|\boldsymbol{\Sigma}_e\|$ *is uniformly bounded, and* $\sigma^2 = \operatorname{tr}(\boldsymbol{\Sigma}_e)/q$ *is fixed over all instances of* n.

Here we assume the spectral distribution of the predictor for the covariance matrix $\boldsymbol{\Sigma}$ has a limiting distribution, which is commonly assumed in the literature of high-dimensional statistics, e.g., in the analysis of asymptotic risks for ridge regression [10]. This assumption facilitates the asymptotic analysis of $(\hat{\rho}^2, \hat{\sigma}^2)$, particularly for the heteroskedastic cases. Specifically, based on some basic results in random matrix theory [30, 34], the spectral distribution of $\boldsymbol{S}_n = \frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}$, denoted by $F_{\boldsymbol{S}_n}$, converges weakly to some limiting empirical spectral distribution F, supported on $[0, \infty)$ with probability one. An important consequence is that for $k = 1, 2, ...,$

$$\hat{g}_k = \frac{1}{p} \operatorname{tr}\left(\boldsymbol{S}_n^k\right) \xrightarrow{p} g_k := \int_{l=0}^{\infty} l^k dF(l). \tag{3.10}$$

We are now ready to introduce our asymptotic distribution results for $(\hat{\rho}^2, \hat{\sigma}^2)$ under the above assumptions for the homoskedastic cases.

Theorem 3.2.2. *Under Assumptions 3.2.4, 3.2.5 and 3.2.6, we have*

$$n^{1/2}\boldsymbol{V}^{-1/2} \begin{bmatrix} \hat{\sigma}^2 - \sigma^2 \\ \hat{\rho}^2 - \rho^2 \end{bmatrix} \Longrightarrow \mathcal{N}(0, \boldsymbol{I}_2),$$

where the 2×2 symmtric matrix $\boldsymbol{V}$ is defined by

$$V_{11} = \frac{1}{(g_2 - \tau g_1^2)^2 q^2} \left(\left(2g_2^2 - 2\tau g_1^2 g_2\right) \|\boldsymbol{\Sigma}_e\|_F^2 + \left(4g_1^2 g_3 - 4g_1 g_2^2\right) \operatorname{tr}(\boldsymbol{\Sigma}_e \boldsymbol{\Sigma}_b) + \left(\frac{2}{\tau} g_2^3 + \frac{2}{\tau} g_1^2 g_4 - \frac{4}{\tau} g_1 g_2 g_3\right) \|\boldsymbol{\Sigma}_b\|_F^2 \right)$$

$$V_{22} = \frac{1}{(g_2 - \tau g_1^2)^2 q^2} \left(\left(2\tau g_2 - 2\tau^2 g_1^2\right) \|\boldsymbol{\Sigma}_e\|_F^2 + \left(4\tau^2 g_1^3 + 4g_3 - 8\tau g_1 g_2\right) \operatorname{tr}(\boldsymbol{\Sigma}_e \boldsymbol{\Sigma}_b) + \left(2\tau g_1^2 g_2 + \frac{2}{\tau} g_4 - 4g_1 g_3\right) \|\boldsymbol{\Sigma}_b\|_F^2 \right)$$

and

$$V_{12} = \frac{1}{(g_2 - \tau g_1^2)^2 q^2} \left(\left(-2\tau g_1 g_2 + 2\tau^2 g_1^3\right) \|\boldsymbol{\Sigma}_e\|_F^2 + \left(-4g_1 g_3 + 4g_2^2\right) \operatorname{tr}(\boldsymbol{\Sigma}_e \boldsymbol{\Sigma}_b) + \left(-2g_1 g_2^2 - \frac{2}{\tau} g_1 g_4 + \frac{2}{\tau} g_2 g_3 + 2g_1^2 g_3\right) \|\boldsymbol{\Sigma}_b\|_F^2 \right).$$

Recall that g_k's are defined in (3.10). Consequently, we have

$$n^{1/2}(\hat{r}^2 - r^2)/\sigma_r \Rightarrow \mathcal{N}(0,1),$$

where

$$\sigma_r^2 = \rho^4/(\rho^2+\sigma^2)^4 V_{11} + \sigma^4/(\rho^2+\sigma^2)^4 V_{22} - 2\rho^2\sigma^2/(\rho^2+\sigma^2)^4 V_{12}. \tag{3.11}$$

REMARK 3.2.5. *Similar to the fixed effects case, to apply Theorem 3.2.2 and construct a confidence interval for the true SNR r^2, we estimate the quantities g_k using their empirical counterparts $\hat{g}_k$ as defined in (3.10). The covariance matrices $\mathbf{\Sigma}_b$ and $\mathbf{\Sigma}_e$ are estimated using the method-of-moments estimators given in (3.5) and (3.6), respectively.*

3.2.2.1. *Extension to Heteroskedastic Random Effects Models.* In this section, we address the problem of estimating the SNR in multivariate linear models in the presence of heteroskedasticity. Specifically, we extend the previous random effects model to accommodate heterogeneous Gaussian noise, where $\boldsymbol{e}_i \overset{ind}{\sim} \mathcal{N}(0, \mathbf{\Sigma}_i)$ for $i = 1, \ldots, n$. In other words, each observation has an individual noise covariance $\mathbf{\Sigma}_i$. Note that this general heteroskedastic model effectively encompasses several non-Gaussian noise settings.

- Consider the heavy-tailed multivariate noise model $\boldsymbol{e}_i = \xi_i \tilde{\boldsymbol{e}}_i$, where $\tilde{\boldsymbol{e}}_i \overset{i.i.d.}{\sim} \mathcal{N}(0, \mathbf{\Sigma}_e)$, and ξ_i's are i.i.d. heavy-tailed random variables with $\mathbb{E}[\xi_i^2] = 1$. This model includes some commonly-used multivariate heavy-tailed models, such as multivariate Student-t distributions. Conditional on the values of ξ_i's, we have $\boldsymbol{e}_i \sim \mathcal{N}(\mathbf{0}, \xi_i^2 \mathbf{\Sigma}_e)$, which is a specific case of our heterogeneous Gaussian noise model.
- Consider the Gaussian mixture model $\boldsymbol{e}_i \overset{i.i.d.}{\sim} \sum_{m=1}^M \phi_m \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_m^*)$. Conditional on the group labels, there holds $\boldsymbol{e}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_m^*)$, where m is the group label of the i-th observation.
- Combine the above two cases together, our generic heterogeneous model also covers some mixtures of multivariate heavy-tailed distributions, such as mutivariate Student-t mixture models. In this case, conditional on the labels and heavy-tailed

scalar random variables, we have $\boldsymbol{e}_i \sim \mathcal{N}(\mathbf{0}, \xi_i^2 \boldsymbol{\Sigma}_m^*)$, where m is the label of i and ξ_i is defined the same as above.

The SNR under the homoskedastic multivariate random effects model given in Section 3.1 can be naturally extended to the heteroskedastic case. First, define the average noise covariance as $\overline{\boldsymbol{\Sigma}}_e := \frac{1}{n}\sum_{i=1}^n \boldsymbol{\Sigma}_i$. Denote $\rho^2 = \operatorname{tr}(\boldsymbol{\Sigma}_b)/q$ and $\sigma^2 = \operatorname{tr}(\overline{\boldsymbol{\Sigma}}_e)/q$, we define the SNR as $r^2 := \rho^2/(\rho^2+\sigma^2) = \operatorname{tr}(\boldsymbol{\Sigma}_b)/\operatorname{tr}(\boldsymbol{\Sigma}_b+\overline{\boldsymbol{\Sigma}}_e)$. We consider the same method-of-moments estimator of the SNR defined in (3.5), (3.6) and (3.7). In order to establish the asymptotic result under the above heteroskedastic setting, we replace Assumption 3.2.6 with the following assumption.

Assumption 3.2.7. *Assume* $\boldsymbol{E} = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n]^\top \in \mathbb{R}^{n\times p}$ *where* $\boldsymbol{e}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma}_i)$ *for* $i = 1, \ldots, n$ *and* $\boldsymbol{\Sigma}_i = (\sigma_{i,kl})_{1\le k,l,\le q} \in \mathbb{R}^{q\times q}$. *Assume* $\max_{1\le i\le n} \|\boldsymbol{\Sigma}_i\|$ *is uniformly bounded over all instances. In addition, we make the following notations*

- $\boldsymbol{\Lambda}_{kl} = \operatorname{diag}(\sigma_{1,kl}, \cdots, \sigma_{n,kl}) \in \mathbb{R}^{n\times n},\ 1 \le k, l \le q$;
- $\overline{\sigma}_{kl} = \frac{1}{n}(\sigma_{1,kl} + \cdots + \sigma_{n,kl})$;
- $\overline{\boldsymbol{\Sigma}}_e = \frac{1}{n}(\boldsymbol{\Sigma}_1 + \cdots + \boldsymbol{\Sigma}_n) = (\overline{\sigma}_{kl})_{1\le k,l\le q}$;
- $\kappa_{kl} = \frac{1}{n}\sum_{i=1}^n (\sigma_{i,kl} - \overline{\sigma}_{kl})^2$;
- $\kappa_{tot} = \sum_{1\le k,l\le q} \kappa_{kl}$.

Theorem 3.2.2 is then generalized to the following result for the heteroskedastic random effects models.

Theorem 3.2.3. *Under Assumptions 3.2.4, 3.2.5 and 3.2.7, we have*

$$n^{1/2}\boldsymbol{V}^{-1/2}\begin{bmatrix}\hat{\sigma}^2 - \sigma^2 \\ \hat{\rho}^2 - \rho^2\end{bmatrix} \Longrightarrow \mathcal{N}(0, \boldsymbol{I}_2),$$

where V_{12} and V_{22} are the same as those in Theorem 3.2.2 by replacing $\mathbf{\Sigma}_e$ with $\overline{\mathbf{\Sigma}}_e$, while

$$V_{11} = \frac{1}{(g_2 - \tau g_1^2)^2 q^2} \left(\left(2g_2^2 - 2\tau g_1^2 g_2\right) \|\overline{\mathbf{\Sigma}}_e\|_F^2 + \left(2g_2^2 + 2\tau^2 g_1^4 - 4\tau g_1^2 g_2\right) \kappa_{tot} \right. \tag{3.12}$$

$$\left. + \left(4g_1^2 g_3 - 4g_1 g_2^2\right) \operatorname{tr}\left(\overline{\mathbf{\Sigma}}_e \mathbf{\Sigma}_b\right) + \left(\frac{2}{\tau} g_2^3 + \frac{2}{\tau} g_1^2 g_4 - \frac{4}{\tau} g_1 g_2 g_3\right) \|\mathbf{\Sigma}_b\|_F^2 \right). \tag{3.13}$$

Here g_k's are defined in (3.10) and κ_{tot} is defined in Assumption 3.2.7. Again, we have $n^{1/2}(\hat{r}^2 - r^2)/\sigma_r \Rightarrow \mathcal{N}(0,1)$, where σ_r has the same form as in (3.11) with the new V_{11} defined above.

Remark 3.2.6. *(Heteroskedasticity) When $\mathbf{\Sigma}_e$ is replaced by $\overline{\mathbf{\Sigma}}_e$, the expression for V_{11} in Theorem 3.2.3 includes an additional term, $(2g_2^2 + 2\tau^2 g_1^4 - 4\tau g_1^2 g_2)\,\kappa_{tot}$, compared to that in Theorem 3.2.2. In the homogeneous noise case, we have $\kappa_{\mathrm{tot}} = 0$, in which case Theorem 3.2.3 reduces exactly to the result in Theorem 3.2.2.*

Remark 3.2.7. *(Inference) Similar to Remark 3.2.5, to construct a confidence interval for r^2, we estimate the standard error by plugging n, p, q, $\hat{g}_k$, $\widehat{\overline{\mathbf{\Sigma}}}_e$, $\widehat{\mathbf{\Sigma}}_b$, and the estimated variance components $(\hat{\sigma}^2, \hat{\rho}^2)$ into the asymptotic variance formula of $\hat{r}^2$. An additional parameter required for standard error estimation is κ_{tot}. Although this quantity is difficult to estimate for general heteroskedasticity, it is estimable under certain structured noise models. In particular, we propose consistent estimators of κ_{tot} for both the scalar heterogeneity noise model and the subgroup noise model. Details are given in Section 3.3.2.*

To justify the asymptotic inference based on plugging $\widehat{\overline{\mathbf{\Sigma}}}_e$ and $\widehat{\mathbf{\Sigma}}_b$ into the asymptotic variance formula of $\hat{r}^2$, we establish the operator norm consistency of these two covariance matrix estimators under additional conditions on q.

Theorem 3.2.4. *Under Assumptions 3.2.4, 3.2.5 and 3.2.7, assuming further that $q = o(n)$, we have*

$$\left\|\widehat{\overline{\mathbf{\Sigma}}}_e - \overline{\mathbf{\Sigma}}_e\right\| = o_P(1), \;\; \text{and} \;\; \left\|\widehat{\mathbf{\Sigma}}_b - \mathbf{\Sigma}_b\right\| = o_P(1),$$

where $\widehat{\mathbf{\Sigma}}_b$ and $\widehat{\overline{\mathbf{\Sigma}}}_e$ are estimated as in (3.5) and (3.6), respectively.

Remark 3.2.8. *In the above result, requiring $q = o(n)$ is equivalent to demanding that the dimensionality of these covariance matrices grows strictly slower than the sample size. In fact, under our framework, the plug-in estimators $(\widehat{\widehat{\boldsymbol{\Sigma}}}_e, \widehat{\boldsymbol{\Sigma}}_b)$ are within estimation error that scales in $O_P(q/n)$, which vanishes in probability when $q/n \to 0$. By contrast, the asymptotic normality in Theorem 3.2.3 does not require the condition $q = o(n)$.*

3.3. Simulations

In this section, we aim to demonstrate the empirical properties of $\hat{r}^2$ and its uncertainty quantification under both fixed effects and random effects models, based on the asymptotic results established in Theorems 3.2.1, 3.2.2 and 3.2.3. All computations are carried out on an Intel Xeon 72-core CPU server.

3.3.1. Fixed Effects Models. To estimate the standard error of $\hat{r}^2$ given in Theorem 3.2.1, we need to estimate $\boldsymbol{W}_b$, $\boldsymbol{\Sigma}_e$, ρ^2 and σ^2, whose estimators are given in (3.2), (3.3) and (3.4).

We consider the following two ways to generate the $n \times p$ design matrix $\boldsymbol{X}$:

(1) The entries of $\boldsymbol{X}$ are i.i.d. standard Gaussian variables.

(2) (SNP design) The standardized genotype model proposed in [24]: First generate $f_j \sim \text{Unif}[0.05, 0.5]$ for $j = 1, \ldots, p$ independently. Then, generate a $n \times p$ matrix $\boldsymbol{U} \in \{0, 1, 2\}^{n \times p}$ with independent entries, such that each entry in the j-th column follows a discrete distribution over $\{0, 1, 2\}$ with assigned probabilities $(1 - f_j)^2$, $2f_j(1 - f_j)$ and f_j^2, respectively. Finally, the $n \times p$ matrix $\boldsymbol{X}$ is generated by standardizing each column of $\boldsymbol{U}$.

We also consider two different cases for the coefficient matrix $\boldsymbol{B}$:

(1) (Sparse Case) We generate $\boldsymbol{B}$ based on $\boldsymbol{B}_{ij} \propto 0.8^{|i-j|}$.

(2) (Dense Case) We generate $\boldsymbol{B}$ from the distribution $\mathcal{N}(\boldsymbol{0}, \frac{1}{p}\boldsymbol{I}_p \otimes \boldsymbol{\Sigma}_b)$, where $\boldsymbol{\Sigma}_b = (\sigma_{b,ij})_{q \times q} = (0.8^{|i-j|})_{q \times q}$. We then rescale $\boldsymbol{B}$ to ensure that ρ^2 fixed. We keep this $\boldsymbol{B}$ fixed across all 500 Monte Carlo simulations.

The noise matrix is generated based on $(\boldsymbol{\Sigma}_e)_{ij} \propto \boldsymbol{\Gamma}^{1/2}(0.5^{|i-j|})_{q\times q}\boldsymbol{\Gamma}^{1/2}$, where the diagonal matrix $\boldsymbol{\Gamma}$ consists of diagonal entries being a random permutation of $(1, 2^{-0.5}, \dots, q^{-0.5})$.

In all simulation settings, we fix $\rho^2 = \mathrm{tr}(\boldsymbol{B}^\top\boldsymbol{B})/q = 1$ and $\sigma^2 = \mathrm{tr}(\boldsymbol{\Sigma}_e)/q = 0.5$, which implies a true signal-to-noise ratio of $r^2 = 0.667$. We conduct separate simulations for sparse and dense coefficient matrices $\boldsymbol{B}$, varying the sample size n from 200 to 5000 and the dimension p from 100 to 1000, while keeping $q = 20$ fixed. The results are reported in Table 3.1 and Table 3.2. For each configuration, we perform 500 Monte Carlo replications.

The simulation results demonstrate that the estimator $\hat{r}^2$, defined in (3.8), is consistent and that the nominal 95% confidence intervals achieve satisfactory coverage of the true r^2 value when n is sufficiently large and n and p grow proportionally. Furthermore, although the theoretical guarantees in Theorem 3.2.1 assume a standard Gaussian design for $\boldsymbol{X}$, the estimator also performs well when the entries of $\boldsymbol{X}$ follow an SNP-like distribution. These findings suggest that it may be possible to extend our theoretical framework to accommodate sub-Gaussian design matrices.

3.3.2. Random Effects Models. For our simulations under the random effects models with either homoskedastic or heteroskedastic noise, the design matrix $\boldsymbol{X}$ is generated as $\boldsymbol{X} = \boldsymbol{Z}\boldsymbol{\Sigma}^{1/2}$, where the $n \times p$ matrix $\boldsymbol{Z}$ either follows the previously mentioned SNP design or has i.i.d. standardized t_7 entries. We always set $\boldsymbol{\Sigma} = (\sigma_{ij}) = (0.5^{|i-j|})p \times p$. The random coefficient matrix $\boldsymbol{B}$ is generated with covariance $\boldsymbol{\Sigma}_b = (0.8^{|i-j|})_{q\times q}$.

The above setting implies that $\rho^2 = \mathrm{tr}(\boldsymbol{\Sigma}_b)/q = 1$. In all subsequent noise generation settings, we fix $\sigma^2 = 0.5$, which further implies $r^2 = 0.667$. Simulated experiments are conducted for various values of n and p, as shown in Tables 3.3, 3.4, and 3.5, with $q = 20$ held fixed. For each simulation configuration, we perform 500 Monte Carlo replications.

In following three subsections, we will present results of homogeneity model, scalar heterogeneity and subgroup models, and details of estimating the asymptotic variance of $\hat{r}^2$ as in (3.11) and (3.12), especially the estimator of κ_{tot} in scalar heterogeneity and subgroup models.

Table 3.1. SNR estimation under fixed effects model and sparse $\boldsymbol{B}$. We fix $q = 20$. The columns provide: (1) the sample size n and dimension p for each setting, (2) the average of SNR estimates $\hat{r}^2$, (3) the empirical standard error of $\hat{r}^2$, (4) the average of estimated standard errors of $\hat{r}^2$, and (5) the coverage probability of the nominal 95% confidence intervals for r^2. Results are shown for the design matrix generated using standard Gaussian and SNP distributions across various n and p configurations.

	n, p	Mean	Emp.se (10^{-2})	Ave.$\hat{\text{se}}$ (10^{-2})	Coverage
Gaussian	200,100	0.663	6.80	7.20	96.8%
	400,100	0.664	4.80	4.71	94.2%
	100,1000	0.654	23.9	25.5	97.4%
	500,1000	0.665	5.82	6.07	95.6%
	1000,1000	0.667	3.67	3.60	96.0%
	2000,1000	0.667	2.15	2.26	95.2%
	5000,1000	0.666	1.27	1.31	94.2%
SNP	200,100	0.650	6.78	7.16	95.2%
	400,100	0.657	4.36	4.70	96.6%
	100,1000	0.597	25.3	25.4	95.2%
	500,1000	0.654	6.07	6.06	94.4%
	1000,1000	0.662	3.54	3.59	95.4%
	2000,1000	0.667	2.15	2.26	96.6%
	5000,1000	0.667	1.19	1.31	96.4%

Table 3.2. SNR estimation under fixed effects model and dense $\boldsymbol{B}$.

	n, p	Mean	Emp.se (10^{-2})	Ave.$\hat{\text{se}}$ (10^{-2})	Coverage
Gaussian	200,100	0.642	5.47	5.51	94.4%
	400,100	0.668	3.12	3.12	95.6%
	100,1000	0.653	20.0	19.9	95.6%
	500,1000	0.666	4.31	4.43	95.4%
	1000,1000	0.668	2.46	2.57	96.4%
	2000,1000	0.666	1.52	1.58	96.4%
	5000,1000	0.666	0.944	0.903	93.8%
SNP	200,100	0.653	4.36	4.91	96.6%
	400,100	0.669	2.97	3.10	95.0%
	100,1000	0.608	19.0	19.8	94.2%
	500,1000	0.662	4.37	4.44	96.2%
	1000,1000	0.663	2.54	2.56	95.4%
	2000,1000	0.665	1.65	1.58	95.0%
	5000,1000	0.667	0.937	0.905	93.6%

3.3.2.1. *Homoskedastic Random Effects Models.* In the homoskedastic random effects models, the noise is generated in essentially the same way as in Section 3.3.1, with an additional scaling to ensure that $\sigma^2 = \mathrm{tr}(\boldsymbol{\Sigma}_e)/q = 0.5$.

Table 3.3 evaluates the performance of the SNR estimator $\hat{r}^2$ under the random effects model with homogeneous noise for design matrices drawn from SNP and t_7 distributions. Across all (n, p) configurations, the average $\hat{r}^2$ is nearly unbiased (maximum deviation < 0.005). The empirical standard error of $\hat{r}^2$ contracts from approximately 0.051 at $(200, 100)$ to 0.010 at $(5000, 1000)$, closely matching the average estimated standard error in every setting. Nominal 95% confidence intervals achieve coverage between 95.0% and 96.2% for SNP designs, and between 93.2% and 96.8% for t_7 designs, with slight undercoverage ($\approx$ 94%) only at the largest sample sizes under heavy-tailed covariates.

TABLE 3.3. SNR estimation under random effects model with homogeneous noise.

	n, p	Mean	Emp.se (10^{-2})	Ave.$\hat{\text{se}}$ (10^{-2})	Coverage
SNP	200,100	0.662	5.10	5.45	96.0%
	400,100	0.665	3.54	3.66	96.2%
	500,1000	0.663	4.26	4.44	95.2 %
	1000,1000	0.666	2.54	2.60	95.4%
	2000,1000	0.666	1.63	1.64	95.0%
	5000,1000	0.666	1.01	1.01	95.4%
t_7	200,100	0.665	6.90	7.83	96.8%
	400,100	0.666	5.96	6.29	95.6%
	500,1000	0.663	4.72	4.45	94.4 %
	1000,1000	0.667	2.94	2.99	95.2%
	2000,1000	0.665	2.24	2.22	93.2%
	5000,1000	0.666	1.80	1.74	93.8%

3.3.2.2. *Scalar Heterogeneity Model.* Recall that in the heteroskedastic cases, we assume the individual noise satisfies $\boldsymbol{e}_i \sim \mathcal{N}(0, \boldsymbol{\Sigma}_i)$ for $i = 1, \ldots, n$. However, it is difficult or perhaps impossible to estimate κ_{tot} for such a generic setting of heterogeneity. Here we consider a particular case, where there is a simple moment estimator for κ_{tot}. On the other hand, this particular model of heterogeneity is related to some important examples of non-Gaussian noise, such as multivariate t-distributions.

In the scalar heterogeneity model, we assume that the individual noise covariance matrices satisfy $\mathbf{\Sigma}_i = \nu_i \mathbf{\Sigma}_e$, where $\mathbf{\Sigma}_e = (\sigma_{kl})_{1\le k,l\le q}$ is generated in the same way as before. Let $\boldsymbol{V}_n = \mathrm{diag}(\nu_1, \nu_2, \ldots, \nu_n)$, where $\nu_i = \theta|w_i|$ and $w_i \overset{i.i.d.}{\sim} \mathcal{N}(0,9)$. The scaling factor θ is chosen such that $\mathrm{tr}(\boldsymbol{V}_n) = n$, which implies that $\overline{\mathbf{\Sigma}}_e = \mathbf{\Sigma}_e$.

It turns out that the total degree heterogeneity has a simple form under this model. If we denote the heteroskedasticity parameter as $\eta = \frac{1}{n}\sum_{i=1}^n (\nu_i - 1)^2$, straightforward calculation yields $\kappa_{tot} = \sum_{1\le k,l\le q} \frac{1}{n}\sum_{i=1}^n (\nu_i\sigma_{kl} - \sigma_{kl})^2 = \eta\|\mathbf{\Sigma}_e\|_F^2$. Given $\mathbf{\Sigma}_e = \overline{\mathbf{\Sigma}}_e$ can be estimated as in (3.6), the estimation of κ_{tot} reduces to the estimation of η.

We now introduce an estimator of η based on a method of moments. It is easy to obtain

$$\begin{aligned}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \left(\boldsymbol{y}_i^\top \boldsymbol{y}_i\right)^2 |\boldsymbol{X}\right] &= (2\|\mathbf{\Sigma}_e\|_F^2 + q^2\sigma^4)(\eta+1) + \frac{2}{np^2}\sum_{i=1}^n \|\boldsymbol{x}_i\|_2^4 \|\mathbf{\Sigma}_b\|_F^2 \\ &+ \frac{4}{np}\,\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{V}_n\right)\mathrm{tr}(\mathbf{\Sigma}_b\mathbf{\Sigma}_e) \\ &+ \frac{1}{np^2}\sum_{i=1}^n \|\boldsymbol{x}_i\|_2^4 q^2\rho^4 + \frac{2}{np}\,\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{V}_n\right) q^2\sigma^2\rho^2.\end{aligned}$$

From Lemma 3.5.0.1 in the Appendix, We actually have the approximation $\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{V}_n\right) \approx \mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top\right)$, which implies the following method-of-moments estimator of heteroskedasticity parameter η

$$\begin{aligned}\hat{\eta} := \frac{1}{2\|\widehat{\overline{\mathbf{\Sigma}}}_e\|_F^2 + q^2\hat{\sigma}^4}\Bigg(&\frac{1}{n}\sum_{i=1}^n \left(\boldsymbol{y}_i^\top \boldsymbol{y}_i\right)^2 - \frac{2}{np^2}\sum_{i=1}^n \|\boldsymbol{x}_i\|_2^4 \|\widehat{\mathbf{\Sigma}}_b\|_F^2 \\ &- \frac{4}{np}\,\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top\right)\mathrm{tr}(\widehat{\mathbf{\Sigma}}_b\widehat{\overline{\mathbf{\Sigma}}}_e) - \frac{1}{np^2}\sum_{i=1}^n \|\boldsymbol{x}_i\|_2^4 q^2\hat{\rho}^4 - \frac{2}{np}\,\mathrm{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top\right) q^2\hat{\sigma}^2\hat{\rho}^2\Bigg) - 1. \qquad (3.14)\end{aligned}$$

Based on the proofs of Theorems 3.2.3 and 3.2.4, it is easy to show that under certain mild assumptions, $\hat{\eta}$ is a consistent estimator of η. Consequently, we estimate κ_{tot} by $\hat{\kappa}_{tot} := \hat{\eta}\|\widehat{\overline{\mathbf{\Sigma}}}_e\|_F^2$.

Table 3.4 examines the SNR estimator $\hat{r}^2$ under the scalar heterogeneity model (with $q = 20$), using SNP and t_7 design matrices across varying (n, p). The mean $\hat{r}^2$ remains essentially unbiased—ranging from 0.662 at $(200, 100)$ to 0.666 at $(5000, 1000)$ under SNP and similarly under t_7. The empirical standard error of $\hat{r}^2$ decreases from about 0.050 at $(200, 100)$ to approximately 0.018 at $(5000, 1000)$, and in each setting the average plug-in standard error tracks the empirical value closely. Nominal 95% confidence intervals achieve coverage between 95.2% and 96.8% under SNP, and between 94.6% and 96.6% under t_7, with only minor undercoverage (around 94%) in the largest heavy-tailed scenarios. In addition, relative to Table 3.3, the empirical standard error under the scalar heterogeneity model is uniformly larger, reflecting the extra heteroskedasticity contribution κ_{tot} identified in Remark 3.2.6.

TABLE 3.4. SNR estimation under scalar heterogeneity model.

	n, p	Mean	Emp.se (10^{-2})	Ave.$\hat{\text{se}}$ (10^{-2})	Coverage
SNP	200,100	0.662	5.24	5.46	95.8%
	400,100	0.664	3.57	3.67	96.2%
	500,1000	0.664	4.30	4.43	95.2 %
	1000,1000	0.666	2.50	2.60	95.4%
	2000,1000	0.666	1.59	1.64	95.0%
	5000,1000	0.666	1.00	1.00	95.2%
t_7	200,100	0.664	6.76	7.85	97.0%
	400,100	0.663	5.62	6.28	96.8%
	500,1000	0.662	0.473	4.45	93.8 %
	1000,1000	0.666	2.83	2.99	95.6%
	2000,1000	0.665	2.23	2.22	94.0%
	5000,1000	0.666	1.81	1.74	94.0%

3.3.2.3. *Subgroup Model.* In the second example of heterogeneous noise, we assume that all individuals fall into M distinct groups. Subgroup structures are very common in high-dimensional data analysis. In genomic studies, for example, the environmental noise may exhibit different covariance structures across different locations, based on which the individuals can be grouped.

To be specific, assume there are n_m individuals in the m-th group, which implies $\sum_{m=1}^M n_m = n$. The model in each group is represented as

$$\boldsymbol{Y}^{(m)} = \boldsymbol{X}^{(m)}\boldsymbol{B} + \boldsymbol{E}^{(m)}, \text{ for } m = 1, ..., M.$$

Moreover, in each subgroup, we assume the the noise satisfies the scalar heterogeneity model discussed in Section 3.3.2.2, i.e. $\boldsymbol{E}^{(m)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{V}_n^{(m)} \otimes \boldsymbol{\Sigma}_e^{(m)})$, with the corresponding heterogeneity parameter denoted as $\eta^{(m)}$.

By denoting $r_m = n_m/n$, this model implies $\overline{\boldsymbol{\Sigma}}_e = \sum_{m=1}^M r_m \boldsymbol{\Sigma}_e^{(m)}$, and further

$$\kappa_{tot} = \sum_{m=1}^M r_m \left\| \boldsymbol{\Sigma}_e^{(m)} - \overline{\boldsymbol{\Sigma}}_e \right\|_F^2 + \sum_{m=1}^M r_m \eta^{(m)} \left\| \boldsymbol{\Sigma}_e^{(m)} \right\|_F^2.$$

Note that in each subgroup, the relevant parameters can be estimated by the method for the scalar heterogeneity model. Therefore, we get the estimator of κ_{tot} as

$$\hat{\kappa}_{tot} := \sum_{m=1}^M r_m \left\| \widehat{\boldsymbol{\Sigma}}_e^{(m)} - \widehat{\overline{\boldsymbol{\Sigma}}}_e \right\|_F^2 + \sum_{m=1}^M r_m \hat{\eta}^{(m)} \left\| \widehat{\boldsymbol{\Sigma}}_e^{(m)} \right\|_F^2.$$

We can also show the consistency of $\hat{\kappa}_{tot}$ under certain mild assumptions including fixed M and fixed r_m for $m = 1, \ldots, M$. Here we omit the proof. Specifically, assume the observations are evenly divided into $M = 10$ groups. For the m-th group, the noise matrix is generated as $\boldsymbol{E}^{(m)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I} \otimes \boldsymbol{\Sigma}_e^{(m)})$, i.e., each group satisfies residual homoskedasticity. Set $\boldsymbol{\Sigma}_e^{(m)} = \theta \boldsymbol{\Gamma}^{1/2} (\phi_m^{|i-j|})_{q\times q} \boldsymbol{\Gamma}^{1/2}$, where $\boldsymbol{\Gamma}$ is generated in the same way as it in Section 3.3.2.2, $\phi_m \overset{i.i.d.}{\sim} \text{Unif}[0.2, 0.6]$, and θ is chosen such that $\sigma^2 = \text{tr}(\overline{\boldsymbol{\Sigma}}_e)/q = 0.5$.

Table 3.5 evaluates the SNR estimator $\hat{r}^2$ under the subgroup heterogeneity model (with $q = 20$), again comparing SNP and t_7 design matrices across various (n, p). The average $\hat{r}^2$ remains essentially unbiased in all regimes—deviations stay below 0.005, for instance from 0.664 at $(2000, 100)$ to 0.666 at $(20000, 1000)$ under SNP. Coverage of nominal 95% confidence intervals lies between 94.0% and 96.6% under SNP, and between 93.4% and 95.6% under t_7, with only slight undercoverage (around 94%) in the largest heavy-tailed configurations.

TABLE 3.5. SNR estimation under subgroup model.

	n, p	Mean	Emp.se (10^{-2})	Ave.$\hat{\text{se}}$ (10^{-2})	Coverage
SNP	2000,100	0.664	1.87	1.99	96.6%
	4000,100	0.666	1.77	1.73	94.0%
	10000,1000	0.666	0.772	0.758	94.0%
	20000,1000	0.666	0.601	0.616	94.6%
t_7	2000,100	0.665	4.70	5.02	95.6%
	4000,100	0.663	4.66	4.84	94.8%
	10000,1000	0.666	1.66	1.57	93.4%
	20000,1000	0.665	1.52	1.49	94.2%

3.3.3. Performance. In all the above data generation setups, we consistently have $\rho^2 = 1$ and $\sigma^2 = 0.5$, which means the true SNR is always set as $r^2 = 0.667$. Simulated experiments are carried out for various values of n and p as shown in Table 3.1 to 3.5, while $q = 20$ is fixed. In every simulation configuration, we perform 500 Monte Carlo simulations. We report the empirical means and standard errors of $\hat{r}^2$, the average estimated standard errors based on the proposed methods given in Section 3.2, and the empirical coverage results for the asymptotic 95% confidence intervals.

From our simulation results, we observe that $\hat{r}^2$ is in general a consistent estimator of the true SNR. Further, the average estimated standard errors align with the empirical standard errors over most configurations of n and p, unless p is much greater than n, such as $n = 100$ and $p = 1000$ in the Scalar Heterogeneity model. Similarly, the nominal 95% confidence intervals exhibit desirable empirical properties for most cases.

3.4. Proof of Theorem 3.2.1

In this section, we give a proof for Theorem 3.2.1 and the proofs of Lemma 3.4.0.2 and Lemma 3.4.0.3 are deferred to the Appendix. Let $\hat{\theta} = (\hat{\sigma}^2, \hat{\rho}^2)$ and let $\boldsymbol{S} = \left(\frac{1}{n}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{Y}), \frac{1}{n^2}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y})\right)$, then $\hat{\theta} = \boldsymbol{A}\boldsymbol{S}$ where

$$\boldsymbol{A} = \begin{pmatrix} \frac{p+n+1}{q(n+1)} & -\frac{n}{q(n+1)} \\ -\frac{p}{q(n+1)} & \frac{n}{q(n+1)} \end{pmatrix}. \tag{3.15}$$

It follows that $\operatorname{cov}(\hat{\theta}) = \boldsymbol{A}\operatorname{cov}(\boldsymbol{S})\boldsymbol{A}^\top$. To compute the covariance of $\hat{\sigma}^2$ and $\hat{\rho}^2$, we could compute the covariance of S first, which will use the following lemma

LEMMA 3.4.0.1. *Suppose that $\boldsymbol{X}$ is an $n \times p$ matrix with iid entris $x_{ij} \sim \mathcal{N}(0,1)$, then $\boldsymbol{W} = \boldsymbol{X}^\top\boldsymbol{X}$ is a Wishart$(n, \boldsymbol{I}_p)$ random matrix. Let $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^p$, then we have*

$$\mathbb{E}\left[\boldsymbol{\alpha}^\top\boldsymbol{W}\boldsymbol{\alpha}\boldsymbol{\beta}^\top\boldsymbol{W}\boldsymbol{\beta}\right] = 2n(\boldsymbol{\alpha}^\top\boldsymbol{\beta})^2 + n^2\|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2 \tag{3.16}$$

$$\mathbb{E}\left[\boldsymbol{\alpha}^\top\boldsymbol{W}\boldsymbol{\alpha}\boldsymbol{\beta}^\top\boldsymbol{W}^2\boldsymbol{\beta}\right] = (2n + 2np + 4n^2)(\boldsymbol{\alpha}^\top\boldsymbol{\beta})^2 + (4n + n^2 + n^2p + n^3)\|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2 \tag{3.17}$$

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\alpha}^\top\boldsymbol{W}^2\boldsymbol{\alpha}\boldsymbol{\beta}^\top\boldsymbol{W}^2\boldsymbol{\beta}\right] &= (2np^2 + 10n^2p + 8n^3 + 8np + 4n^2 + 20n)(\boldsymbol{\alpha}^\top\boldsymbol{\beta})^2\\ &\quad + (n^2p^2 + n^4 + 2n^3p + 2n^2p + 2n^3 + 27n^2 + 8np + 10n)\|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2.\end{aligned} \tag{3.18}$$

The covariance result of S is given in the next lemma

LEMMA 3.4.0.2. *We have*

$$Var\left(\frac{1}{n}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{Y})\right) = \frac{2}{n}\|\boldsymbol{B}^\top\boldsymbol{B}\|_F^2 + \frac{4}{n}\sum_{i=1}^{q}\sigma_{ii}^2\|\boldsymbol{\beta}_i\|^2 + \frac{2}{n}\operatorname{tr}(\boldsymbol{\Sigma}_e^2). \tag{3.19}$$

$$\begin{aligned}Var\left(\frac{1}{n^2}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y})\right) &= \frac{2}{n}\left(\left(\frac{p}{n}\right)^2 + \frac{p}{n} + \frac{p}{n^2}\right)\operatorname{tr}(\boldsymbol{\Sigma}_e^2) + \frac{2}{n}\frac{p}{n^2}\operatorname{tr}^2(\boldsymbol{\Sigma}_e)\\ &\quad + \frac{2}{n}\left(4 + \frac{2}{n} + \frac{5p}{n} + \frac{p^2}{n^2} + \frac{4p}{n^2} + \frac{10}{n^2}\right)\|\boldsymbol{B}^\top\boldsymbol{B}\|_F^2\\ &\quad + \frac{2}{n}\left(\frac{13}{n} + \frac{4p}{n^2} + \frac{5}{n^2}\right)\|\boldsymbol{B}\|_F^4\\ &\quad + \frac{2}{n}\left(2\frac{p^2}{n^2} + \frac{6p}{n} + \frac{6p}{n^2} + 2 + \frac{6}{n} + \frac{8}{n^2}\right)\sum_{i=1}^{p}\sigma_{ii}^2\|\boldsymbol{\beta}_i\|^2\\ &\quad + \frac{2}{n}\left(\frac{4p}{n^2} + \frac{4}{n} + \frac{4}{n^2}\right)\|\boldsymbol{B}\|_F^2\operatorname{tr}(\boldsymbol{\Sigma}_e)\end{aligned} \tag{3.20}$$

$$Cov\left(\frac{1}{n}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{Y}), \frac{1}{n^2}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y})\right) = \frac{2}{n}\left\{\left(\frac{1}{n}+\frac{p}{n}+2\right)\|\boldsymbol{B}^\top\boldsymbol{B}\|_F^2 + \frac{2}{n}\|\boldsymbol{B}\|_F^4 + \left(\frac{2}{n}+2+\frac{2p}{n}\right)\sum_{i=1}^p \sigma_{ii}^2\|\boldsymbol{\beta}_i\|^2 + \frac{1}{n}\|\boldsymbol{B}\|_F^2\operatorname{tr}(\boldsymbol{\Sigma}_e) + \frac{p}{n}(\operatorname{tr}(\boldsymbol{\Sigma}_e))^2\right\} \tag{3.21}$$

The detailed proof of Lemma 3.4.0.1 and Lemma 3.4.0.2 are shown in the Supplementary Material. Asymptotic result for the entries of $\operatorname{cov}(\hat{\theta})$ follow directly from Lemma 3.4.0.1 and Lemma 3.4.0.2.

Lemma 3.4.0.3.

$$Var(\hat{\sigma}^2) = \frac{2}{q^2(n+1)^2 n}\Big\{\left(n^2+np-2n+2p+9\right)\|\boldsymbol{B}^\top\boldsymbol{B}\|_F^2 + (9n+1)\|\boldsymbol{B}\|_F^4 + (2pn+2p+2n+6)\sum_{i=1}^p \sigma_{ii}^2\|\boldsymbol{\beta}_i\|^2 + (2n+2p+2)\|\boldsymbol{B}\|_F^2\operatorname{tr}(\boldsymbol{\Sigma}_e) + \left(n^2+np+2n+p+1\right)\operatorname{tr}(\boldsymbol{\Sigma}_e^2) + p\operatorname{tr}^2(\boldsymbol{\Sigma}_e)\Big\}$$

$$Var(\hat{\rho}^2) = \frac{2}{q^2(n+1)^2 n}\Big\{\left(4n^2+np+2n+2p+10\right)\|\boldsymbol{B}^\top\boldsymbol{B}\|_F^2 + (13n+5)\|\boldsymbol{B}\|_F^4 + \left(2n^2+2pn+6n+2p+8\right)\sum_{i=1}^p \sigma_{ii}^2\|\boldsymbol{\beta}_i\|^2 + (4n+2p+4)\|\boldsymbol{B}\|_F^2\operatorname{tr}(\boldsymbol{\Sigma}_e) + (pn+p)\operatorname{tr}(\boldsymbol{\Sigma}_e^2) + p\operatorname{tr}^2(\boldsymbol{\Sigma}_e)\Big\}$$

$$Cov(\hat{\sigma}^2, \hat{\rho}^2) = -\frac{2}{q^2(n+1)^2 n}\Big\{\left(2n^2+np-n+2p+9\right)\|\boldsymbol{B}^\top\boldsymbol{B}\|_F^2 + (11n+3)\|\boldsymbol{B}\|_F^4 + (2np+2n+2p+6)\sum_{i=1}^p \sigma_{ii}^2\|\boldsymbol{\beta}_i\|^2 + (3n+2p+3)\|\boldsymbol{B}\|_F^2\operatorname{tr}(\boldsymbol{\Sigma}_e) + (pn+p)\operatorname{tr}(\boldsymbol{\Sigma}_e^2) + p\operatorname{tr}^2(\boldsymbol{\Sigma}_e)\Big\}$$

By the assumption that $p(n)/n \to \tau$, we can have the asymptotic approximation

$$\mathrm{Var}(\hat{\sigma}^2) = \frac{\left\{2 + O\left(\frac{1}{n}\right)\right\}}{q^2(n+1)^2 n} \left\{ \left(n^2 + np\right) \|\boldsymbol{B}^\top \boldsymbol{B}\|_F^2 + 2pn \operatorname{tr}\left(\boldsymbol{\Sigma}_e \boldsymbol{B}^\top \boldsymbol{B}\right) + \left(n^2 + np\right) \operatorname{tr}(\boldsymbol{\Sigma}_e^2) \right\},$$

$$\mathrm{Var}(\hat{\rho}^2) = \frac{\left\{2 + O\left(\frac{1}{n}\right)\right\}}{q^2(n+1)^2 n} \left\{ \left(4n^2 + np\right) \|\boldsymbol{B}^\top \boldsymbol{B}\|_F^2 + \left(2n^2 + 2pn\right) \operatorname{tr}\left(\boldsymbol{\Sigma}_e \boldsymbol{B}^\top \boldsymbol{B}\right) + pn \operatorname{tr}(\boldsymbol{\Sigma}_e^2) \right\},$$

(3.22)

$$\mathrm{Cov}(\hat{\sigma}^2, \hat{\rho}^2) = -\frac{\left\{2 + O\left(\frac{1}{n}\right)\right\}}{q^2(n+1)^2 n} \left\{ \left(2n^2 + np\right) \|\boldsymbol{B}^\top \boldsymbol{B}\|_F^2 + 2np \operatorname{tr}\left(\boldsymbol{\Sigma}_e \boldsymbol{B}^\top \boldsymbol{B}\right) + pn \operatorname{tr}(\boldsymbol{\Sigma}_e^2) \right\}.$$

By a similar argument used to prove Theorem 1 in [7], we can know that the asymptotic behavior of the upper bound of the total variation distance $d_{TV}(h(S), w)$ where $h : \mathbb{R} \to \mathbb{R}$ is a function with continuous second order partial derivatives, $S = (n^{-1} \operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{Y}), n^{-2} \operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{X} \boldsymbol{X}^\top \boldsymbol{Y}))$ and w is the standard normal variable, is determined by the function h. For the function h considered in this paper, similar to [7], when $p(n)/n \to \tau$ we can have $d_{TV}(h(S), w) = O(n^{-1/2})$.

Therefore, by (3.22) and the asymptotic normality, Theorem 3.2.1 is proved.

3.5. Proof of Theorem 3.2.2

In this section, We give important lemmas used in this paper and the proof of Theorem 3.2.3. The proofs of these key lemmas are deferred to [28]. Note that Theorem 3.2.2 is a direct corollary of Theorem 3.2.3, so we omit its proof.

3.5.1. Supporting Lemmas. To dicuss the cases with heteroskedastic noise, we denote $\boldsymbol{T}_n = \operatorname{diag}(\nu_1, ..., \nu_n)$ as a $n \times n$ diagonal matrix, where all ν_i's are assumed to be positive and $\max_{1 \le i \le n} \nu_i = O(1)$. Denote

$$\kappa = \frac{1}{n} \sum_{i=1}^{n} (\nu_i - \overline{\nu})^2,$$

where $\overline{\nu} = \frac{1}{n} \sum_{i=1}^{n} \nu_i$. We further assume that κ is fixed. Then we introduce the key lemma that will be used repeatedly in the proof of Theorem 3.2.3.

Lemma 3.5.0.1. *Under Assumption 3.2.4, we have*

(1)

$$\frac{1}{n}\operatorname{tr}\left(\left(\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{T}_n\right)^2\right) - \frac{\overline{\nu}^2}{n}\operatorname{tr}\left(\left(\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^\top\right)^2\right) - \frac{\kappa}{n^2}\operatorname{tr}^2\left(\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^\top\right) = o_p\left(1\right). \tag{3.23}$$

(2)

$$\frac{1}{n}\operatorname{tr}\left(\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{T}_n\right) - \frac{\overline{\nu}}{n}\operatorname{tr}\left(\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^\top\right) = O_p\left(\sqrt{\frac{1}{n^2}}\right). \tag{3.24}$$

(3)

$$\frac{1}{n}\operatorname{tr}\left(\left(\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^\top\right)^k\boldsymbol{T}_n\right) - \frac{\overline{\nu}}{n}\operatorname{tr}\left(\left(\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^\top\right)^k\right) = O_p\left(\frac{1}{\sqrt{n}}\right), \ \ for\ k = 2, 3. \tag{3.25}$$

Then as far as the consistency of method-of-moments estimators $\widehat{\boldsymbol{\Sigma}}_b$ and $\widehat{\overline{\boldsymbol{\Sigma}}}_e$ is concerned, we need the following Lemmas (Lemma 3.5.0.2 to Lemma 3.5.0.3) to prove Theorem 3.2.4.

Lemma 3.5.0.2. *Under Assumptions 3.2.4, 3.2.5 and 3.2.7, there hold*

$$\mathbb{E}\left[\frac{1}{n}\boldsymbol{Y}^\top\boldsymbol{Y}|\boldsymbol{X}\right] = \overline{\boldsymbol{\Sigma}}_e + \frac{1}{p}\operatorname{tr}\left(\boldsymbol{S}_n\right)\boldsymbol{\Sigma}_b,$$
$$\mathbb{E}\left[\frac{1}{n^2}\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y}|\boldsymbol{X}\right] = \frac{1}{n^2}\left(\operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{\Lambda}_{kl}\right)\right)_{1\le k,l\le q} + \frac{1}{p}\operatorname{tr}\left(\boldsymbol{S}_n^2\right)\boldsymbol{\Sigma}_b.$$

Recall that $\boldsymbol{\Lambda}_{kl} = \operatorname{diag}(\sigma_{1,kl}, \cdots, \sigma_{n,kl})$ *as defined in Assumption 3.2.7.*

Lemma 3.5.0.3. *Under Assumption 3.2.4, we have*

$$\mathbb{E}\left[\frac{1}{p}\operatorname{tr}(\boldsymbol{S}_n^2) - \frac{1}{np}\operatorname{tr}^2(\boldsymbol{S}_n)\right] = \frac{1}{p}\operatorname{tr}(\boldsymbol{\Sigma}^2) + O\left(\frac{1}{n}\right),$$

and

$$var\left(\frac{1}{p}\operatorname{tr}(\boldsymbol{S}_n^2) - \frac{1}{np}\operatorname{tr}^2(\boldsymbol{S}_n)\right) = O\left(\frac{n^2+p^2}{n^3}\right).$$

With probability at least $1 - c(n^2 + p^2)/n^3$,

$$\frac{1}{p}\operatorname{tr}(\boldsymbol{S}_n^2) - \frac{1}{np}\operatorname{tr}^2(\boldsymbol{S}_n) \geq C.$$

3.5.2. Proof of Theorem 3.2.2. By definition, $\widehat{\sigma}^2$ and $\widehat{\rho}^2$ are linear combinations of $\operatorname{tr}\left(\boldsymbol{Y}^\top \boldsymbol{Y}\right)$ and $\operatorname{tr}\left(\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}\right)$. Therefore we first aim to introduce asymptotic results on $\operatorname{tr}\left(\boldsymbol{Y}^\top \boldsymbol{Y}\right)$ and $\operatorname{tr}\left(\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}\right)$. We define the block matrix $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda}_{kl})_{1\leq k,l\leq q}$ with

$$\boldsymbol{\Lambda}_{kl} = \operatorname{diag}(\sigma_{1,kl}, \cdots, \sigma_{n,kl}) \tag{3.26}$$

where $(\boldsymbol{\Sigma}_i)_{kl} = \sigma_{i,kl}$. This definition implies that $vec(\boldsymbol{E}) \sim \mathcal{N}(0, \boldsymbol{\Lambda})$. By $vec(\boldsymbol{B}) \sim \mathcal{N}(\boldsymbol{0}, \frac{1}{p}\boldsymbol{I}_p \otimes \boldsymbol{\Sigma}_b)$, there exist some $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{q(p+n)})$ such that

$$\begin{bmatrix} vec(\boldsymbol{B}) \\ vec(\boldsymbol{E}) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{p}}\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{I}_p & 0 \\ 0 & \boldsymbol{\Lambda}^{1/2} \end{bmatrix} \boldsymbol{z}.$$

We have

$$\operatorname{tr}\left(\boldsymbol{Y}^\top \boldsymbol{Y}\right) = \boldsymbol{z}^\top \boldsymbol{Q}_1 \boldsymbol{z}, \text{ where } \boldsymbol{z} \sim \mathcal{N}(0, \boldsymbol{I}_{q(n+p)}),$$

and

$$\begin{aligned} \boldsymbol{Q}_1 &= \begin{bmatrix} \boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{I}_p & 0 \\ 0 & \boldsymbol{\Lambda}^{1/2} \end{bmatrix} \begin{bmatrix} \frac{1}{p}\boldsymbol{I}_q \otimes \boldsymbol{X}^\top \boldsymbol{X} & \frac{1}{\sqrt{p}}\boldsymbol{I}_q \otimes \boldsymbol{X}^\top \\ \frac{1}{\sqrt{p}}\boldsymbol{I}_q \otimes \boldsymbol{X} & \boldsymbol{I}_q \otimes \boldsymbol{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{I}_p & 0 \\ 0 & \boldsymbol{\Lambda}^{1/2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{p}\boldsymbol{\Sigma}_b \otimes \boldsymbol{X}^\top \boldsymbol{X} & \frac{1}{\sqrt{p}}\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2} \\ \frac{1}{\sqrt{p}}\boldsymbol{\Lambda}^{1/2}\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}\right) & \boldsymbol{\Lambda} \end{bmatrix}. \end{aligned}$$

Similarly we have

$$\operatorname{tr}\left(\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}\right) = \boldsymbol{z}^\top \boldsymbol{Q}_2 \boldsymbol{z}, \quad \text{where} \quad \boldsymbol{z} \sim \mathcal{N}(0, \boldsymbol{I}_{q(n+p)}),$$

and

$$
\boldsymbol{Q}_2 = \begin{bmatrix} \boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{I}_p & 0 \\ 0 & \boldsymbol{\Lambda}^{1/2} \end{bmatrix} \begin{bmatrix} \frac{1}{p}\boldsymbol{I}_q \otimes (\boldsymbol{X}^\top \boldsymbol{X})^2 & \frac{1}{\sqrt{p}}\boldsymbol{I}_q \otimes \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \\ \frac{1}{\sqrt{p}}\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X} & \boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{I}_p & 0 \\ 0 & \boldsymbol{\Lambda}^{1/2} \end{bmatrix}
$$

$$
= \begin{bmatrix} \frac{1}{p}\boldsymbol{\Sigma}_b \otimes \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^2 & \frac{1}{\sqrt{p}}\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2} \\ \frac{1}{\sqrt{p}}\boldsymbol{\Lambda}^{1/2}\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\right) & \boldsymbol{\Lambda}^{1/2}\left(\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2} \end{bmatrix}.
$$

Conditional Mean Analysis. To get the conditional mean of $\operatorname{tr}\left(\boldsymbol{Y}^\top \boldsymbol{Y}\right)$ and $\operatorname{tr}\left(\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}\right)$, we will introduce the useful lemma:

LEMMA 3.5.0.4 (Lemma S8 in [9]). *Let $\boldsymbol{z} \in \mathbb{R}^p$ be a random vector with i.i.d. entries with finite fourth moments satisfying $\mathbb{E}[z_i] = 0$ and $var[z_i] = 1$, and let $\boldsymbol{x} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{z}$ for a fixed positive definite matrix $\boldsymbol{\Sigma}$. Assume $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ to be a fixed symmetric matrix. We have*

$$
\mathbb{E}\left[\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}\right] = \operatorname{tr}(\boldsymbol{A}\boldsymbol{\Sigma}),
$$

$$
\mathbb{E}\left[(\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x})^2\right] = (\mathbb{E}[z_i^4] - 3)\sum_{i=1}^{p}(\boldsymbol{\Sigma}^{1/2}\boldsymbol{A}\boldsymbol{\Sigma}^{1/2})_{ii}^2 + 2\operatorname{tr}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}) + \operatorname{tr}^2(\boldsymbol{A}\boldsymbol{\Sigma}),
$$

and

$$
var(\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}) = (\mathbb{E}[z_i^4] - 3)\sum_{i=1}^{p}(\boldsymbol{\Sigma}^{1/2}\boldsymbol{A}\boldsymbol{\Sigma}^{1/2})_{ii}^2 + 2\operatorname{tr}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}\boldsymbol{\Sigma}).
$$

Here $(\boldsymbol{\Sigma}^{1/2}\boldsymbol{A}\boldsymbol{\Sigma}^{1/2})_{ii}$ is the i-th diagonal entry of $\boldsymbol{\Sigma}^{1/2}\boldsymbol{A}\boldsymbol{\Sigma}^{1/2}$.

By Lemma 3.5.0.4 we can have

$$
\mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{Y}^\top \boldsymbol{Y}\right)|\boldsymbol{X}\right] = \operatorname{tr}\left(\boldsymbol{Q}_1\right) = \operatorname{tr}\left(\frac{1}{p}\boldsymbol{\Sigma}_b \otimes \boldsymbol{X}^\top \boldsymbol{X}\right) + \operatorname{tr}\left(\boldsymbol{\Lambda}\right) = \frac{1}{p}\operatorname{tr}\left(\boldsymbol{X}^\top \boldsymbol{X}\right)q\rho^2 + nq\sigma^2,
$$

and

$$
\begin{aligned}
&\mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y} | \boldsymbol{X}\right)\right] \\
&= \operatorname{tr}\left(\boldsymbol{Q}_2\right) = \operatorname{tr}\left(\frac{1}{p}\boldsymbol{\Sigma}_b \otimes \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^2\right) + \operatorname{tr}\left(\boldsymbol{\Lambda}^{1/2}\left(\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2}\right).
\end{aligned}
$$

By properties of Kronecker product and the definition of ρ^2, we have that

$$
\operatorname{tr}\left(\frac{1}{p}\boldsymbol{\Sigma}_b \otimes \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^2\right) = \frac{1}{p}\operatorname{tr}\left((\boldsymbol{X}^\top \boldsymbol{X})^2\right) q\rho^2.
$$

By the fact

$$
\sum_{k=1}^{q} \boldsymbol{\Lambda}_{kk} = \operatorname{diag}\left(\operatorname{tr}(\boldsymbol{\Sigma}_1), ..., \operatorname{tr}(\boldsymbol{\Sigma}_n)\right),
$$

we have that

$$
\begin{aligned}
&\operatorname{tr}\left(\boldsymbol{\Lambda}^{1/2}\left(\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2}\right) \\
&= \operatorname{tr}\left(\left(\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}\right) \\
&= \operatorname{tr}\left(\begin{bmatrix} \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{\Lambda}_{11} & \cdots & \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{\Lambda}_{1q} \\ \vdots & \ddots & \vdots \\ \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{\Lambda}_{q1} & \cdots & \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{\Lambda}_{qq} \end{bmatrix}\right) \\
&= \sum_{k=1}^{q} \operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{\Lambda}_{kk}\right) \\
&= \operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{D}_n\right) q\sigma^2,
\end{aligned}
$$

where $\boldsymbol{D}_n = \operatorname{diag}\left(\operatorname{tr}(\boldsymbol{\Sigma}_1), ..., \operatorname{tr}(\boldsymbol{\Sigma}_n)\right)/(q\sigma^2)$. Therefore

$$
\mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y} | \boldsymbol{X}\right)\right] = \frac{1}{p}\operatorname{tr}\left((\boldsymbol{X}^\top \boldsymbol{X})^2\right) q\rho^2 + \operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{D}_n\right) q\sigma^2.
$$

In fact, since $\widehat{\sigma}^2$, and $\widehat{\rho}^2$ are taking linear combination of $\operatorname{tr}\left(\boldsymbol{Y}^\top\boldsymbol{Y}\right)$ and $\operatorname{tr}\left(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y}\right)$, the corresponding conditional expectation follows as

$$\mathbb{E}\left[\widehat{\sigma}^2|\boldsymbol{X}\right] = \sigma^2 + \frac{1}{\frac{1}{p}\operatorname{tr}(\boldsymbol{S}_n^2) - \frac{1}{np}\operatorname{tr}^2(\boldsymbol{S}_n)}\frac{1}{p}\operatorname{tr}(\boldsymbol{S}_n)\left(\frac{1}{n}\operatorname{tr}(\boldsymbol{S}_n) - \frac{1}{n}\operatorname{tr}(\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{D}_n)\right)\sigma^2,$$
$$\mathbb{E}\left[\widehat{\rho}^2|\boldsymbol{X}\right] = \rho^2 + \frac{1}{\frac{1}{p}\operatorname{tr}(\boldsymbol{S}_n^2) - \frac{1}{np}\operatorname{tr}^2(\boldsymbol{S}_n)}\left(\frac{1}{n}\operatorname{tr}(\boldsymbol{S}_n) - \frac{1}{n}\operatorname{tr}(\frac{1}{n}\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{D}_n)\right)\sigma^2.$$

By (3.10), we have

$$\frac{1}{p}\operatorname{tr}(\boldsymbol{S}_n) = O_p(1) \quad \text{and} \quad \frac{1}{p}\operatorname{tr}(\boldsymbol{S}_n^2) - \frac{1}{np}\operatorname{tr}^2(\boldsymbol{S}_n) = O_p(1).$$

Combining Lemma 3.5.0.1, we have

$$\begin{cases} \mathbb{E}\left[\widehat{\sigma}^2|\boldsymbol{X}\right] = \sigma^2 + O_p(\frac{1}{n}) \\ \mathbb{E}\left[\widehat{\rho}^2|\boldsymbol{X}\right] = \rho^2 + O_p(\frac{1}{n}). \end{cases} \tag{3.27}$$

Conditional Variance Analysis. Denote

$$\tilde{\boldsymbol{D}}_n = \operatorname{diag}\left(\operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_b), ..., \operatorname{tr}(\boldsymbol{\Sigma}_n\boldsymbol{\Sigma}_b)\right) / \operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e\boldsymbol{\Sigma}_b\right),$$

and $\boldsymbol{\Sigma}_b = (\sigma_{b,kl})_{1\le k,l\le q}$. For the conditional variance, based on Lemma 3.5.0.4, since $\boldsymbol{Q}_1$ is symmetric, we have

$$\operatorname{var}\left(\operatorname{tr}\left(\boldsymbol{Y}^\top\boldsymbol{Y}\right)|\boldsymbol{X}\right) = 2\operatorname{tr}\left(\boldsymbol{Q}_1^2\right) = 2\left\|\boldsymbol{Q}_1\right\|_F^2.$$

By the form of $\boldsymbol{Q}_1$, we have that

$$\left\|\boldsymbol{Q}_1\right\|_F^2 = \left\|\frac{1}{p}\boldsymbol{\Sigma}_b\otimes\boldsymbol{X}^\top\boldsymbol{X}\right\|_F^2 + \left\|\frac{1}{\sqrt{p}}\left(\boldsymbol{\Sigma}_b^{1/2}\otimes\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2}\right\|_F^2$$
$$+ \left\|\frac{1}{\sqrt{p}}\boldsymbol{\Lambda}^{1/2}\left(\boldsymbol{\Sigma}_b^{1/2}\otimes\boldsymbol{X}\right)\right\|_F^2 + \left\|\boldsymbol{\Lambda}\right\|_F^2.$$

By the property of Kronecker product, we have

$$\left\|\frac{1}{p}\boldsymbol{\Sigma}_b \otimes \boldsymbol{X}^\top \boldsymbol{X}\right\|_F^2 = \frac{1}{p^2}\operatorname{tr}\left((\boldsymbol{X}^\top \boldsymbol{X})^2\right)\operatorname{tr}\left(\boldsymbol{\Sigma}_b^2\right).$$

By the definition of $\boldsymbol{\Lambda}$, we have

$$\begin{aligned}
&\left\|\frac{1}{\sqrt{p}}\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2}\right\|_F^2 \\
&= \frac{1}{p}\operatorname{tr}\left(\left(\boldsymbol{\Sigma}_b \otimes \boldsymbol{X}\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}\right) \\
&= \frac{1}{p}\operatorname{tr}\left(\begin{bmatrix} \sum_{k=1}^q \sigma_{b,1k}\boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{\Lambda}_{k1} & & * \\ & \ddots & \\ * & & \sum_{k=1}^q \sigma_{b,qk}\boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{\Lambda}_{kq} \end{bmatrix}\right) \\
&= \frac{1}{p}\operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top \sum_{1\le k,l\le q} \sigma_{b,lk}\boldsymbol{\Lambda}_{kl}\right) \\
&= \frac{1}{p}\operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{D}}_n\right)\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e \boldsymbol{\Sigma}_b\right).
\end{aligned}$$

The last line is due to the fact that

$$\sum_{1\le k,l\le q} \sigma_{b,lk}\boldsymbol{\Lambda}_{kl} = \operatorname{diag}\left(\operatorname{tr}(\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_b), ..., \operatorname{tr}(\boldsymbol{\Sigma}_n\boldsymbol{\Sigma}_b)\right).$$

We have

$$\|\boldsymbol{\Lambda}\|_F^2 = \sum_{i=1}^n \operatorname{tr}\left(\boldsymbol{\Sigma}_i^2\right) = n\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e^2\right) + n\sum_{1\le k,l\le q}\kappa_{kl}.$$

Due to

$$\left\|\frac{1}{\sqrt{p}}\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2}\right\|_F^2 = \left\|\frac{1}{\sqrt{p}}\boldsymbol{\Lambda}^{1/2}\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}\right)\right\|_F^2,$$

we have

$$\operatorname{var}\left(\operatorname{tr}\left(\boldsymbol{Y}^\top\boldsymbol{Y}\right)|\boldsymbol{X}\right)$$

(3.28)

$$= \frac{2}{p^2}\operatorname{tr}\left((\boldsymbol{X}^\top\boldsymbol{X})^2\right)\operatorname{tr}\left(\boldsymbol{\Sigma}_b^2\right) + \frac{4}{p}\operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top\tilde{\boldsymbol{D}}_n\right)\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e\boldsymbol{\Sigma}_b\right) + 2n\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e^2\right) + 2n\sum_{1\le k,l\le q}\kappa_{kl}.$$

Based on Lemma 3.5.0.4, since $\boldsymbol{Q}_2$ is symmetric, we have

$$\operatorname{var}\left(\operatorname{tr}\left(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y}\right)|\boldsymbol{X}\right) = 2\operatorname{tr}\left(\boldsymbol{Q}_2^2\right) = 2\left\|\boldsymbol{Q}_2\right\|_F^2.$$

By the form of $\boldsymbol{Q}_2$, we have

$$\begin{aligned}\|\boldsymbol{Q}_2\|_F^2 = &\left\|\frac{1}{p}\boldsymbol{\Sigma}_b\otimes\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^2\right\|_F^2 + \left\|\frac{1}{\sqrt{p}}\left(\boldsymbol{\Sigma}_b^{1/2}\otimes\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2}\right\|_F^2\\ &+\left\|\frac{1}{\sqrt{p}}\boldsymbol{\Lambda}^{1/2}\left(\boldsymbol{\Sigma}_b^{1/2}\otimes\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{X}\right)\right\|_F^2 + \left\|\boldsymbol{\Lambda}^{1/2}\left(\boldsymbol{I}_q\otimes\boldsymbol{X}\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2}\right\|_F^2.\end{aligned}$$

By the property of Kronecker product, we have

$$\left\|\frac{1}{p}\boldsymbol{\Sigma}_b\otimes\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^2\right\|_F^2 = \frac{1}{p^2}\operatorname{tr}\left((\boldsymbol{X}^\top\boldsymbol{X})^4\right)\operatorname{tr}\left(\boldsymbol{\Sigma}_b^2\right).$$

It is similar with the part of $\left\|\frac{1}{\sqrt{p}}\left(\boldsymbol{\Sigma}_b^{1/2}\otimes\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2}\right\|_F^2$, to show that

$$\begin{aligned}&\left\|\frac{1}{\sqrt{p}}\left(\boldsymbol{\Sigma}_b^{1/2}\otimes\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2}\right\|_F^2\\ &=\frac{1}{p}\operatorname{tr}\left(\left(\boldsymbol{\Sigma}_b\otimes\left(\boldsymbol{X}\boldsymbol{X}^\top\right)^3\right)\boldsymbol{\Lambda}\right)\\ &=\frac{1}{p}\operatorname{tr}\left(\left(\boldsymbol{X}\boldsymbol{X}^\top\right)^3\tilde{\boldsymbol{D}}_n\right)\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e\boldsymbol{\Sigma}_b\right).\end{aligned}$$

Due to the form of the matrix $\left(\left(\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}\right)^2$ and $\boldsymbol{\Lambda}_{kl} = \boldsymbol{\Lambda}_{lk}$, we have

$$\begin{aligned}
&\left\|\boldsymbol{\Lambda}^{1/2}\left(\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2}\right\|_F^2 \\
&= \operatorname{tr}\left(\left(\left(\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}\right)^2\right) \\
&= \operatorname{tr}\left(\begin{bmatrix} \sum_{k=1}^q \boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{\Lambda}_{1k}\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{\Lambda}_{k1} & & * \\ & \ddots & \\ * & & \sum_{k=1}^q \boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{\Lambda}_{qk}\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{\Lambda}_{kq} \end{bmatrix}\right) \\
&= \sum_{1\le k,l\le q} \operatorname{tr}\left((\boldsymbol{X}^\top\boldsymbol{\Lambda}_{kl}\boldsymbol{X})^2\right).
\end{aligned}$$

Since

$$\left\|\frac{1}{\sqrt{p}}\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2}\right\|_F^2 = \left\|\frac{1}{\sqrt{p}}\boldsymbol{\Lambda}^{1/2}\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{X}\right)\right\|_F^2,$$

we have

$$\begin{aligned}
&\operatorname{var}\left(\operatorname{tr}\left(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y}\right)|\boldsymbol{X}\right) \\
&= \frac{2}{p^2}\operatorname{tr}\left((\boldsymbol{X}^\top\boldsymbol{X})^4\right)\operatorname{tr}\left(\boldsymbol{\Sigma}_b^2\right) + \frac{4}{p}\operatorname{tr}\left((\boldsymbol{X}\boldsymbol{X}^\top)^3\tilde{\boldsymbol{D}}_n\right)\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e\boldsymbol{\Sigma}_b\right) \\
&\quad + 2\sum_{1\le k,l\le q}\operatorname{tr}\left((\boldsymbol{X}^\top\boldsymbol{\Lambda}_{kl}\boldsymbol{X})^2\right). \qquad (3.29)
\end{aligned}$$

For the covariance term, we use the following lemma:

LEMMA 3.5.0.5. *Let $\boldsymbol{x} \in \mathbb{R}^k$ and $\boldsymbol{y} \in \mathbb{R}^l$ be two independent random vectors satisfying $\boldsymbol{x} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_x)$ and $\boldsymbol{y} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_y)$. Let $\boldsymbol{A}_1, \boldsymbol{B}_1 \in \mathbb{R}^{k\times k}$ and $\boldsymbol{A}_2, \boldsymbol{B}_2 \in \mathbb{R}^{k\times l}$ be fixed matrices. Then we have*

$$cov(\boldsymbol{x}^\top\boldsymbol{A}_1\boldsymbol{x}, \boldsymbol{x}^\top\boldsymbol{B}_1\boldsymbol{x}) = 2\operatorname{tr}(\boldsymbol{A}_1\boldsymbol{\Sigma}_x\boldsymbol{B}_1\boldsymbol{\Sigma}_x)$$

and

$$cov(\boldsymbol{x}^\top\boldsymbol{A}_2\boldsymbol{y}, \boldsymbol{x}^\top\boldsymbol{B}_2\boldsymbol{y}) = \operatorname{tr}(\boldsymbol{\Sigma}_x\boldsymbol{A}_2\boldsymbol{\Sigma}_y\boldsymbol{B}_2^\top).$$

By Lemma 3.5.0.5 in the appendix, we have that

$$\operatorname{cov}\left(\operatorname{tr}\left(\boldsymbol{Y}^\top \boldsymbol{Y}\right), \operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}) \,|\boldsymbol{X}\right) = 2\operatorname{tr}\left(\boldsymbol{Q}_1\boldsymbol{Q}_2\right).$$

By the form of $\boldsymbol{Q}_1$ and $\boldsymbol{Q}_2$, we have

$$\begin{aligned}\operatorname{tr}(\boldsymbol{Q}_1\boldsymbol{Q}_2) &= \frac{1}{p^2}\boldsymbol{\Sigma}_b^2 \otimes \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^3 + \frac{1}{p}\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}^\top\right)\boldsymbol{\Lambda}\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\right) \\ &\quad + \frac{1}{p}\boldsymbol{\Lambda}^{1/2}\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}^\top\right)\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\right)\boldsymbol{\Lambda}^{1/2} + \boldsymbol{\Lambda}^{3/2}\left(\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2}\end{aligned}$$

By the property of Kronecker product, we have

$$\frac{1}{p^2}\boldsymbol{\Sigma}_b^2 \otimes \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^3 = \frac{1}{p^2}\operatorname{tr}\left((\boldsymbol{X}^\top \boldsymbol{X})^3\right)\operatorname{tr}(\boldsymbol{\Sigma}_b^2).$$

It is similar with the part of $\left\|\frac{1}{\sqrt{p}}\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}^\top\right)\boldsymbol{\Lambda}^{1/2}\right\|_F^2$, to show that

$$\begin{aligned}&\operatorname{tr}\left(\frac{1}{p}\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}^\top\right)\boldsymbol{\Lambda}\left(\boldsymbol{\Sigma}_b^{1/2} \otimes \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\right)\right) \\ &= \frac{1}{p}\operatorname{tr}\left(\left(\boldsymbol{\Sigma}_b \otimes \left(\boldsymbol{X}\boldsymbol{X}^\top\right)^2\right)\boldsymbol{\Lambda}\right) \\ &= \frac{1}{p}\operatorname{tr}\left(\left(\boldsymbol{X}\boldsymbol{X}^\top\right)^2 \tilde{\boldsymbol{D}}_n\right)\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e \boldsymbol{\Sigma}_b\right).\end{aligned}$$

Since $\boldsymbol{\Lambda}_{kl} = \boldsymbol{\Lambda}_{lk}$ and

$$\sum_{1 \le k,l \le q} \boldsymbol{\Lambda}_{kl}^2 = \operatorname{diag}\left(\operatorname{tr}(\boldsymbol{\Sigma}_1^2), ..., \operatorname{tr}(\boldsymbol{\Sigma}_n^2)\right),$$

we have

$$
\begin{aligned}
&\operatorname{tr}\left(\mathbf{\Lambda}^{3/2}\left(\boldsymbol{I}_q\otimes\boldsymbol{X}\boldsymbol{X}^\top\right)\mathbf{\Lambda}^{1/2}\right)\\
&=\operatorname{tr}\left(\left(\boldsymbol{I}_q\otimes\boldsymbol{X}\boldsymbol{X}^\top\right)\mathbf{\Lambda}^{2}\right)\\
&=\operatorname{tr}\left(\begin{bmatrix}\boldsymbol{X}\boldsymbol{X}^\top\sum_{k=1}^q\mathbf{\Lambda}_{1k}\mathbf{\Lambda}_{k1} & & *\\ & \ddots & \\ * & & \boldsymbol{X}\boldsymbol{X}^\top\sum_{k=1}^q\mathbf{\Lambda}_{qk}\mathbf{\Lambda}_{kq}\end{bmatrix}\right)\\
&=\operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top\operatorname{diag}\left(\operatorname{tr}(\mathbf{\Sigma}_1^2),...,\operatorname{tr}(\mathbf{\Sigma}_n^2)\right)\right).
\end{aligned}
$$

By the fact that

$$
\begin{aligned}
&\operatorname{tr}\left(\frac{1}{p}\left(\mathbf{\Sigma}_b^{1/2}\otimes\boldsymbol{X}^\top\right)\mathbf{\Lambda}\left(\mathbf{\Sigma}_b^{1/2}\otimes\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{X}\right)\right)\\
&=\operatorname{tr}\left(\frac{1}{p}\mathbf{\Lambda}^{1/2}\left(\mathbf{\Sigma}_b^{1/2}\otimes\boldsymbol{X}^\top\right)\left(\mathbf{\Sigma}_b^{1/2}\otimes\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{X}\right)\mathbf{\Lambda}^{1/2}\right),
\end{aligned}
$$

we have

$$
\begin{aligned}
&\operatorname{cov}\left(\operatorname{tr}\left(\boldsymbol{Y}^\top\boldsymbol{Y}\right),\operatorname{tr}\left(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y}\right)|\boldsymbol{X}\right)\\
&=\frac{2}{p^2}\operatorname{tr}\left((\boldsymbol{X}^\top\boldsymbol{X})^3\right)\operatorname{tr}(\mathbf{\Sigma}_b^2)+\frac{4}{p}\operatorname{tr}\left((\boldsymbol{X}\boldsymbol{X}^\top)^2\tilde{\boldsymbol{D}}_n\right)\operatorname{tr}\left(\overline{\mathbf{\Sigma}}_e\mathbf{\Sigma}_b\right)\\
&\quad+2\operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top\operatorname{diag}\left(\operatorname{tr}(\mathbf{\Sigma}_1^2),...,\operatorname{tr}(\mathbf{\Sigma}_n^2)\right)\right).
\end{aligned}
\tag{3.30}
$$

Let w_1, w_2 be $\operatorname{tr}\left(\boldsymbol{Y}^\top\boldsymbol{Y}\right),\operatorname{tr}\left(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y}\right)$ after being centered and standardized, defined as

$$
w_1=\frac{1}{n^{1/2}q}\left(\boldsymbol{z}^\top\boldsymbol{Q}_1\boldsymbol{z}-\operatorname{tr}(\boldsymbol{Q}_1)\right)\quad\text{and}\quad w_2=\frac{1}{n^{3/2}q}\left(\boldsymbol{z}^\top\boldsymbol{Q}_2\boldsymbol{z}-\operatorname{tr}(\boldsymbol{Q}_2)\right).
$$

and denote $\boldsymbol{w}=(w_1,w_2)^\top$. Combining equation (3.28) and Leave-One-Out analysis in Lemma 3.5.0.1, under Assumptions 3.2.4, 3.2.5, and 3.2.7, one could approximate traces containing $\tilde{\boldsymbol{D}}_n$ and $\boldsymbol{\Lambda}_{kl}$ to obtain approximation on $\operatorname{var}(w_1|\boldsymbol{X})$ as

$$\operatorname{var}(w_1|\boldsymbol{X}) = \frac{2}{np^2q^2}\operatorname{tr}\left((\boldsymbol{X}^\top\boldsymbol{X})^2\right)\operatorname{tr}\left(\boldsymbol{\Sigma}_b^2\right) + \frac{4}{npq^2}\operatorname{tr}\left(\boldsymbol{X}^\top\boldsymbol{X}\right)\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e\boldsymbol{\Sigma}_b\right) + \frac{2}{q^2}\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e^2\right) + \frac{2}{q^2}\sum_{1\le k,l\le q}\kappa_{kl} + o_p(1).$$

Based on (3.10), as $p/n \to \tau$, we have

$$\operatorname{var}(w_1|\boldsymbol{X}) - \left(\frac{2}{\tau q^2}g_2\|\boldsymbol{\Sigma}_b\|_F^2 + \frac{4}{q^2}g_1\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e\boldsymbol{\Sigma}_b\right) + \frac{2}{q^2}\|\overline{\boldsymbol{\Sigma}}_e\|_F^2 + \frac{2}{q^2}\sum_{1\le k,l\le q}\kappa_{kl}\right) \xrightarrow{P} 0. \tag{3.31}$$

By (3.29) and Leave-One-Out analysis in Lemma 3.5.0.1, under Assumptions 3.2.4, 3.2.5, and 3.2.7, it is similar to have

$$\operatorname{var}(w_2|\boldsymbol{X}) = \frac{2}{n^3p^2q^2}\operatorname{tr}\left((\boldsymbol{X}^\top\boldsymbol{X})^4\right)\operatorname{tr}\left(\boldsymbol{\Sigma}_b^2\right) + \frac{4}{n^3pq^2}\operatorname{tr}\left((\boldsymbol{X}\boldsymbol{X}^\top)^3\right)\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e\boldsymbol{\Sigma}_b\right) + \frac{2}{n^3q^2}\operatorname{tr}\left((\boldsymbol{X}^\top\boldsymbol{X})^2\right)\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e^2\right) + \frac{2}{n^3q^2}\operatorname{tr}^2\left(\boldsymbol{X}^\top\boldsymbol{X}\right)\sum_{1\le k,l\le q}\kappa_{kl} + o_p(1).$$

Based on (3.10), as $p/n \to \tau$, we have

$$\operatorname{var}(w_2|\boldsymbol{X}) - \left(\frac{2}{\tau q^2}g_4\|\boldsymbol{\Sigma}_b\|_F^2 + \frac{4}{q^2}g_3\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e\boldsymbol{\Sigma}_b\right) + \frac{2}{q^2}\tau g_2\|\overline{\boldsymbol{\Sigma}}_e\|_F^2 + \frac{2}{q^2}\tau^2 g_1^2\sum_{1\le k,l\le q}\kappa_{kl}\right) \xrightarrow{P} 0. \tag{3.32}$$

By (3.30) and Leave-One-Out analysis in Lemma 3.5.0.1, under Assumptions 3.2.4, 3.2.5, and 3.2.7, one could approximate $\operatorname{cov}(w_1, w_2|\boldsymbol{X})$ by

$$\operatorname{cov}(w_1, w_2|\boldsymbol{X}) = \frac{2}{n^2p^2q^2}\operatorname{tr}\left((\boldsymbol{X}^\top\boldsymbol{X})^3\right)\operatorname{tr}(\boldsymbol{\Sigma}_b^2) + \frac{4}{n^2pq^2}\operatorname{tr}\left((\boldsymbol{X}\boldsymbol{X}^\top)^2\right)\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e\boldsymbol{\Sigma}_b\right) + \frac{2}{n^2q^2}\operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top\right)\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e^2\right) + \frac{2}{n^2q^2}\operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top\right)\sum_{1\le k,l\le q}\kappa_{kl} + o_p(1).$$

Based on (3.10), as $p/n \to \tau$, we have

(3.33)

$$\operatorname{cov}(w_1, w_2|\boldsymbol{X}) - \left(\frac{2}{\tau q^2} g_3 \|\boldsymbol{\Sigma}_b\|_F^2 + \frac{4}{q^2} g_2 \operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e \boldsymbol{\Sigma}_b\right) + \frac{2}{q^2}\tau g_1 \|\overline{\boldsymbol{\Sigma}}_e\|_F^2 + \frac{2}{q^2}\tau g_1 \sum_{1\le k,l\le q} \kappa_{kl}\right) \xrightarrow{P} 0.$$

Asymptotic Distribution. By the definition of $\boldsymbol{\Lambda}$, one could obtain that $\boldsymbol{\Lambda}$ and $\operatorname{diag}(\boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_n)$ are similar. Due to the assumption that $\max_{1\le i\le n} \|\boldsymbol{\Sigma}_i\|_2 = O(1)$, we have

$$\|\boldsymbol{\Lambda}\|_2 \le \max_{1\le i\le n} \|\boldsymbol{\Sigma}_i\|_2 = O(1).$$

Then we introduce the following lemma:

LEMMA 3.5.0.6 (Theorem 6.5 in [42]). *Let $\boldsymbol{X}$ be an $n \times p$ matrix whose rows $\boldsymbol{x}_i$ are i.i.d. sub-gaussian random vectors in $\mathbb{R}^n$, $\|\boldsymbol{x}_i\|_{\psi_2} = C$, and $\operatorname{cov}(\boldsymbol{x}_i) = \boldsymbol{\Sigma}$. There are universal constants $\{c_j\}_{j=0}^3$ such that, the sample covariance matrix $\boldsymbol{S}_n = \frac{1}{n}\boldsymbol{X}^\top \boldsymbol{X}$ satisfies the bounds*

$$\mathbb{P}\left(\frac{1}{C}\|\boldsymbol{S}_n - \boldsymbol{\Sigma}\| \ge c_1\left(\sqrt{\frac{p}{n}} + \frac{p}{n}\right) + \delta\right) \le c_2 \exp\{-c_3 n \min\{\delta, \delta^2\}\}, \quad \forall \delta \ge 0.$$

LEMMA 3.5.0.7 (Theorem 6.6.1 in [38]). *Let $X_1, \ldots, X_n$ be mean-zero, symmetric, $d \times d$ random matrices such that $\|X_i\| \le C$ almost surely for all $i \in \{1, \ldots, n\}$. Then for all $t \ge 0$,*

$$\mathbb{P}\left\{\left\|\sum_{i=1}^n X_i\right\| \ge t\right\} \le 2d \exp\left\{\frac{-t^2}{2\left(\sigma^2 + Ct/3\right)}\right\},$$

where $\sigma^2 = \|\sum_{i=1}^n \mathbb{E}[X_i^2]\|$ is the norm of the matrix variance of the sum.

Therefore by Lemma 3.5.0.6 and the assumption $\|\boldsymbol{\Sigma}\|_2 = O(1)$, we have

$$\begin{aligned}\frac{1}{n^{1/2}q}\|\boldsymbol{Q}_1\| &\le \frac{1}{n^{1/2}q}\left(\|\boldsymbol{\Sigma}_b^{1/2}\| + \|\boldsymbol{\Lambda}^{1/2}\|\right)^2\left(\|\frac{1}{p}\boldsymbol{X}^\top \boldsymbol{X}\| + \frac{2}{p^{1/2}}\|\boldsymbol{X}\| + 1\right)\\ &\le \frac{2}{n^{1/2}q}\left(\|\boldsymbol{\Sigma}_b\| + \|\boldsymbol{\Lambda}\|\right)\left(\|\frac{1}{p}\boldsymbol{X}^\top \boldsymbol{X}\| + \frac{2}{p^{1/2}}\|\boldsymbol{X}\| + 1\right) = O_P\left(\frac{1}{n^{1/2}q}\right).\end{aligned}$$

The second line is due to Cauchy-Schwarz inequality. It is simlar to have that

$$
\begin{aligned}
&\frac{1}{n^{3/2}q}\|\boldsymbol{Q}_2\| \\
&\leq \frac{1}{n^{3/2}q}\left(\|\boldsymbol{\Sigma}_b^{1/2}\| + \|\boldsymbol{\Lambda}^{1/2}\|\right)^2 \left(\left\|\frac{1}{p}(\boldsymbol{X}^\top\boldsymbol{X})^2\right\| + \frac{2}{p^{1/2}}\|\boldsymbol{X}\|\|\boldsymbol{X}^\top\boldsymbol{X}\| + \|\boldsymbol{X}^\top\boldsymbol{X}\|\right) \\
&\leq \frac{2}{n^{3/2}q}\left(\|\boldsymbol{\Sigma}_b\| + \|\boldsymbol{\Lambda}\|\right) \left(\left\|\frac{1}{p}(\boldsymbol{X}^\top\boldsymbol{X})^2\right\| + \frac{2}{p^{1/2}}\|\boldsymbol{X}\|\|\boldsymbol{X}^\top\boldsymbol{X}\| + \|\boldsymbol{X}^\top\boldsymbol{X}\|\right) \\
&= O_P\left(\frac{1}{n^{1/2}q}\right).
\end{aligned}
$$

Then we introduce the following lemma to establish the normality:

LEMMA 3.5.0.8 ([9]). *Let $\zeta_1, ..., \zeta_d$ be i.i.d sub-Gaussian random variables with mean 0, variance 1, and sub-Gaussian parameter bounded by C_0. Let $\boldsymbol{\zeta} = (\zeta_1, ..., \zeta_d)^\top$ and $\boldsymbol{Q}_k$ be an $d \times d$ positive semidefinite matrix, for $k = 1, ..., K$. Define $w_k = \zeta^\top \boldsymbol{Q}_k \zeta - \mathrm{tr}(\boldsymbol{Q}_k)$, and $\boldsymbol{w} = (w_1, ..., w_K)^\top$. Let $\boldsymbol{z} \sim \mathcal{N}(0, \boldsymbol{I}_K)$ and $\boldsymbol{V} = cov(\boldsymbol{w})$. There is an absolute constant $0 < C_1 < \infty$ such that*

$$
\begin{aligned}
\left|\mathbb{E}[f(\boldsymbol{w})] - \mathbb{E}[f(\boldsymbol{V}^{1/2}\boldsymbol{z})]\right| \leq & C_1(C_0+1)^8 K^{3/2} d^{1/2} |f|_2 (\max_{k=1,...,K} \|\boldsymbol{Q}_k\|)^2 \\
& + C_1(C_0+1)^8 K^3 d |f|_3 (\max_{k=1,...,K} \|\boldsymbol{Q}_k\|)^3,
\end{aligned}
$$

for all three-times differentiable functions $f : \mathbb{R}^K \to \mathbb{R}$.

By Lemma 3.5.0.8, for all three-times differentiable functions $f : \mathbb{R}^2 \to \mathbb{R}$, and $\boldsymbol{U}_n = \mathrm{cov}(\boldsymbol{w}|\boldsymbol{X}) \in \mathbb{R}^{2\times 2}$, we have

$$
\begin{aligned}
&\left|\mathbb{E}\left[f(\boldsymbol{w})|\boldsymbol{X}\right] - \mathbb{E}\left[f(\boldsymbol{U}_n^{1/2}\boldsymbol{t})|\boldsymbol{X}\right]\right| \\
&\leq C_1(C_0+1)^8 2^{3/2} \left(q(n+p)\right)^{1/2} |f|_2 \left(\max\left\{\frac{1}{n^{1/2}q}\|\boldsymbol{Q}_1\|, \frac{1}{n^{3/2}q}\|\boldsymbol{Q}_2\|\right\}\right)^2 \\
&\quad + C_1(C_0+1)^8 2^3 q(n+p)|f|_3 \left(\max\left\{\frac{1}{n^{1/2}q}\|\boldsymbol{Q}_1\|, \frac{1}{n^{3/2}q}\|\boldsymbol{Q}_2\|\right\}\right)^3 \\
&= O_P\left(\frac{1}{n^{1/2}q^{3/2}}\right) = o_P(1). \qquad (3.34)
\end{aligned}
$$

where $\boldsymbol{t} \sim \mathcal{N}(0, \boldsymbol{I}_2)$. Based on former calculations on $\text{cov}(\boldsymbol{w}|\boldsymbol{X})$ in equations (3.31), (3.32) and (3.33), we have $U_{n,ij} - \tilde{U}_{n,ij} \xrightarrow{P} 0$, where

$$\begin{aligned}
\tilde{U}_{n,11} &= \frac{2}{\tau q^2} g_2 \|\boldsymbol{\Sigma}_b\|_F^2 + \frac{4}{q^2} g_1 \operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e \boldsymbol{\Sigma}_b\right) + \frac{2}{q^2} \|\overline{\boldsymbol{\Sigma}}_e\|_F^2 + \frac{2}{q^2} \sum_{1 \le k,l \le q} \kappa_{kl} \\
\tilde{U}_{n,12} &= \frac{2}{\tau q^2} g_3 \|\boldsymbol{\Sigma}_b\|_F^2 + \frac{4}{q^2} g_2 \operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e \boldsymbol{\Sigma}_b\right) + \frac{2}{q^2} \tau g_1 \|\overline{\boldsymbol{\Sigma}}_e\|_F^2 + \frac{2}{q^2} \tau g_1 \sum_{1 \le k,l \le q} \kappa_{kl} \\
\tilde{U}_{n,22} &= \frac{2}{\tau q^2} g_4 \|\boldsymbol{\Sigma}_b\|_F^2 + \frac{4}{q^2} g_3 \operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e \boldsymbol{\Sigma}_b\right) + \frac{2}{q^2} \tau g_2 \|\overline{\boldsymbol{\Sigma}}_e\|_F^2 + \frac{2}{q^2} \tau^2 g_1^2 \sum_{1 \le k,l \le q} \kappa_{kl}.
\end{aligned}$$

Denote $\Phi(t, s)$ as the c.d.f. of $\mathcal{N}(0, \boldsymbol{I}_2)$. Based on the expressions (3.31), (3.32) and (3.33), as well as Assumptions 2.1 and 2.4, we can see $\boldsymbol{U}_n$ satisfies $\|\boldsymbol{U}_n\| = O_P(1)$ and $\|\boldsymbol{U}_n^{-1}\| = O_P(1)$. In fact, Assumption 2.1 guarantees $g_2 g_4 - g_3^2$, $g_1 g_3 - g_2^2$ and $g_2 - g_1^2$ are lower bounded by quantities determined by the limit spectral distribution H of the predictor population covariance matrix and the aspect ratio τ. Also, the boundedness of $\|\boldsymbol{\Sigma}_i\|$ for $i = 1, \ldots, n$ actually implies $\sum_{1 \le k,l \le q} \kappa_{kl} = O(q)$. Therefore by (3.34), we have

$$\mathbb{P}\left\{\boldsymbol{U}_n^{-1/2} \boldsymbol{w} \in (-\infty, t] \times (-\infty, s] | \boldsymbol{X}\right\} \xrightarrow{p} \Phi(t, s).$$

By DCT,

$$\boldsymbol{U}_n^{-1/2} \boldsymbol{w} \Rightarrow \mathcal{N}(0, \boldsymbol{I}_2).$$

By Slutsky Theorem, we have

$$\tilde{\boldsymbol{U}}_n^{-1/2} \boldsymbol{w} \Rightarrow \mathcal{N}(0, \boldsymbol{I}_2).$$

By definition, $\widehat{\sigma}^2$ and $\widehat{\rho}^2$ are linear combinations of $\operatorname{tr}\left(\boldsymbol{Y}^\top\boldsymbol{Y}\right)$ and $\operatorname{tr}\left(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y}\right)$. Combining the forms of w_1, w_2, we have

$$\begin{aligned}&\begin{pmatrix}\widehat{\sigma}^2 - \mathbb{E}[\widehat{\sigma}^2|\boldsymbol{X}] \\ \widehat{\rho}^2 - \mathbb{E}[\widehat{\rho}^2|\boldsymbol{X}]\end{pmatrix} \\ &= n^{-1/2}\left(\frac{1}{p}\operatorname{tr}(\boldsymbol{S}_n^2) - \frac{1}{np}\operatorname{tr}^2(\boldsymbol{S}_n)\right)^{-1}\begin{bmatrix}\frac{1}{p}\operatorname{tr}(\boldsymbol{S}_n^2) & -\frac{1}{p}\operatorname{tr}(\boldsymbol{S}_n) \\ -\frac{1}{n}\operatorname{tr}(\boldsymbol{S}_n^2) & 1\end{bmatrix}\begin{bmatrix}w_1 \\ w_2\end{bmatrix}.\end{aligned}$$

Combining with (3.27), by equation (3.10) and Slutsky Theorem we obtain the asymptotic distribution for $\widehat{\sigma}^2$ and $\widehat{\rho}^2$ as following:

$$n^{1/2}\boldsymbol{V}_n^{-1/2}\begin{bmatrix}\widehat{\sigma}^2 - \sigma^2 \\ \widehat{\rho}^2 - \rho^2\end{bmatrix} \Longrightarrow \mathcal{N}(0, \boldsymbol{I}_2).$$

Here $\boldsymbol{V}_n \in \mathbb{R}^{2\times 2}$ is the asymptotic covariance matrix of $\widehat{\sigma}^2, \widehat{\rho}^2$ as

$$\boldsymbol{V}_n = \left(g_2 - \tau g_1^2\right)^{-2}\begin{bmatrix}g_2 & -g_1 \\ -\tau g_2 & 1\end{bmatrix}\tilde{\boldsymbol{U}}_n\begin{bmatrix}g_2 & -g_1 \\ -\tau g_2 & 1\end{bmatrix}^\top,$$

where

$$\begin{cases} V_{n,11} = \frac{1}{(g_2-\tau g_1^2)^2 q^2}\Big((2g_2^2 - 2\tau g_1^2 g_2)\,\|\overline{\boldsymbol{\Sigma}}_e\|_F^2 + (2g_2^2 + 2\tau^2 g_1^4 - 4\tau g_1^2 g_2)\sum_{1\le k,l\le q}\kappa_{kl} \\ \quad + (4g_1^2 g_3 - 4g_1 g_2^2)\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e\boldsymbol{\Sigma}_b\right) + \left(\frac{2}{\tau}g_2^3 + \frac{2}{\tau}g_1^2 g_4 - \frac{4}{\tau}g_1 g_2 g_3\right)\|\boldsymbol{\Sigma}_b\|_F^2\Big) \\[2ex] V_{n,22} = \frac{1}{(g_2-\tau g_1^2)^2 q^2}\left((2\tau g_2 - 2\tau^2 g_1^2)\,\|\overline{\boldsymbol{\Sigma}}_e\|_F^2 + (4\tau^2 g_1^3 + 4g_3 - 8\tau g_1 g_2)\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e\boldsymbol{\Sigma}_b\right)\right. \\ \quad \left. + \left(2\tau g_1^2 g_2 + \frac{2}{\tau}g_4 - 4g_1 g_3\right)\|\boldsymbol{\Sigma}_b\|_F^2\right) \\[2ex] V_{n,12} = V_{n,21} = \frac{1}{(g_2-\tau g_1^2)^2 q^2}\left((-2\tau g_1 g_2 + 2\tau^2 g_1^3)\,\|\overline{\boldsymbol{\Sigma}}_e\|_F^2 + (-4g_1 g_3 + 4g_2^2)\operatorname{tr}\left(\overline{\boldsymbol{\Sigma}}_e\boldsymbol{\Sigma}_b\right)\right. \\ \quad \left. + \left(-2g_1 g_2^2 - \frac{2}{\tau}g_1 g_4 + \frac{2}{\tau}g_2 g_3 + 2g_1^2 g_3\right)\|\boldsymbol{\Sigma}_b\|_F^2\right). \end{cases}$$

CHAPTER 4

Conclusion and Discussion

This dissertation focuses on signal-to-noise ratio (SNR) estimation in high-dimensional linear models, covering both the univariate and multivariate cases. In Chapter 2, we study high-dimensional linear models with heteroscedastic and correlated noise (model (2.1)). We derive consistency and asymptotic distributions for the REML estimator of the SNR under a general fixed coefficient vector, without assuming i.i.d. Gaussian priors, and in the presence of heteroscedastic and correlated errors. These theoretical findings are supported by extensive numerical simulations.

In Chapter 3, we consider the multiple-response high-dimensional linear model (3.1) under both fixed and random effects settings. For the random effects model, we further extend our framework to accommodate residual heteroskedasticity. We propose definitions of SNR tailored to each model and establish the asymptotic distributions of the corresponding estimators. We also show how to make inference about SNR and demonstrate the practical validity of our methods through extensive simulations.

There remain several avenues for future work. For instance, the simulation results in Section 2.3 suggest that the symmetry (skew-free) assumption on the design matrix entries—crucial for the main results in Section 2.2—might be relaxed. This assumption enabled the use of double Rademacher sequences and a leave-k-out analysis to derive asymptotic properties by dealing with conditional mean and variance. However, given that skewness may occur in real high-dimensional data, it would be of practical and theoretical interest to extend our analysis to designs with asymmetric distributions.

Another promising direction is the estimation of group-wise SNR in models with grouped features, which is especially relevant in GWAS applications where genes can be grouped naturally (e.g., by chromosome; see [44]). While asymptotic analysis for mixed effects models with feature groups has been studied in the literature [21], extending our misspecification framework to this setting is nontrivial. As discussed in Section 2.5, we have made preliminary progress, but technical challenges remain.

Chapter 3 also opens the door to exploring likelihood-based methods for SNR estimation in multivariate random effects models. While maximum likelihood estimation (MLE) is statistically efficient, it becomes computationally prohibitive as the response dimension q increases, due to the $O(q^2)$ parameters involved in estimating $\mathbf{\Sigma}_b$ and $\mathbf{\Sigma}_e$. Since our primary interest lies in estimating the SNR, a full MLE approach may be unnecessarily expensive. Pseudo-likelihood or profile-likelihood methods offer a promising alternative by avoiding direct estimation of large covariance matrices.

Furthermore, in Chapter 3, we examine two specific random effects models with heteroskedasticity, under which the asymptotic variance of $\hat{r}^2$ can be consistently estimated. Several open questions arise in this context: Can we extend these results to more general heteroskedastic settings? For the two specific models considered, are there improved variance estimators beyond those we have proposed?

Lastly, while our theoretical results for the fixed effects model require the design matrix to have i.i.d. Gaussian entries, simulation results in Section 3.3.1 suggest that this assumption could potentially be relaxed. Investigating the theoretical implications of such robustness would also be a valuable direction for future research.

APPENDIX A

Supporting Proofs of Chapter 2

In the appendix we give detailed proofs of the technical lemmas that appear in Section 2.4. As mentioned earlier, the proofs of Lemmas 2.4.1.1, 2.4.1.2, 2.4.1.3, 2.4.1.5 and 2.4.1.10 basically follow the proof ideas in [24], but we provide self-contained proofs here for completeness. Interested readers are recommended to read [24] for deeper insights.

A.1. Preliminaries

Let's first recall the famous Marčenko-Pastur law in random matrix theory.

THEOREM A.1.1 (Marčenko-Pastur law, [39]). *Let $\boldsymbol{Z}$ be an $n \times p$ random matrix whose entries are i.i.d. random variables with mean 0 and variance 1 in which $n/p \to \tau \in (0, \infty)$ as $n, p \to \infty$. Then the empirical spectral distribution (ESD) of $S = p^{-1}\boldsymbol{Z}\boldsymbol{Z}^\top$, which is defined as $F^{\boldsymbol{S}}$, converges almost surely (a.s.) in distribution to F_τ, whose p.d.f. is given by*

$$f_\tau(x) = \begin{cases} \max\{\tau - 1, 0\}\delta_0(x) + \dfrac{1}{2\pi\tau x}\sqrt{(b_+(\tau) - x)\,(x - b_-(\tau))} & b_-(\tau) \le x \le b_+(\tau) \\ 0 & \textit{elsewhere} \end{cases}$$

where $b_\pm(\tau) = (1 \pm \sqrt{\tau})^2$ and $\delta_0(x)$ is a point mass τ^{-1} at the origin.

Note that in our settings, the entries of the design matrix are not necessarily identically distributed. To this end, we consider the following extension of Marčenko-Pastur law.

THEOREM A.1.2 ([2], Theorem 2.8). *Let $\boldsymbol{Z}$ be an $n \times p$ random matrix whose entries are independent random variables with mean 0 and variance 1. Assume that $n/p \to \tau \in (0, \infty)$ and that for any $\delta > 0$,*

$$\frac{1}{\delta^2 np}\sum_{i,j}\mathbb{E}\left[|z_{ij}^{(n)}|^2 I_{(|z_{ij}^{(n)}| \ge \delta\sqrt{n})}\right] \to 0.$$

Then $F^{\boldsymbol{S}}$, defined as in Theorem A.1.1, tends almost surely to the Marčenko-Pastur law with ratio index τ.

COROLLARY A.1.1. *Under the assumption of Theorem A.1.1 or A.1.2, for any integer l, we have*

$$\frac{1}{n}\operatorname{trace}(\boldsymbol{S}^l) \xrightarrow{a.s.} \int_{b_-(\tau)}^{b_+(\tau)} x^l f_\tau(x)\mathrm{d}x \quad \textit{as} \quad n,p\to\infty.$$

Define the sub-Gaussian norm of a random variable ζ as

$$\|\zeta\|_{\psi_2} \equiv \sup_{q\geq 1}\{q^{-1/2}(\mathbb{E}|\zeta|^q)^{1/q}\}.$$

A random variable ζ is sub-Gaussian if and only if its sub-Gaussian norm $\|\zeta\|_{\psi_2} < \infty$. We have the following equivalent characterizations on the sub-Gaussianity of a random variable:

LEMMA A.1.2.1 ([40], Lemma 5.5). *A random variable ζ is sub-Gaussian if and only if*

1) $\|\zeta\|_{\psi_2} < \infty$; or

2) $\mathbb{P}\{|\zeta| > t\} \leq \exp(1 - t^2/K^2)$ for some parameter $K > 0$ and all $t > 0$.

Part 2) actually implies that the design matrix under the setting of Theorem 3.2.4, in which the entries have sub-Gaussian norms that are uniformly upper bounded, satisfies the conditions in Theorem A.1.2. In fact, if ζ is sub-Gaussian random variable, then by the identity $\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > t)\mathrm{d}t$ for any nonnegative random variable X, we have

$$\begin{aligned}
\mathbb{E}\left[|\zeta|^2 I_{(|\zeta|\geq\delta\sqrt{n})}\right] &= \int_{\delta\sqrt{n}}^\infty \mathbb{P}\{|\zeta| > t\}2t\mathrm{d}t + \delta^2 n\mathbb{P}\{|\zeta| > \delta\sqrt{n}\} \\
&\leq 2\int_{\delta\sqrt{n}}^\infty e^{1-\frac{t^2}{K^2}} t\mathrm{d}t + \delta^2 n e^{1-\frac{\delta^2 n}{K^2}} \\
&= (K^2 + \delta^2 n)e^{1-\frac{\delta^2 n}{K^2}}.
\end{aligned}$$

This implies that for $n\times p$ random matrices $\boldsymbol{Z}$ whose entries have uniformly upper bounded sub-Gaussian norms,

$$\frac{1}{\delta^2 np}\sum_{i,j}\mathbb{E}\left[|z_{ij}^{(n)}|^2 I_{(|z_{ij}^{(n)}|\geq\delta\sqrt{n})}\right] \to 0,$$

as $n, p \to \infty$, for any $\delta > 0$.

Our proof also relies crucially on the following fundamental concentration inequalities.

Proposition A.1.1 (Hanson–Wright inequality, [33]). *Let* $\boldsymbol{\zeta} = (\zeta_1, \cdots, \zeta_n)^\top$*, where the* ζ_i*'s are independent random variables satisfying* $\mathbb{E}(\zeta_i) = 0$ *and* $\|\zeta_i\|_{\psi_2} \le K < \infty$*. Let* $\boldsymbol{A}$ *be an* $n \times n$ *deterministic matrix. Then we have for any* $t > 0$,

$$\mathbb{P}\{|\boldsymbol{\zeta}^\top \boldsymbol{A} \boldsymbol{\zeta} - \mathbb{E}(\boldsymbol{\zeta}^\top \boldsymbol{A} \boldsymbol{\zeta})| > t\} \le 2 \exp\left\{-c \min\left(\frac{t^2}{K^4 \|\boldsymbol{A}\|_F^2}, \frac{t}{K^2 \|\boldsymbol{A}\|}\right)\right\},$$

where $c > 0$ *is an absolute constant. Here* $\|\boldsymbol{A}\|$ *and* $\|\boldsymbol{A}\|_F$ *denote the operator and Frobenius norms of* $\boldsymbol{A}$*, respectively.*

Proposition A.1.2 (Hoeffding-type inequality, [40], Proposition 5.10). *Let* $\boldsymbol{\zeta} = (\zeta_1, \cdots, \zeta_n)^\top$*, where the* ζ_i*'s are independent centered sub-Gaussian random variables. Let* $K = \max_{1 \le i \le n} \|\zeta_i\|_{\psi_2}$ *and* $\boldsymbol{a} = (a_1, \cdots, a_N)^\top \in \mathbb{R}^N$*. Then we have for any* $t \ge 0$,

$$\mathbb{P}\{|\boldsymbol{a}^\top \boldsymbol{\zeta}| > t\} \le e \exp\left\{-c \frac{t^2}{K^2 \|\boldsymbol{a}\|_2^2}\right\},$$

where $c > 0$ *is an absolute constant.*

Proposition A.1.3 (Bernstein-type inequality, [40], Proposition 5.16). *Let* $\boldsymbol{\zeta} = (\zeta_1, \cdots, \zeta_n)^\top$*, where the* ζ_i*'s are independent centered sub-exponential random variables. Let* $K = \max_{1 \le i \le n} \|\zeta_i\|_{\psi_2}$ *and* $\boldsymbol{a} = (a_1, \cdots, a_N)^\top \in \mathbb{R}^N$*. Then we have for any* $t \ge 0$,

$$\mathbb{P}\{|\boldsymbol{a}^\top \boldsymbol{\zeta}| > t\} \le 2 \exp\left\{-c \min\left(\frac{t^2}{K^2 \|\boldsymbol{a}\|_2^2}, \frac{t}{K \|\boldsymbol{a}\|_\infty}\right)\right\},$$

where $c > 0$ *is an absolute constant.*

The next result, the famous Sherman-Morrison-Woodbury formula in matrix analysis is repeatedly used in our proofs, as the corner stone of leave-one-out analysis.

THEOREM A.1.3 (Sherman-Morrison-Woodbury formula, [17], Page 19). *Let* $\boldsymbol{P}$ *and* $\boldsymbol{Q}$ *be* n*-dimensional non-singular matrices such that* $\boldsymbol{Q} = \boldsymbol{P} + \boldsymbol{U}\boldsymbol{V}^\top$*, where* $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{R}^{n\times q}$*. Then*

$$\boldsymbol{Q}^{-1} = (\boldsymbol{P} + \boldsymbol{U}\boldsymbol{V}^\top)^{-1} = \boldsymbol{P}^{-1} - \boldsymbol{P}^{-1}\boldsymbol{U}(\boldsymbol{I}_q + \boldsymbol{V}^\top\boldsymbol{P}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}^\top\boldsymbol{P}^{-1}.$$

The following results, implied by [4] and [5], are conditions for the normality of quadratic forms.

THEOREM A.1.4 ([4], Proposition 3.1). *Let* $X = (X_1, \dots, X_n)$ *be i.i.d. Rademacher random variables and* $A = (a_{ij})1 \le i, j \le n$ *be a real symmetric matrix. Let* $W = X^\top \boldsymbol{A} X$ *and*

$$\sigma^2 = Var(W) = \frac{1}{2}\operatorname{trace}(\boldsymbol{A}^2).$$

Let μ *be the law of* $(W - \mathbb{E}(W))/\sqrt{Var(W)}$ *and let* ν *be the standard Gaussian law. We define*

$$d_W := \mathcal{W}(\mu, \nu),$$

where $\mathcal{W}$ *is the Kantorovich–Wasserstein distance between two probability measures with*

$$\mathcal{W}(\mu, \nu) = \sup\left\{\left|\int h d\mu - \int h d\nu\right| : h \text{ Lipschitz, with } \|h\|_{Lip} \le 1\right\}$$

Then,

$$d_W \le \left(\frac{\operatorname{trace}(\boldsymbol{A}^4)}{2\sigma^4}\right)^{1/2} + \frac{5}{2\sigma^3}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij}^2\right)^{3/2} \le 6\sqrt{2}\frac{\|\boldsymbol{A}\|^2}{\|\boldsymbol{A}\|_F^2}.$$

THEOREM A.1.5 ([5]). *Suppose* $\boldsymbol{x}$ *is a gaussian random vector with mean 0 and covariance matrix* $\boldsymbol{\Sigma}$*. Take any* $g \in C^2(\mathbb{R})$ *and let* ∇g *and* $\nabla^2 g$ *denote the gradient and Hessian of* g*. Let*

$$\varsigma_1 = \left(\mathbb{E}\,\|\nabla g(\boldsymbol{x})\|^4\right)^{\frac{1}{4}}, \quad \varsigma_2 = \left(\mathbb{E}\,\|\nabla^2 g(\boldsymbol{x})\|^4\right)^{\frac{1}{4}}.$$

Then let $W = g(\boldsymbol{x})$ have a finite fourth moment and U be a normal random variable having the same mean and variance as W,

$$d_{TV}(W, U) \leq \frac{2\sqrt{5}\|\boldsymbol{\Sigma}\|^{\frac{3}{2}}\varsigma_1\varsigma_2}{Var[W]}.$$

Here d_{TV} is the total variation distance between random variables u and v,

$$d_{TV}(u, v) = \sup_{B \in \mathcal{B}(\mathbb{R})} |\mathbb{P}(u \in B) - \mathbb{P}(v \in B)|,$$

where $\mathcal{B}(\mathbb{R})$ denotes the collection of Borel sets in $\mathbb{R}$.

Next, there is a famous result for the bounds of eigenvalues of the sub-gaussian random matrix.

THEOREM A.1.6 (Theorem 5.39, [40]). *Let $\boldsymbol{Z}$ be an $n \times p$ matrix whose rows are independent sub-gaussian isotropic random vectors. Then for every $t \geq 0$, with probability at least $1 - 2\exp(-ct^2)$ one has*

$$\sqrt{n} - C\sqrt{p} - t \leq \lambda_{\min}(\boldsymbol{Z}) \leq \lambda_{\max}(\boldsymbol{Z}) \leq \sqrt{n} + C\sqrt{p} + t$$

Here $C = C_K$, $c = c_K > 0$ depend only on the subgaussian norm K of the rows.

A.2. Proofs of Lemmas in Section 2.4.1

A.2.1. Proof of Lemma 2.4.0.1. Since $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$, there holds $(\varepsilon_i/\sigma_i)^2 \sim \chi_1^2$. By the standard Laurent-Massart bound ([25]), there holds that

$$\mathbb{P}\{\varepsilon_i^2/\sigma_i^2 - 1 \geq 2\sqrt{t} + 2t\} \leq \exp(-t) \text{ and } \mathbb{P}\{1 - \varepsilon_i^2/\sigma_i^2 \geq 2\sqrt{t}\} \leq \exp(-t).$$

Taking $t = 2\log n$, we can have for any $i = 1, \ldots, n$,

$$\mathbb{P}\{\max\left|\varepsilon_i^2/\sigma_i^2 - 1\right| \geq 2\sqrt{2\log n} + 4\log n\} \leq \frac{1}{n}, \tag{A.1}$$

which implies (2.12).

Since $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon)$, we can rewrite $\sum_{i=1}^n \varepsilon_i^2$ as $\boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon}$, then

$$\mathbb{E}\left[\boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon}\right] = \text{trace}\left(\boldsymbol{\Sigma}_\varepsilon\right) = \sum_{i=1}^n \sigma_i^2 = n\sigma_0^4,$$

and

$$\text{Var}\left[\boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon}\right] = \text{trace}\left(\boldsymbol{\Sigma}_\varepsilon^2\right) = 2\|\boldsymbol{\Sigma}_\varepsilon\|_F^2.$$

By the assumption that $\|\boldsymbol{\Sigma}_\varepsilon\|_F = o(n)$,

$$\text{Var}\left[\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2\right] = o\left(1\right).$$

Then we can have (2.13).

By applying Theorem A.1.5 directly, we can take

$$g(\boldsymbol{x}) = \frac{1}{\sqrt{n}}\sum_{i=1}^n x_i^2,$$

then

$$d_{TV}\left(g(\boldsymbol{\varepsilon}), U\right) \leq \frac{\sqrt{5}\|\boldsymbol{\Sigma}_\varepsilon\|^{\frac{3}{2}}\varsigma_1\varsigma_2}{\frac{1}{n}\|\boldsymbol{\Sigma}_\varepsilon\|_F^2},$$

where $U \sim \mathcal{N}(\sqrt{n}\sigma_0^2, \frac{1}{n}\|\boldsymbol{\Sigma}_\varepsilon\|_F^2)$. Since

$$\frac{\partial g}{\partial x_i} = \frac{2}{\sqrt{n}}x_i,$$

$$\frac{\partial g}{\partial x_i \partial x_j} = \begin{cases} \frac{2}{\sqrt{n}} & i = j, \\ 0 & i \neq j, \end{cases}$$

by the assumption that $\|\boldsymbol{\Sigma}_\varepsilon\|_F^2 = n\kappa\sigma_0^4$, it follows that

$$\begin{aligned}
\varsigma_1 &= \left(\mathbb{E}\,\|\nabla g(\boldsymbol{\varepsilon})\|^4\right)^{\frac{1}{4}} = \left(\mathbb{E}\left(\sum_{i=1}^{n}\left(\frac{2}{\sqrt{n}}\varepsilon_i\right)^2\right)^2\right)^{\frac{1}{4}} = \frac{2}{\sqrt{n}}\left(\mathbb{E}\left(\sum_{i=1}^{n}\varepsilon_i^2\right)^2\right)^{\frac{1}{4}} \\
&= \frac{2}{\sqrt{n}}\left(\mathrm{Var}\left[\boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon}\right] + \left(\mathbb{E}\left[\boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon}\right]\right)^2\right)^{\frac{1}{4}} \\
&= O(1),
\end{aligned}$$

and $\varsigma_2 = (\mathbb{E}\,\|\nabla^2 g(\boldsymbol{\varepsilon})\|^4)^{\frac{1}{4}} = O(1/\sqrt{n})$.Then by the assumption that $\|\boldsymbol{\Sigma}_\varepsilon\|$ is uniformly bounded, we can have that as $n \to \infty$

$$d_{TV}\left(g(\boldsymbol{\varepsilon}), U\right) = O(1/\sqrt{n}) = o(1),$$

which implies (2.14).

A.2.2. Proof of Lemma 2.4.1.1. For convenience, define

$$\text{(A.2)}\qquad \begin{cases} \rho_k := \eta_{kk,k}^{(1)} := \boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_k, \\ \phi_k := \eta_{kk,k}^{(2)} := \boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-2}\boldsymbol{z}_k, \\ \psi_k := \eta_{kk,k}^{(3)} := \boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-3}\boldsymbol{z}_k. \end{cases}$$

First, there is a simple relationship: $\psi_k \le \phi_k \le \rho_k$. In fact, since $\boldsymbol{I}_n - \boldsymbol{V}_{\gamma,-k}^{-1} \succeq \mathbf{0}$, we know that

$$\rho_k - \phi_k = \boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-1/2}(I - \boldsymbol{V}_{\gamma,-k}^{-1})\boldsymbol{V}_{\gamma,-k}^{-1/2}\boldsymbol{z}_k \ge 0.$$

i.e., $\phi_k \le \rho_k$. We can similarly obtain $\psi_k \le \phi_k$.

Using Sherman-Morrison-Woodbury formula (Theorem A.1.3), we have

$$\text{(A.3)}\qquad \boldsymbol{V}_\gamma^{-1} = \boldsymbol{V}_{\gamma,-k}^{-1} - \frac{\gamma}{p}(1 + \frac{\gamma}{p}\rho_k)^{-1}\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_k\boldsymbol{z}_k^\top\boldsymbol{V}_{\gamma,-k}^{-1},$$

and

$$
\begin{aligned}
\boldsymbol{V}_\gamma^{-2} &= \left(\boldsymbol{V}_{\gamma,-k}^{-1} - \frac{\gamma}{p}(1+\frac{\gamma}{p}\rho_k)^{-1}\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_k\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-1}\right)^2 \\
&= \boldsymbol{V}_{\gamma,-k}^{-2} - \frac{\gamma}{p}(1+\frac{\gamma}{p}\rho_k)^{-1}\boldsymbol{V}_{\gamma,-k}^{-2}\boldsymbol{z}_k\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-1} - \frac{\gamma}{p}(1+\frac{\gamma}{p}\rho_k)^{-1}\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_k\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-2} \\
&\quad + \left(\frac{\gamma}{p}\right)^2(1+\frac{\gamma}{p}\rho_k)^{-2}\phi_k\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_k\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-1}. \qquad \text{(A.4)}
\end{aligned}
$$

By (A.3) and (A.4), we can also have

$$
\operatorname{trace}(\boldsymbol{V}_\gamma^{-1}) = \operatorname{trace}(\boldsymbol{V}_{\gamma,-k}^{-1}) - \frac{\gamma}{p}(1+\frac{\gamma}{p}\rho_k)^{-1}\phi_k,
$$

and

$$
\operatorname{trace}(\boldsymbol{V}_\gamma^{-2}) = \operatorname{trace}(\boldsymbol{V}_{\gamma,-k}^{-2}) - \frac{2\gamma}{p}(1+\frac{\gamma}{p}\rho_k)^{-1}\psi_k + (\frac{\gamma}{p})^2(1+\frac{\gamma}{p}\rho_k)^{-2}\phi_k^2.
$$

Then

$$
\left|\operatorname{trace}(\boldsymbol{V}_\gamma^{-1}) - \operatorname{trace}(\boldsymbol{V}_{\gamma,-k}^{-1})\right| = \frac{\gamma}{p}(1+\frac{\gamma}{p}\rho_k)^{-1}\phi_k \le \frac{\gamma}{p}(1+\frac{\gamma}{p}\rho_k)^{-1}\rho_k < 1, \qquad \text{(A.5)}
$$

and

$$
\left|\operatorname{trace}(\boldsymbol{V}_\gamma^{-2}) - \operatorname{trace}(\boldsymbol{V}_{\gamma,-k}^{-2})\right| \le \frac{2\gamma}{p}(1+\frac{\gamma}{p}\rho_k)^{-1}\rho_k + (\frac{\gamma}{p})^2(1+\frac{\gamma}{p}\rho_k)^{-2}\rho_k^2 < 3. \qquad \text{(A.6)}
$$

Similarly, we can also prove that

$$
\left|\operatorname{trace}(\boldsymbol{V}_\gamma^{-3}) - \operatorname{trace}(\boldsymbol{V}_{\gamma,-k}^{-3})\right| \le 7 \text{ and } \left|\operatorname{trace}(\boldsymbol{V}_\gamma^{-4}) - \operatorname{trace}(\boldsymbol{V}_{\gamma,-k}^{-4})\right| \le 15.
$$

Since the entries of $\boldsymbol{Z}$ are independent sub-Gaussian and $\mathbb{E}(z_{ik}) = 0$, using Proposition A.1.1, we have, for any $1 \le k \le p$ and $t > 0$:

$$
\mathbb{P}\{|\rho_k - \operatorname{trace}(\boldsymbol{V}_{\gamma,-k}^{-1})| > t|\boldsymbol{V}_{\gamma,-k}\} \le 2\exp\left\{-c\min\left(\frac{t^2}{K^4\|\boldsymbol{V}_{\gamma,-k}^{-1}\|_F^2}, \frac{t}{K^2\|\boldsymbol{V}_{\gamma,-k}^{-1}\|}\right)\right\},
$$

where c and K are positive constants. If we set

$$t = t_k = K^2 \max\left(\sqrt{\frac{2\log p}{c}}\|\boldsymbol{V}_{\gamma,-k}^{-1}\|_F, \frac{2\log p}{c}\|\boldsymbol{V}_{\gamma,-k}^{-1}\|\right),$$

it follows that $\mathbb{P}\{|\rho_k - \text{trace}(\boldsymbol{V}_{\gamma,-k}^{-1})| > t_k|\boldsymbol{V}_{\gamma,-k}^{-1}\} \leq 2/p^2$. Thus

$$\mathbb{P}\left\{\max_{1\leq k\leq p} t_k^{-1}|\rho_k - \text{trace}(\boldsymbol{V}_{\gamma,-k}^{-1})| > 1\right\} \leq \frac{2}{p}.$$

By Lemma 2.4.0.2, $\|\boldsymbol{V}_{\gamma,-k}^{-1}\| \leq 1$, and $\|\boldsymbol{V}_{\gamma,-k}^{-1}\|_F \leq \sqrt{n}\|\boldsymbol{V}_{\gamma,-k}^{-1}\| \leq \sqrt{n}$, we can obtain that

$$t_k \leq K^2 \max\left(\sqrt{\frac{2}{c}}\sqrt{n\log p}, \frac{2}{c}\log p\right),$$

which implies

$$\mathbb{P}\left\{\max_{1\leq k\leq p}|\rho_k - \text{trace}(\boldsymbol{V}_{\gamma,-k}^{-1})| > C\sqrt{n\log p}\right\} \leq 2/p$$

for some constant $C > 0$. Then, it follows that

$$\max_{1\leq k\leq p}|\rho_k - \text{trace}(\boldsymbol{V}_{\gamma,-k}^{-1})| = O_P(\sqrt{n\log n}). \tag{A.7}$$

By a similar argument, we have

$$\max_{1\leq k\leq p}|\phi_k - \text{trace}(\boldsymbol{V}_{\gamma,-k}^{-2})| = O_P(\sqrt{n\log n}). \tag{A.8}$$

Combining (A.5), (A.7), (A.6) and (A.8), we have

$$\max_{1\leq k\leq p}|\rho_k - \text{trace}(\boldsymbol{V}_{\gamma}^{-1})| = O_P(\sqrt{n\log n}), \quad \text{and} \tag{A.9}$$

$$\max_{1\leq k\leq p}|\phi_k - \text{trace}(\boldsymbol{V}_{\gamma}^{-2})| = O_P(\sqrt{n\log n}). \tag{A.10}$$

A.2.3. Proof of Lemma 2.4.1.2. Based on (A.3) and (A.4), there holds

$$\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-1}\boldsymbol{z}_k = (1 + \frac{\gamma}{p}\rho_k)^{-1}\rho_k, \tag{A.11}$$

$$\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-2}\boldsymbol{z}_k = (1 + \frac{\gamma}{p}\rho_k)^{-2}\phi_k, \tag{A.12}$$

where ρ_k and ϕ_k are defined in (A.2).

Let's now come back to find approximations of $\mathbb{E}[A_1|\boldsymbol{Z}]$ and $\mathbb{E}[A_2|\boldsymbol{Z}]$. We define the following intermediate quantities

$$\theta_1 = \frac{1}{n}\frac{\text{trace}(\boldsymbol{V}_\gamma^{-1})}{1+\frac{\gamma}{p}\,\text{trace}(V_\gamma^{-1})} \text{ and } \theta_2 = \frac{1}{n}\frac{\text{trace}(\boldsymbol{V}_\gamma^{-2})}{\left(1+\frac{\gamma}{p}\,\text{trace}(V_\gamma^{-1})\right)^2}. \tag{A.13}$$

Then by (A.11) and (A.9), we can have

$$\max_{1\le k\le p}\left|\theta_1 - \frac{\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-1}\boldsymbol{z}_k}{n}\right| \le \max_{1\le k\le p}\left|\frac{\text{trace}(\boldsymbol{V}_\gamma^{-1})-\rho_k}{n}\right| = O_P\left(\sqrt{\frac{\log n}{n}}\right), \tag{A.14}$$

which implies (2.18). Similarly, by (A.12) and (A.10), there holds

$$\begin{aligned}
\max_{1\le k\le p}\left|\theta_2 - \frac{\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-2}\boldsymbol{z}_k}{n}\right| &\le \max_{1\le k\le p}\frac{1}{n}\left|\text{trace}(\boldsymbol{V}_\gamma^{-2})\left(1+\frac{\gamma}{p}\rho_k\right)^2 - \left(1+\frac{\gamma}{p}\,\text{trace}(V_\gamma^{-1})\right)^2\phi_k\right| \\
&\le \frac{1}{n}\Bigg[\max_{1\le k\le p}\left|\text{trace}(\boldsymbol{V}_\gamma^{-2})-\phi_k\right| + \max_{1\le k\le p}\frac{2\gamma}{p}\rho_k\left|\text{trace}(\boldsymbol{V}_\gamma^{-2})-\phi_k\right| \\
&\quad + \max_{1\le k\le p}\frac{2\gamma}{p}\phi_k\left|\rho_k - \text{trace}(\boldsymbol{V}_\gamma^{-1})\right| + \max_{1\le k\le p}\frac{\gamma^2}{p^2}\rho_k^2\left|\text{trace}(\boldsymbol{V}_\gamma^{-2})-\phi_k\right| \\
&\quad + \max_{1\le k\le p}\frac{\gamma^2}{p^2}\phi_k\left(\left|\rho_k-\text{trace}(\boldsymbol{V}_\gamma^{-1})\right|\left|\rho_k+\text{trace}(\boldsymbol{V}_\gamma^{-1})\right|\right)\Bigg].
\end{aligned} \tag{A.15}$$

It follows, by the facts $\text{trace}(\boldsymbol{V}_\gamma^{-1}) = O_P(n)$, $\text{trace}(\boldsymbol{V}_\gamma^{-2}) = O_P(n)$, $\rho_k = O_P(n)$ and $\phi_k = O_P(n)$ (Lemma 2.4.0.2 and Lemma 2.4.1.1),(2.19) is true.

Then we can have for $l = 1, 2$,

$$\begin{aligned}
&\left|\frac{\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-l}\boldsymbol{z}_k}{n} - \frac{1}{np}\,\text{trace}\left(\boldsymbol{V}_\gamma^{-1}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right| \\
&= \left|\frac{\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-l}\boldsymbol{z}_k}{n} - \theta_1 + \frac{1}{p}\sum_{i=1}^p\left(\theta_1 - \sum_{i=1}^p\frac{1}{n}\boldsymbol{z}_i^\top \boldsymbol{V}_\gamma^{-1}\boldsymbol{z}_i\right)\right| = O_P\left(\sqrt{\frac{\log n}{n}}\right),
\end{aligned} \tag{A.16}$$

which implies (2.20). Then by the definition of $\boldsymbol{B}_\gamma$, when $l = 1$ we can get (2.21) from (2.20).

When $l=2$, for $\left(\boldsymbol{z}_k^\top \boldsymbol{B}_{\gamma_0} \boldsymbol{z}_k\right)^2$,

$$\left(\boldsymbol{z}_k^\top \boldsymbol{B}_{\gamma} \boldsymbol{z}_k\right)^2 = \left(\frac{1}{n}\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-1}\boldsymbol{z}_k - \frac{\frac{1}{n}\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-2}\boldsymbol{z}_k}{\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})}\right)^2$$

$$= \left(\frac{1}{n}\eta_{kk}^{(1)}\right)^2 - 2\frac{\frac{1}{n}\eta_{kk}^{(1)}\frac{1}{n}\eta_{kk}^{(2)}}{\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})} + \frac{\left(\frac{1}{n}\eta_{kk}^{(2)}\right)^2}{\left(\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})\right)^2}. \tag{A.17}$$

Then by triangle inequality, for any $l, m = 1, 2$,

$$\begin{aligned}
&\frac{1}{n}\eta_{kk}^{(l)}\frac{1}{n}\eta_{kk}^{(m)} - \frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-l}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-m}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right) \\
&\le \frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-l}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\left|\frac{1}{n}\eta_{kk}^{(m)} - \frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-m}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right| \\
&\quad + \frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-m}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\left|\frac{1}{n}\eta_{kk}^{(l)} - \frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-l}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right| \\
&\quad + \left|\frac{1}{n}\eta_{kk}^{(m)} - \frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-m}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right|\left|\frac{1}{n}\eta_{kk}^{(l)} - \frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-l}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right|.
\end{aligned}$$

By (A.16) and the fact that

$$\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-l}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right) = O_P(1), \quad \frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-m}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right) = O_P(1),$$

we can have

$$\max_{1\le k\le p}\left|\frac{1}{n}\eta_{kk}^{(l)}\frac{1}{n}\eta_{kk}^{(m)} - \frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-l}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-m}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right| = O_P\left(\sqrt{\frac{\log n}{n}}\right). \tag{A.18}$$

Since

$$
\begin{aligned}
&\left(\operatorname{trace}\left(\boldsymbol{B}_\gamma \frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right)^2 \\
&= \left(\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right) - \frac{\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)}{\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})}\right)^2 \\
&= \left(\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right)^2 - 2\frac{\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)}{\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})} \\
&+ \frac{\left(\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right)^2}{\left(\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})\right)^2},
\end{aligned}
$$

then by (A.17) and (A.18), there holds that

$$
\max_{1\le k\le p}\left|\left(\boldsymbol{z}_k^\top \boldsymbol{B}_\gamma \boldsymbol{z}_k\right)^2 - \left(\operatorname{trace}\left(\boldsymbol{B}_\gamma\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right)^2\right| = O_P\left(\sqrt{\frac{\log n}{n}}\right). \tag{A.19}
$$

Finally, as we can know from (2.18), (2.20) and Lemma 2.4.0.2 that $(np)^{-1}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\boldsymbol{Z}\boldsymbol{Z}^\top\right)$ converges to the same limit as

$$
\frac{n^{-1}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})}{1+\gamma p^{-1}\operatorname{trace}(V_\gamma^{-1})},
$$

which means

$$
\frac{1}{\gamma}(1-h_1(\gamma,\tau)) = \frac{h_1(\gamma,\tau)}{1+\gamma\tau h_1(\gamma,\tau)}.
$$

And Lemma 2.4.0.2 shows that

$$
\left|\frac{1}{np}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\boldsymbol{Z}\boldsymbol{Z}^\top\right) - \frac{1}{\gamma}(1-h_1(\gamma,\tau))\right| = O_P\left(\frac{1}{n}\right),
$$

and

$$
\left|\frac{1}{n}\frac{\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})}{1+\frac{\gamma}{p}\operatorname{trace}(V_\gamma^{-1})} - \frac{h_1(\gamma,\tau)}{1+\gamma\tau h_1(\gamma,\tau)}\right| = O_P\left(\frac{1}{n}\right).
$$

Combine the above two inequalities, we can get (2.22). Similarly, by (2.19), (2.20) and Lemma 2.4.0.2 we can get (2.23).

A.2.4. Proof of Lemma 2.4.1.3. By (2.22) and (A.11), we can know that

$$
\begin{aligned}
&\max_{1\le k\le p}\Bigg|\frac{1}{n}\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-1}\boldsymbol{z}_k-\frac{1}{np}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\\
&-\frac{1}{\left(1+\frac{\gamma}{p}\operatorname{trace}(V_\gamma^{-1})\right)^2}\left(\frac{1}{n}\eta^{(1)}_{kk,k}-\frac{1}{n}\operatorname{trace}(V_\gamma^{-1})\right)\Bigg|\\
&=\max_{1\le k\le p}\Bigg|\frac{1}{n}\frac{\eta^{(1)}_{kk,k}}{1+\frac{\gamma}{p}\eta^{(1)}_{kk,k}}-\frac{1}{n}\frac{\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})}{1+\frac{\gamma}{p}\operatorname{trace}(V_\gamma^{-1})}\\
&-\frac{1}{\left(1+\frac{\gamma}{p}\operatorname{trace}(V_\gamma^{-1})\right)^2}\left(\frac{1}{n}\eta^{(1)}_{kk,k}-\frac{1}{n}\operatorname{trace}(V_\gamma^{-1})\right)\Bigg|+O_P\left(\frac{1}{n}\right),
\end{aligned}
\tag{A.20}
$$

and similarly by (2.23) and (A.12), there holds that

$$
\begin{aligned}
&\max_{1\le k\le p}\Bigg|\frac{1}{n}\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-2}\boldsymbol{z}_k-\frac{1}{np}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\\
&+\frac{\operatorname{trace}(\boldsymbol{V}_\gamma^{-2})}{\left(1+\frac{\gamma}{p}\operatorname{trace}(V_\gamma^{-1})\right)^3}\frac{2\gamma}{p}\left(\frac{1}{n}\eta^{(1)}_{kk,k}-\frac{1}{n}\operatorname{trace}(V_\gamma^{-1})\right)\\
&-\frac{1}{\left(1+\frac{\gamma}{p}\operatorname{trace}(V_\gamma^{-1})\right)^2}\left(\frac{1}{n}\eta^{(2)}_{kk,k}-\frac{1}{n}\operatorname{trace}(V_\gamma^{-2})\right)\Bigg|\\
&=\max_{1\le k\le p}\Bigg|\frac{\eta^{(2)}_{kk,k}}{\left(1+\frac{\gamma}{p}\eta^{(1)}_{kk,k}\right)^2}-\frac{\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\right)}{\left(1+\frac{\gamma}{p}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right)^2}\\
&+\frac{\operatorname{trace}(\boldsymbol{V}_\gamma^{-2})}{\left(1+\frac{\gamma}{p}\operatorname{trace}(V_\gamma^{-1})\right)^3}\frac{2\gamma}{p}\left(\frac{1}{n}\eta^{(1)}_{kk,k}-\frac{1}{n}\operatorname{trace}(V_\gamma^{-1})\right)\\
&-\frac{1}{\left(1+\frac{\gamma}{p}\operatorname{trace}(V_\gamma^{-1})\right)^2}\left(\frac{1}{n}\eta^{(2)}_{kk,k}-\frac{1}{n}\operatorname{trace}(V_\gamma^{-2})\right)\Bigg|+O_P\left(\frac{1}{n}\right).
\end{aligned}
$$

Define

$$z_\gamma(x) = \frac{x}{1 + \gamma \frac{n}{p} x}, \quad w_\gamma(x, y) = \frac{x}{\left(1 + \gamma \frac{n}{p} y\right)^2}.$$

By the Taylor series expansion, as $\frac{1}{n}\eta^{(1)}_{kk,k} \rightarrow \frac{1}{n} \operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)$

$$\begin{aligned} z_\gamma\left(\frac{1}{n}\eta^{(1)}_{kk,k}\right) &= z_\gamma\left(\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right) + z'_\gamma\left(\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right)\left(\frac{1}{n}\eta^{(1)}_{kk,k} - \frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right) \\ &\quad + R_1\left(\frac{1}{n}\eta^{(1)}_{kk,k}, \frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right). \end{aligned}$$

Here R_1 is the remainder term

$$R_1 = \frac{1}{2} z''_\gamma\left(c_k\right)\left(\frac{1}{n}\eta^{(1)}_{kk,k} - \frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right)^2$$

where c_k is some constant between $\frac{1}{n}\eta^{(1)}_{kk,k}$ and $\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)$. Then

$$\begin{aligned} &\frac{\frac{1}{n}\eta^{(1)}_{kk,k}}{1 + \frac{\gamma}{p}\eta^{(1)}_{kk,k}} - \frac{\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)}{1 + \frac{\gamma}{p}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)} \\ &= z_\gamma\left(\frac{1}{n}\eta^{(1)}_{kk,k}\right) - z_\gamma\left(\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right) \\ &= z'_\gamma\left(\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right)\left(\frac{1}{n}\eta^{(1)}_{kk,k} - \frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right) + R_1\left(\frac{1}{n}\eta^{(1)}_{kk,k}, \frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right), \end{aligned}$$

where

$$z'_\gamma\left(x\right) = \frac{1}{\left(1 + \frac{\gamma n}{p} x\right)^2}, \quad z''_\gamma\left(x\right) = \frac{\gamma n}{p}\frac{1}{\left(1 + \frac{\gamma n}{p} x\right)^3}.$$

By (A.20), this implies that

$$
\begin{aligned}
&\max_{1\le k\le p}\Bigg|\frac{1}{n}\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-1}\boldsymbol{z}_k-\frac{1}{np}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\\
&\quad-\frac{1}{\left(1+\frac{\gamma}{p}\operatorname{trace}(V_\gamma^{-1})\right)^2}\left(\frac{1}{n}\eta_{kk,k}^{(1)}-\frac{1}{n}\operatorname{trace}(V_\gamma^{-1})\right)\Bigg|\\
&=\max_{1\le k\le p}\left|R_1\left(\frac{1}{n}\eta_{kk,k}^{(1)},\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right)\right|+O_P\left(\frac{1}{n}\right),
\end{aligned}
$$

and by Lemma 2.4.1.1,

$$
\begin{aligned}
&\max_{1\le k\le p}\left|R_1\left(\frac{1}{n}\eta_{kk,k}^{(1)},\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right)\right|\\
&=\max_{1\le k\le p}\left|\frac{1}{2}z_\gamma^{''}(c_k)\left(\frac{1}{n}\eta_{kk,k}^{(1)}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right)^2\right|\\
&\le\frac{1}{2}\max_{1\le k\le p}\left|\frac{\gamma n}{p}\frac{1}{\left(1+\frac{\gamma n}{p}c_k\right)^3}\right|\max_{1\le k\le p}\left|\left(\frac{1}{n}\eta_{kk,k}^{(1)}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right)^2\right|\\
&\le\frac{\gamma n}{2p}\max_{1\le k\le p}\left|\left(\frac{1}{n}\eta_{kk,k}^{(1)}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right)^2\right|\\
&=O_P\left(\frac{\log n}{n}\right).
\end{aligned}
$$

Combine the above two inequalities, we can get (2.24).

Similarly, by the Taylor series expansion, as $\frac{1}{n}\eta^{(2)}_{kk,k} \to \frac{1}{n}\,\text{trace}\left(\boldsymbol{V}_\gamma^{-2}\right)$, we can have

$$\begin{aligned}
&\max_{1\le k\le p}\left|\frac{1}{n}\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-2}\boldsymbol{z}_k - \frac{1}{np}\,\text{trace}\left(\boldsymbol{V}_\gamma^{-2}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right.\\
&+\frac{\text{trace}(\boldsymbol{V}_\gamma^{-2})}{\left(1+\frac{\gamma}{p}\,\text{trace}(V_\gamma^{-1})\right)^3}\frac{2\gamma}{p}\left(\frac{1}{n}\eta^{(1)}_{kk,k}-\frac{1}{n}\,\text{trace}(V_\gamma^{-1})\right)\\
&\left.-\frac{1}{\left(1+\frac{\gamma}{p}\,\text{trace}(V_\gamma^{-1})\right)^2}\left(\frac{1}{n}\eta^{(2)}_{kk,k}-\frac{1}{n}\,\text{trace}(V_\gamma^{-2})\right)\right|\\
&=\max_{1\le k\le p}\left|\tilde{R}_1\left(\frac{1}{n}\eta^{(2)}_{kk,k},\frac{1}{n}\,\text{trace}\left(\boldsymbol{V}_\gamma^{-2}\right),\frac{1}{n}\eta^{(1)}_{kk,k},\frac{1}{n}\,\text{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right)\right|+O_P\left(\frac{1}{n}\right).
\end{aligned}$$

Since

$$\begin{aligned}
&\frac{\partial^2 w_\gamma}{\partial x^2}(x,y)=0,\\
&\left|\frac{\partial^2 w_\gamma}{\partial x\partial y}(x,y)\right| = \left|-2\frac{\gamma n}{p}\frac{1}{\left(1+\gamma\frac{n}{p}y\right)^3}\right| \le 2\frac{\gamma n}{p}\quad \text{for } y\ge 0,\\
&\left|\frac{\partial^2 w_\gamma}{\partial y^2}(x,y)\right| = 6\left(\frac{\gamma n}{p}\right)^2\frac{x}{\left(1+\gamma\frac{n}{p}y\right)^3}\le 6\left(\frac{\gamma n}{p}\right)^2 x\quad \text{for } x,y\ge 0,
\end{aligned}$$

by Lemma 2.4.1.1 we can have

$$\begin{aligned}
&\max_{1\le k\le q}\left|\tilde{R}_1\left(\frac{1}{n}\eta^{(2)}_{kk,k},\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\right),\frac{1}{n}\eta^{(1)}_{kk,k},\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right)\right|\\
&\le \max_{1\le k\le q}\left|\frac{\partial^2 w_\gamma}{\partial x\partial y}(c_{k1},c_{k2})\left(\frac{1}{n}\eta^{(2)}_{kk,k}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\right)\right)\left(\frac{1}{n}\eta^{(1)}_{kk,k}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right)\right|\\
&\ +\max_{1\le k\le q}\left|\frac{\frac{\partial^2 w_\gamma}{\partial y^2}(c_{k1},c_{k2})}{2}\left(\frac{1}{n}\eta^{(1)}_{kk,k}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right)^2\right|\\
&\le 2\frac{\gamma n}{p}\max_{1\le k\le q}\left|\left(\frac{1}{n}\eta^{(2)}_{kk,k}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\right)\right)\right|\max_{1\le k\le q}\left|\left(\frac{1}{n}\eta^{(1)}_{kk,k}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right)\right|\\
&\ +3\left(\frac{\gamma n}{p}\right)^2 c_{k1}\max_{1\le k\le q}\left|\left(\frac{1}{n}\eta^{(1)}_{kk,k}-\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right)^2\right|\\
&= O_P\left(\frac{\log n}{n}\right),
\end{aligned}$$

where c_{k1} is some constant between $\frac{1}{n}\eta^{(2)}_{kk,k}$ and $\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\right)$ and c_{k2} is some constant between $\frac{1}{n}\eta^{(1)}_{kk,k}$ and $\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)$.

A.2.5. Proof of Lemma 2.4.1.5. Note that

$$(\boldsymbol{z}_k^\top\boldsymbol{V}_\gamma^{-1}\boldsymbol{z}_j)^2=(1+\frac{\gamma}{p}\rho_k)^{-2}(\boldsymbol{z}_k^\top\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_j)^2\le\left(\boldsymbol{z}_k^\top\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_j\right)^2$$

and

$$\begin{aligned}
&(\boldsymbol{z}_k^\top\boldsymbol{V}_\gamma^{-2}\boldsymbol{z}_j)^2 \qquad\text{(A.21)}\\
&=\left((1+\frac{\gamma}{p}\rho_k)^{-1}\boldsymbol{z}_k^\top\boldsymbol{V}_{\gamma,-k}^{-2}\boldsymbol{z}_j+\left(-\frac{\gamma}{p}\phi_k(1+\frac{\gamma}{p}\rho_k)^{-2}\boldsymbol{z}_k^\top\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_j\right)\right)^2\\
&\le\frac{1}{2}(1+\frac{\gamma}{p}\rho_k)^{-2}\left(\boldsymbol{z}_k^\top\boldsymbol{V}_{\gamma,-k}^{-2}\boldsymbol{z}_j\right)^2+\frac{1}{2}\left(\frac{\gamma}{p}\phi_k\right)^2(1+\frac{\gamma}{p}\rho_k)^{-4}\left(\boldsymbol{z}_k^\top\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_j\right)^2\\
&\le 2\left(\left(\boldsymbol{z}_k^\top\boldsymbol{V}_{\gamma,-k}^{-2}\boldsymbol{z}_j\right)^2+\left(\boldsymbol{z}_k^\top\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_j\right)^2\right) \qquad\text{(A.22)}
\end{aligned}$$

where the last inequality is due to $0<\phi_k\le\rho_k$.

Denote $\boldsymbol{Z}_{-k}=[\boldsymbol{z}_1,\dots,\boldsymbol{z}_{k-1},\boldsymbol{z}_{k+1},\dots,\boldsymbol{z}_p]$. Note that the components of $\boldsymbol{z}_k$ are independent mean-zero sub-Gaussian random variables, conditional on $\boldsymbol{Z}_{-k}$, by Proposition A.1.2, we have, for any $k \neq j$ and $t \geq 0$,

$$\mathbb{P}\left\{|\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_j| \geq t \Big| \boldsymbol{Z}_{-k}\right\} \leq e\exp\left\{-c\frac{t^2}{K^2\|\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_j\|_2^2}\right\},$$

where c and K are some positive constants. By letting $t = K\sqrt{\frac{3\log p}{c}}\|\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_j\|_2$, it follows

$$\mathbb{P}\left\{|\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_j| \geq C\sqrt{\log p}\|\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_j\| \Big| \boldsymbol{Z}_{-k}\right\} \leq \frac{e}{p^3},$$

where C is some positive constant. It further implies the unconditional probability inequality

$$\mathbb{P}\left\{|\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_j| \geq C\sqrt{\log p}\|\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_j\|\right\} \leq \frac{e}{p^3}. \tag{A.23}$$

By the fact $\|\boldsymbol{V}_{\gamma,-k}^{-1}\| \leq 1$, we have $\|\boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_j\| \leq \|\boldsymbol{z}_j\|$. Note that $z_{1j}^2-1, z_{2j}^2-1, \ldots, z_{nj}^2-1$ are independent centered sub-exponential random variables, by the Proposition A.1.3, we have

$$\mathbb{P}\left\{\left|\sum_{i=1}^n (z_{ij}^2 - 1)\right| \geq t\right\} \leq 2\exp\left\{-c\min\left(\frac{t^2}{K^2 n}, \frac{t}{K}\right)\right\},$$

where c and K are some positive constants. Take $t = K\sqrt{3n\log p}$, then we get

$$\mathbb{P}\left\{\left|\|\boldsymbol{z}_j\|^2 - n\right| \geq C\sqrt{n\log p}\right\} \leq \frac{2}{p^3}, \tag{A.24}$$

for some constant C. Combining the above (A.23) and (A.24) together, with probability at least $1-(2+e)/p^3$, there holds

$$(\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-1}\boldsymbol{z}_j)^2 \leq (\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_j)^2 \leq C(n + C\sqrt{n\log p})\log p.$$

By a similar argument with the fact that $\|\boldsymbol{V}_{\gamma,-k}^{-2}\| \leq 1$, we can have with probability at least $1-(2+e)/p^3$

$$(\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-2}\boldsymbol{z}_j)^2 \leq 2\left(\left(\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-2}\boldsymbol{z}_j\right)^2 + \left(\boldsymbol{z}_k^\top \boldsymbol{V}_{\gamma,-k}^{-1}\boldsymbol{z}_j\right)^2\right) \leq 4C(n + C\sqrt{n\log p})\log p,$$

for some constant C.

Above inequalities imply that with probability at least $1-(2+e)/p$, there hold

$$\max_{k\neq j}|\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-1}\boldsymbol{z}_j|^2 \le C(n+C\sqrt{n\log p})\log p, \quad \text{and}$$
$$\max_{k\neq j}|\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-2}\boldsymbol{z}_j|^2 \le 4C(n+C\sqrt{n\log p})\log p.$$

Then, there follows that

$$\max_{k\neq j}|\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-1}\boldsymbol{z}_j|^2 = O_P(n\log p) \quad \text{and} \quad \max_{k\neq j}|\boldsymbol{z}_k^\top \boldsymbol{V}_\gamma^{-2}\boldsymbol{z}_j|^2 = O_P(n\log p).$$

Next, define

$$\overline{\theta}_1(\gamma,\tau)=\kappa_{1,1}(\gamma,\tau)-2\frac{\kappa_{1,2}(\gamma,\tau)}{h_1(\gamma,\tau)}+\frac{\kappa_{2,2}(\gamma,\tau)}{h_1^2(\gamma,\tau)},$$

where

$$\kappa_{m,l}(\gamma,\tau)=\sum_{q_1=1}^{l}\sum_{q_2=1}^{m}\overline{a}_{q_1}^{(l)}(\gamma,\tau)\overline{a}_{q_2}^{(m)}(\gamma,\tau)h_{q_1+q_2}(\gamma,\tau),$$

and

$$\overline{a}_1^{(1)}(\gamma,\tau)=\frac{1}{\big(1+\tau\gamma h_1(\gamma,\tau)\big)^2},\quad \overline{a}_1^{(2)}(\gamma,\tau)=\frac{-2\tau\gamma h_2(\gamma,\tau)}{\big(1+\tau\gamma h_1(\gamma,\tau)\big)^3},\quad \overline{a}_2^{(2)}(\gamma,\tau)=\frac{1}{\big(1+\tau\gamma h_1(\gamma,\tau)\big)^2}.$$

Recall that $h_1(\gamma,\tau)$ and $h_2(\gamma,\tau)$ are defined in (2.8). Now, by the definition of $\eta_{ij}^{(l)}$ in (2.16), we can rewrite $\big(\boldsymbol{z}_i^\top \boldsymbol{B}_\gamma \boldsymbol{z}_j\big)^2$ as

$$\begin{aligned}\big(\boldsymbol{z}_i^\top \boldsymbol{B}_\gamma \boldsymbol{z}_j\big)^2 &= \left(\frac{1}{n}\boldsymbol{z}_i^\top \boldsymbol{V}_\gamma^{-1}\boldsymbol{z}_j-\frac{\frac{1}{n}\boldsymbol{z}_i^\top \boldsymbol{V}_\gamma^{-2}\boldsymbol{z}_j}{\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})}\right)^2\\ &= \left(\frac{1}{n}\eta_{ij}^{(1)}\right)^2-2\frac{\frac{1}{n}\eta_{ij}^{(1)}\frac{1}{n}\eta_{ij}^{(2)}}{\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})}+\frac{\left(\frac{1}{n}\eta_{ij}^{(2)}\right)^2}{\big(\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})\big)^2}. \qquad \text{(A.25)}\end{aligned}$$

The following results are implied by [24] in the supplementary material.

Proposition A.2.1 ([24]). *For any $i \neq j$ and $i, j \geq 1$, we have*

$$\eta_{ij}^{(1)} = \overline{a}_{1;ij}^{(1)} \eta_{ij;ij}^{(1)},$$

$$\eta_{ij}^{(2)} = \overline{a}_{1;ij}^{(2)} \eta_{ij;ij}^{(1)} + \overline{a}_{2;ij}^{(2)} \eta_{ij;ij}^{(2)},$$

with

$$\begin{aligned}
\overline{a}_{1;ij}^{(1)} =& \frac{1}{\left(1 + \frac{\gamma}{p}\eta_{ii;i}^{(1)}\right)\left(1 + \frac{\gamma}{p}\eta_{jj;ij}^{(1)}\right)}, \\
\overline{a}_{1;ij}^{(2)} =& \frac{-\frac{\gamma}{p}\eta_{ii;i}^{(2)}}{\left(1 + \frac{\gamma}{p}\eta_{ii;i}^{(1)}\right)^2\left(1 + \frac{\gamma}{p}\eta_{jj;ij}^{(1)}\right)} + \frac{-\frac{\gamma}{p}\eta_{jj;ij}^{(2)}}{\left(1 + \frac{\gamma}{p}\eta_{ii;i}^{(1)}\right)\left(1 + \frac{\gamma}{p}\eta_{jj;ij}^{(1)}\right)^2}, \\
\overline{a}_{2;ij}^{(2)} =& \overline{a}_{1;ij}^{(1)}.
\end{aligned}$$

And

$$\max_{1 \leq i \neq j \leq p} \max_{1 \leq l \leq 2} \max_{1 \leq q_1 \leq l} \left| \overline{a}_{q_1;ij}^{(l)} - \overline{a}_{q_1}^{(l)}(\gamma, \tau) \right| = O_{\mathrm{P}}\left(\sqrt{\frac{\log p}{n}}\right).$$

Furthermore,

$$\max_{1 \leq i \neq j \leq p} \max_{1 \leq l, m \leq 2} \max_{1 \leq q_1, q_2 \leq l} \left| \overline{a}_{q_1;ij}^{(l)} \overline{a}_{q_2;ij}^{(m)} - \overline{a}_{q_1}^{(l)}(\gamma, \tau) \overline{a}_{q_2}^{(m)}(\gamma, \tau) \right| = O_{\mathrm{P}}\left(\sqrt{\frac{\log p}{n}}\right) \tag{A.26}$$

Proposition A.2.2 ([24]). *For any $l \geq 1$ and $1 \leq i \neq j$,*

$$\frac{1}{n}\left|\mathrm{trace}\left(\boldsymbol{V}_{\gamma}^{-l}\right) - \mathrm{trace}\left(\boldsymbol{V}_{\gamma,-ij}^{-l}\right)\right| \leq \frac{1}{n} 2^{l+1}. \tag{A.27}$$

Proposition A.2.3 ([24]). *For any $1 \leq q_1, q_2 \leq 2$, define*

$$d_{ij}^{(q_1,q_2)} := \frac{1}{n}\eta_{ij;ij}^{(q_1)}\eta_{ij;ij}^{(q_2)} - \frac{1}{n}\mathrm{trace}\left(\boldsymbol{V}_{\gamma,-ij}^{-(q_1+q_2)}\right).$$

Then the following statements are true.

1) For some constant $K_1 > 0$,

$$\max_{1 \leq i \neq j \leq p} \mathbb{E}\left[\left(d_{ij}^{(q_1,q_2)}\right)^2\right] \leq K_1.$$

2) For any $i \neq j \neq i'$, $j' \neq i'$(either $j = j'$ or not) and some constant $K_2 > 0$,

$$\max_{i \neq j \neq i, j' \neq i'} \left| \mathbb{E}\left[d_{ij}^{(q_1,q_2)} d_{i'j'}^{(q_1,q_2)} \right] \right| \leq \frac{K_2}{\sqrt{n}}.$$

By (A.25), let's first show that for $1 \leq l, m \leq 2$,

$$\text{(A.28)} \quad \frac{1}{p(p-1)} \sum_{i \neq j} \frac{1}{n} \eta_{ij}^{(l)} \eta_{ij}^{(m)} \xrightarrow{P} \kappa_{m.l}(\gamma,\tau), \text{ and } \sum_{i \neq j} \beta_i^2 \beta_j^2 \frac{1}{n} \eta_{ij}^{(l)} \eta_{ij}^{(m)} \xrightarrow{P} \|\boldsymbol{\beta}\|^4 \kappa_{m.l}(\gamma,\tau).$$

Since $\eta_{ij;ij}^{(q_1)} \eta_{ij;ij}^{(q_2)} > 0$ for any $q_1, q_2 = 1, 2, \ldots$, by Proposition A.2.1 we can have

$$\begin{aligned}
\sum_{i \neq j} \beta_i^2 \beta_j^2 \frac{1}{n} \eta_{ij}^{(l)} \eta_{ij}^{(m)} &= \sum_{q_1=1}^{l} \sum_{q_2=1}^{m} \left(\sum_{i \neq j} \overline{a}_{q_1;ij}^{(l)} \overline{a}_{q_2;ij}^{(m)} \beta_i^2 \beta_j^2 \frac{1}{n} \eta_{ij;ij}^{(q_1)} \eta_{ij;ij}^{(q_2)} \right) \\
\text{(A.29)} \qquad &= \sum_{q_1=1}^{l} \sum_{q_2=1}^{m} \overline{a}_{q_1}^{(l)}(\gamma,\tau) \overline{a}_{q_2}^{(m)}(\gamma,\tau) \left(\sum_{i \neq j} \beta_i^2 \beta_j^2 \frac{1}{n} \eta_{ij;ij}^{(q_1)} \eta_{ij;ij}^{(q_2)} \right) + o_P(1),
\end{aligned}$$

and

$$\text{(A.30)} \quad \frac{1}{p(p-1)} \sum_{i \neq j} \frac{1}{n} \eta_{ij}^{(l)} \eta_{ij}^{(m)} = \sum_{q_2=1}^{m} \overline{a}_{q_1}^{(l)}(\gamma,\tau) \overline{a}_{q_2}^{(m)}(\gamma,\tau) \frac{1}{p(p-1)} \sum_{i \neq j} \frac{1}{n} \eta_{ij;ij}^{(q_1)} \eta_{ij;ij}^{(q_2)} + o_P(1).$$

We know that

$$\mathbb{E}\left[\sum_{i \neq j} \beta_i^2 \beta_j^2 d_{ij}^{(q_1,q_2)} \right] = \sum_{i \neq j} \beta_i^2 \beta_j^2 \, \mathbb{E}\left[\mathbb{E}\left[\frac{1}{n} \eta_{ij;ij}^{(q_1)} \eta_{ij;ij}^{(q_2)} - \frac{1}{n} \operatorname{trace}\left(\boldsymbol{V}_{\gamma,-ij}^{(q_1+q_2)} \right) | \boldsymbol{V}_{\gamma,-ij}^{(q_1+q_2)} \right] \right] = 0,$$

and by Proposition A.2.3, we have that

$$\begin{aligned}
\mathbb{E}\left(\sum_{i \neq j} \beta_i^2 \beta_j^2 d_{ij}^{(q_1,q_2)} \right)^2 &= \sum_{i \neq j} \beta_i^4 \beta_j^4 \, \mathbb{E}\left[\left(d_{ij}^{(q_1,q_2)} \right)^2 \right] + 2 \sum_{i \neq i' \neq j} \beta_i^2 \beta_{i'}^2 \beta_j^4 \, \mathbb{E}\left[\left(d_{ij}^{(q_1,q_2)} d_{i'j}^{(q_1,q_2)} \right) \right] \\
&\quad + \sum_{i \neq j \neq i' \neq j'} \beta_i^2 \beta_j^2 \beta_{i'}^2 \beta_{j'}^2 \, \mathbb{E}\left[\left(d_{ij}^{(q_1,q_2)} d_{i'j'}^{(q_1,q_2)} \right) \right] \\
&\leq K_1 \|\boldsymbol{\beta}\|_4^8 + 2 \frac{K_2}{n} \|\boldsymbol{\beta}\|_2^4 \|\boldsymbol{\beta}\|_4^4 + \frac{K_2}{n} \|\boldsymbol{\beta}\|_2^8 \\
&= o_P(1).
\end{aligned}$$

This implies that

$$\sum_{i\neq j}\beta_i^2\beta_j^2\frac{1}{n}\eta_{ij;ij}^{(q_1)}\eta_{ij;ij}^{(q_2)}=\sum_{i\neq j}\beta_i^2\beta_j^2\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_{\gamma,-ij}^{(q_1+q_2)}\right)+o_P(1),$$

where by Proposition A.2.2 and Lemma 2.4.0.2,

$$\left|\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_{\gamma,-ij}^{(q_1+q_2)}\right)-h_{q_1+q_2}(\gamma,\tau)\right|=o_P(1),$$

and $\sum_{i\neq j}\beta_i^2\beta_j^2 h_{q_1+q_2}(\gamma,\tau)=\|\boldsymbol{\beta}\|^4 h_{q_1+q_2}(\gamma,\tau)+o_P(1)$ since $\sum_{i=1}^p\beta_i^4=o_P(1)$. Thus

$$\sum_{i\neq j}\beta_i^2\beta_j^2\frac{1}{n}\eta_{ij;ij}^{(q_1)}\eta_{ij;ij}^{(q_2)}=\|\boldsymbol{\beta}\|^4 h_{q_1+q_2}(\gamma,\tau)+o_P(1). \tag{A.31}$$

Thus by (A.29) and (A.32), we can have

$$\sum_{i\neq j}\beta_i^2\beta_j^2\frac{1}{n}\eta_{ij}^{(l)}\eta_{ij}^{(m)}=\|\boldsymbol{\beta}\|^4\kappa_{m,l}(\gamma,\tau)+o_P(1).$$

Similarly,

$$\mathbb{E}\left[\frac{1}{p(p-1)}\sum_{i\neq j}\frac{1}{n}\eta_{ij;ij}^{(q_1)}\eta_{ij;ij}^{(q_2)}\right]=0,$$

and

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{p(p-1)}\sum_{i\neq j}\frac{1}{n}\eta_{ij;ij}^{(q_1)}\eta_{ij;ij}^{(q_2)}\right)^2\\
&=\frac{1}{p^2(p-1)^2}\left(\sum_{i\neq j}\mathbb{E}\left[\left(d_{ij}^{(q_1,q_2)}\right)^2\right]+2\sum_{i\neq i'\neq j}\mathbb{E}\left[\left(d_{ij}^{(q_1,q_2)}d_{i'j}^{(q_1,q_2)}\right)\right]\right.\\
&\left.\quad+\sum_{i\neq j\neq i'\neq j'}\mathbb{E}\left[\left(d_{ij}^{(q_1,q_2)}d_{i'j'}^{(q_1,q_2)}\right)\right]\right)\\
&\leq\frac{1}{p^2(p-1)^2}\left(p(p-1)K_1+2p(p-1)(p-2)\frac{K_2}{n}+p(p-1)(p-2)(p-3)\frac{K_2}{n}\right)\\
&=o_P(1).
\end{aligned}$$

Then by Proposition A.2.2 and Lemma 2.4.0.2, we can have that

$$\frac{1}{p(p-1)}\sum_{i\neq j}\frac{1}{n}\eta_{ij;ij}^{(q_1)}\eta_{ij;ij}^{(q_2)} = h_{q_1+q_2}(\gamma,\tau), \tag{A.32}$$

which implies

$$\frac{1}{p(p-1)}\sum_{i\neq j}\frac{1}{n}\eta_{ij}^{(q_1)}\eta_{ij}^{(q_2)} = \kappa_{m,l}(\gamma,\tau) + o_P(1),$$

by (A.30).

Now we have proved (A.28), then by (A.25) and Lemma 2.4.0.2, there holds that

$$\begin{cases} \dfrac{n}{p(p-1)}\displaystyle\sum_{i\neq j}\left(\boldsymbol{z}_i^\top \boldsymbol{B}_\gamma \boldsymbol{z}_j\right)^2 & = \overline{\theta}_1(\gamma,\tau) + o_P(1), \\ n\displaystyle\sum_{i\neq j}\beta_i^2\beta_j^2\left(\boldsymbol{z}_i^\top \boldsymbol{B}_\gamma \boldsymbol{z}_j\right)^2 & = \|\boldsymbol{\beta}\|^4\overline{\theta}_1(\gamma,\tau) + o_P(1). \end{cases}$$

A.2.6. Proof of Lemma 2.4.1.6. Let $\tilde{\boldsymbol{z}}_i^\top$ be the ith row of $\boldsymbol{Z}$. By Sherman-Morrison-Woodbury formula, we have

$$\boldsymbol{V}_\gamma^{-1} = \left(\boldsymbol{I}_n + \frac{\gamma}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)^{-1} = \boldsymbol{I}_n - \frac{\gamma}{p}\boldsymbol{Z}\left(\boldsymbol{I}_p + \frac{\gamma}{p}\boldsymbol{Z}^\top\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\top, \tag{A.33}$$

and

$$\begin{aligned} \boldsymbol{V}_\gamma^{-2} &= \left(\boldsymbol{V}_\gamma^{-1}\right)^2 \\ &= \boldsymbol{I}_n - 2\frac{\gamma}{p}\boldsymbol{Z}\left(\boldsymbol{I}_q + \frac{\gamma}{p}\boldsymbol{Z}^\top\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\top + \left(\frac{\gamma}{p}\right)^2\left(\boldsymbol{Z}\left(\boldsymbol{I}_p + \frac{\gamma}{p}\boldsymbol{Z}^\top\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\top\right)^2. \end{aligned} \tag{A.34}$$

Combining (A.33) and (A.34) gives

$$\left(\boldsymbol{V}_\gamma^{-1}\right)_{ii} = 1 - \frac{\gamma}{p}\tilde{\boldsymbol{z}}_i^\top\left(\boldsymbol{I}_p + \frac{\gamma}{p}\boldsymbol{Z}^\top\boldsymbol{Z}\right)^{-1}\tilde{\boldsymbol{z}}_i,$$

and

$$\left(\boldsymbol{V}_\gamma^{-2}\right)_{ii} = 1 - 2\frac{\gamma}{p}\tilde{\boldsymbol{z}}_i^\top \left(\boldsymbol{I}_p + \frac{\gamma}{p}\boldsymbol{Z}^\top\boldsymbol{Z}\right)^{-1}\tilde{\boldsymbol{z}}_i + \left(\frac{\gamma}{p}\right)^2 \tilde{\boldsymbol{z}}_i^\top \left(\boldsymbol{I}_p + \frac{\gamma}{p}\boldsymbol{Z}^\top\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\top\boldsymbol{Z}\left(\boldsymbol{I}_p + \frac{\gamma}{p}\boldsymbol{Z}^\top\boldsymbol{Z}\right)^{-1}\tilde{\boldsymbol{z}}_i.$$

where $\tilde{\boldsymbol{z}}_i$ is the i-th column of $\boldsymbol{Z}^\top$. Define

$$\widetilde{\boldsymbol{V}}_\gamma = \boldsymbol{I}_p + \frac{\gamma}{p}\boldsymbol{Z}^\top\boldsymbol{Z}, \tag{A.35}$$

then we can rewrite $\left(\boldsymbol{V}_\gamma^{-1}\right)_{ii}$ and $\left(\boldsymbol{V}_\gamma^{-2}\right)_{ii}$ as

$$\left(\boldsymbol{V}_\gamma^{-1}\right)_{ii} = 1 - \frac{\gamma}{p}\tilde{\boldsymbol{z}}_i\widetilde{\boldsymbol{V}}_\gamma^{-1}\tilde{\boldsymbol{z}}_i^\top, \quad \left(\boldsymbol{V}_\gamma^{-2}\right)_{ii} = 1 - \frac{\gamma}{p}\tilde{\boldsymbol{z}}_i\widetilde{\boldsymbol{V}}_\gamma^{-1}\tilde{\boldsymbol{z}}_i^\top - \frac{\gamma}{p}\tilde{\boldsymbol{z}}_i\widetilde{\boldsymbol{V}}_\gamma^{-2}\tilde{\boldsymbol{z}}_i^\top.$$

Furthermore, by a similar argument, it can be shown that for $l = 1, 2, \ldots$

$$\boldsymbol{V}_\gamma^{-l} = \boldsymbol{I}_n - \frac{\gamma}{p}\sum_{q=1}^{l}\boldsymbol{Z}\widetilde{\boldsymbol{V}}_\gamma^{-q}\boldsymbol{Z}^\top, \tag{A.36}$$

with

$$\left(\boldsymbol{V}_\gamma^{-l}\right)_{ii} = 1 - \frac{\gamma}{p}\sum_{q=1}^{l}\tilde{\boldsymbol{z}}_i^\top\widetilde{\boldsymbol{V}}_\gamma^{-q}\tilde{\boldsymbol{z}}_i. \tag{A.37}$$

Similar to (2.20) in Lemma 2.4.1.1, combining the leave-one-out technique and Hanson-Wright inequality, taking the uniform bound gives

$$\max_{i\in[n]}\left|\frac{1}{p}\tilde{\boldsymbol{z}}_i^\top\widetilde{\boldsymbol{V}}_\gamma^{-q}\tilde{\boldsymbol{z}}_i - \frac{1}{np}\operatorname{trace}\left(\widetilde{\boldsymbol{V}}_\gamma^{-q}\boldsymbol{Z}^\top\boldsymbol{Z}\right)\right| = O_P\left(\sqrt{\frac{\log n}{n}}\right). \tag{A.38}$$

Together with (A.36) and (A.37) yields (2.27).

Note that we have

$$(\boldsymbol{B}_\gamma)_{ii} = \frac{1}{n}(\boldsymbol{V}_\gamma^{-1})_{ii} - \frac{\frac{1}{n}(\boldsymbol{V}_\gamma^{-2})_{ii}}{\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})},$$

$$(\boldsymbol{B}_\gamma)_{ii}^2 = \left(\frac{1}{n}\right)^2 \left(\boldsymbol{V}_\gamma^{-1}\right)_{ii}^2 - \frac{2\left(\frac{1}{n}\right)^2 \left(\boldsymbol{V}_\gamma^{-1}\right)_{ii}\left(\boldsymbol{V}_\gamma^{-2}\right)_{ii}}{\left(\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})\right)^2} + \frac{\left(\frac{1}{n}\right)^2 \left(\boldsymbol{V}_\gamma^{-2}\right)_{ii}^2}{\left(\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})\right)^2}.$$

By (2.27), we have

$$\begin{aligned}
\max_{i\in[n]}\left|(\boldsymbol{B}_\gamma)_{ii} - \frac{1}{n}\operatorname{trace}(\boldsymbol{B}_\gamma)\right| &\le \max_{i\in[n]}\left|\frac{1}{n}\left(\boldsymbol{V}_\gamma^{-1}\right)_{ii} - \frac{1}{n^2}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right| \\
&\quad + \frac{\max_{i\in[n]}\left|\frac{1}{n}\left(\boldsymbol{V}_\gamma^{-2}\right)_{ii} - \frac{1}{n^2}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\right)\right|}{\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})} \\
&= O_P\left(\frac{1}{n}\sqrt{\frac{\log n}{n}}\right),
\end{aligned}$$

and consequently

$$\begin{aligned}
\max_{i\in[n]}\left|(\boldsymbol{B}_\gamma)_{ii}^2 - \left(\frac{1}{n}\operatorname{trace}(\boldsymbol{B}_\gamma)\right)^2\right| &\le \max_{i\in[n]}\left|(\boldsymbol{B}_\gamma)_{ii} - \frac{1}{n}\operatorname{trace}(\boldsymbol{B}_\gamma)\right|^2 \\
&\quad + 2\frac{1}{n}|\operatorname{trace}(\boldsymbol{B}_\gamma)|\max_{i\in[n]}\left|(\boldsymbol{B}_\gamma)_{ii} - \frac{1}{n}\operatorname{trace}(\boldsymbol{B}_\gamma)\right| \\
&= O_P\left(\frac{1}{n^2}\sqrt{\frac{\log n}{n}}\right),
\end{aligned}$$

which yield (2.28) and (2.29).

From (A.36), we have

$$\left(\boldsymbol{V}_\gamma^{-l}\right)_{ij} = -\frac{\gamma}{p}\sum_{k=1}^{l} \tilde{\boldsymbol{z}}_i^\top \widetilde{\boldsymbol{V}}_\gamma^{-k}\tilde{\boldsymbol{z}}_j.$$

As with (2.25) in Lemma 2.4.1.5, we can have

$$\max_{i\neq j}\left|\tilde{\boldsymbol{z}}_i^\top \widetilde{\boldsymbol{V}}_\gamma^{-k}\tilde{\boldsymbol{z}}_j\right| = O_P(\sqrt{n\log n}).$$

Then (2.30) can be easily obtained by the fact

$$(\boldsymbol{B}_{\gamma_0})_{ij} = \frac{1}{n}\left(\boldsymbol{V}_{\gamma_0}^{-1}\right)_{ij} - \frac{\frac{1}{n}\left(\boldsymbol{V}_{\gamma_0}^{-2}\right)_{ij}}{\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_{\gamma_0}^{-1})}.$$

A.2.7. Proof of Lemma 2.4.1.7. To prove (2.31), we can know that

$$\begin{aligned}(\boldsymbol{B}_{\gamma})_{ij}^2 &= \frac{1}{n^2}\left(\boldsymbol{V}_{\gamma}^{-1}\right)_{ij}^2 - \frac{\frac{2}{n^2}\left(\boldsymbol{V}_{\gamma}^{-1}\right)_{ij}\left(\boldsymbol{V}_{\gamma}^{-2}\right)_{ij}}{\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_{\gamma}^{-1})} + \frac{\frac{1}{n^2}\left(\boldsymbol{V}_{\gamma}^{-2}\right)_{ij}^2}{\left(\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_{\gamma}^{-1})\right)^2} \\ &= \frac{\gamma^2}{n^2}\frac{1}{p^2}\left(\tilde{\eta}_{ij}^{(1)}\right)^2 - 2\frac{\gamma^2}{n^2}\frac{1}{p^2}\frac{\left(\tilde{\eta}_{ij}^{(1)}\right)^2 + \tilde{\eta}_{ij}^{(1)}\tilde{\eta}_{ij}^{(2)}}{\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_{\gamma}^{-1})} + \frac{\gamma^2}{n^2}\frac{1}{p^2}\frac{\left(\tilde{\eta}_{ij}^{(1)}\right)^2 + 2\tilde{\eta}_{ij}^{(1)}\tilde{\eta}_{ij}^{(2)} + \left(\tilde{\eta}_{ij}^{(2)}\right)^2}{\left(\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_{\gamma}^{-1})\right)^2},\end{aligned} \tag{A.39}$$

where

$$\tilde{\eta}_{ij}^{(k)} := \tilde{\boldsymbol{z}}_i^\top \widetilde{\boldsymbol{V}}_{\gamma}^{-k}\tilde{\boldsymbol{z}}_j.$$

Define

$$\overline{\theta}_2 = \gamma^2\tau\left(\tilde{\kappa}_{1,1}(\gamma,\tau) - 2\frac{\tilde{\kappa}_{1,1}(\gamma,\tau) + \tilde{\kappa}_{1,2}(\gamma,\tau)}{h_1(\gamma,\tau)} + \frac{\tilde{\kappa}_{1,1}(\gamma,\tau) + 2\tilde{\kappa}_{1,2}(\gamma,\tau) + \tilde{\kappa}_{2,2}(\gamma,\tau)}{(h_1(\gamma,\tau))^2}\right).$$

where

$$\tilde{\kappa}_{m,l}(\gamma,\tau) = \sum_{q_1=1}^{l}\sum_{q_2=1}^{m}\tilde{\overline{a}}_{q_1}^{(l)}(\gamma,\tau)\tilde{\overline{a}}_{q_2}^{(m)}(\gamma,\tau)\tilde{h}_{q_1+q_2}(\gamma,\tau),$$

and

$$\tilde{\overline{a}}_1^{(1)}(\gamma,\tau) = \frac{1}{\left(1+\gamma\tilde{h}_1(\gamma,\tau)\right)^2}, \quad \tilde{\overline{a}}_1^{(2)}(\gamma,\tau) = \frac{-2\gamma\tilde{h}_2(\gamma,\tau)}{\left(1+\gamma\tilde{h}_1(\gamma,\tau)\right)^3}, \quad \tilde{\overline{a}}_2^{(2)}(\gamma,\tau) = \frac{1}{\left(1+\gamma\tilde{h}_1(\gamma,\tau)\right)^2}.$$

Similar to the definition of $h_l(\gamma,\tau)$, $\tilde{h}_l(\gamma,\tau)$ is the limit of $\frac{1}{p}\operatorname{trace}\left(\widetilde{\boldsymbol{V}}_{\gamma}^{-l}\right)$.

Then similar to Proposition A.2.1, using the leave-two-out technique, there holds that

$$\begin{aligned}\tilde{\eta}_{ij}^{(1)} &= \tilde{\overline{a}}_{1;ij}^{(1)}\tilde{\eta}_{ij;ij}^{(1)}, \\ \eta_{ij}^{(2)} &= \tilde{\overline{a}}_{1;ij}^{(2)}\tilde{\eta}_{ij;ij}^{(1)} + \tilde{\overline{a}}_{2;ij}^{(2)}\tilde{\eta}_{ij;ij}^{(2)},\end{aligned}$$

with

$$\begin{aligned}
\tilde{\bar{a}}_{1;ij}^{(1)} =& \frac{1}{\left(1+\frac{\gamma}{p}\tilde{\eta}_{ii;i}^{(1)}\right)\left(1+\frac{\gamma}{p}\tilde{\eta}_{jj;ij}^{(1)}\right)}, \\
\tilde{\bar{a}}_{1;ij}^{(2)} =& \frac{-\frac{\gamma}{p}\tilde{\eta}_{ii;i}^{(2)}}{\left(1+\frac{\gamma}{p}\tilde{\eta}_{ii;i}^{(1)}\right)^2\left(1+\frac{\gamma}{p}\tilde{\eta}_{jj;ij}^{(1)}\right)} + \frac{-\frac{\gamma}{p}\tilde{\eta}_{jj;ij}^{(2)}}{\left(1+\frac{\gamma}{p}\tilde{\eta}_{ii;i}^{(1)}\right)\left(1+\frac{\gamma}{p}\tilde{\eta}_{jj;ij}^{(1)}\right)^2}, \\
\tilde{\bar{a}}_{2;ij}^{(2)} =& \tilde{\bar{a}}_{1;ij}^{(1)}.
\end{aligned}$$

And

$$\max_{1\le i\neq j\le p}\max_{1\le l\le 2}\max_{1\le q_1\le l}\left|\tilde{\bar{a}}_{q_1;ij}^{(l)} - \tilde{\bar{a}}_{q_1}^{(l)}(\gamma,\tau)\right| = O_{\mathrm{P}}\left(\sqrt{\frac{\log p}{n}}\right).$$

Furthermore,

$$\max_{1\le i\neq j\le p}\max_{1\le l,m\le 2}\max_{1\le q_1,q_2\le l}\left|\tilde{\bar{a}}_{q_1;ij}^{(l)}\tilde{\bar{a}}_{q_2;ij}^{(m)} - \tilde{\bar{a}}_{q_1}^{(l)}(\gamma,\tau)\tilde{\bar{a}}_{q_2}^{(m)}(\gamma,\tau)\right| = O_{\mathrm{P}}\left(\sqrt{\frac{\log p}{n}}\right) \tag{A.40}$$

Similar to (A.32) and (A.31) , we can have that

$$\frac{1}{n(n-1)}\sum_{i\neq j}\frac{1}{p}\tilde{\eta}_{ij;ij}^{(q_1)}\tilde{\eta}_{ij;ij}^{(q_2)} = \frac{1}{p}\operatorname{trace}\left(\widetilde{\boldsymbol{V}}_{\gamma}^{-(q_1+q_2)}\right) + o_P(1),$$

and

$$\frac{1}{n^2}\sum_{i\neq j}\varepsilon_i^2\varepsilon_j^2\frac{1}{p}\tilde{\eta}_{ij;ij}^{(q_1)}\tilde{\eta}_{ij;ij}^{(q_2)} = \sigma_0^4\frac{1}{p}\operatorname{trace}\left(\widetilde{\boldsymbol{V}}_{\gamma}^{-(q_1+q_2)}\right) + o_P(1),$$

by (2.13) in Lemma 2.4.0.1. Then similar to the proof of Lemma 2.4.1.5, (2.31) can be obtained from (A.39).

A.2.8. Proof of Lemma 2.4.1.8. For any $k = 1,\cdots,p$, denote $\boldsymbol{Z}_{-k} = [\boldsymbol{z}_1,\cdots,\boldsymbol{z}_{k-1},\boldsymbol{z}_{k+1},\cdots]$, then

$$\boldsymbol{V}_{\gamma,-k} = \boldsymbol{V}_{\gamma} - \frac{\gamma}{p}\boldsymbol{z}_k\boldsymbol{z}_k^\top = \boldsymbol{I}_n + \frac{\gamma}{p}\boldsymbol{Z}_{-k}\boldsymbol{Z}_{-k}^\top.$$

Similar to the proof of (2.27) in Lemma 2.4.1.6, we can define

$$\widetilde{\boldsymbol{V}}_{\gamma,-k} = \boldsymbol{I}_{p-1} + \frac{\gamma}{p}\boldsymbol{Z}_{-k}^\top\boldsymbol{Z}_{-k}, \tag{A.41}$$

and $\tilde{\boldsymbol{z}}_{i,-k}^\top$ is the i-th row of $\boldsymbol{Z}_{-k}$. Then we can have that

$$\left(\widetilde{\boldsymbol{V}}_{\gamma,-k}\right)_{ii} = 1 - \frac{\gamma}{p}\sum_{q=1}^{l} \tilde{\boldsymbol{z}}_{i,-k}^\top \widetilde{\boldsymbol{V}}_{\gamma,-k}^{-q} \tilde{\boldsymbol{z}}_{i,-k}. \tag{A.42}$$

Again, similar to (2.20) in Lemma 2.4.1.1, combining the leave-one-out technique and Hanson-Wright inequality (taking $t = \sqrt{\frac{3\log p}{c}}\|\widetilde{\boldsymbol{V}}_{\gamma,-ik}^{-q}\|_F$), taking the uniform bound gives

$$\max_{k\in[p]}\max_{i\in[n]}\left|\frac{1}{p}\tilde{\boldsymbol{z}}_{i,-k}^\top \widetilde{\boldsymbol{V}}_{\gamma,-k}^{-q}\tilde{\boldsymbol{z}}_{i,-k} - \frac{1}{np}\,\mathrm{trace}\left(\widetilde{\boldsymbol{V}}_{\gamma,-k}^{-q}\boldsymbol{Z}_{-k}^\top\boldsymbol{Z}_{-k}\right)\right| = O_P\left(\sqrt{\frac{\log n}{n}}\right). \tag{A.43}$$

Then (A.36) and (A.37) implies

$$\max_{k\in[p]}\max_{i\in[n]}\left|\left(\boldsymbol{V}_{\gamma,-k}^{-l}\right)_{ii} - \frac{1}{n}\,\mathrm{trace}\left(\boldsymbol{V}_{\gamma,-k}^{-l}\right)\right| = O_P\left(\sqrt{\frac{\log n}{n}}\right), \quad l = 1,2,3,4. \tag{A.44}$$

By (2.17) in Lemma 2.4.1.1 and (A.44) we can get (2.32).

A.2.9. Proof of Lemma 2.4.1.9. In this section, we focus on the conditional variance $\mathrm{Var}[\Delta(\gamma_0)|\boldsymbol{Z},\boldsymbol{\varepsilon}]$. With $\boldsymbol{y}$ defined in (2.33), we have

$$\Delta(\gamma_0) = \underbrace{\boldsymbol{\xi}^\top\boldsymbol{\Lambda}_{\boldsymbol{\beta}}\boldsymbol{Z}^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{Z}\boldsymbol{\Lambda}_{\boldsymbol{\beta}}\boldsymbol{\xi}}_{M_1} + \underbrace{2\boldsymbol{\xi}^\top\boldsymbol{\Lambda}_{\boldsymbol{\beta}}\boldsymbol{Z}^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}}\boldsymbol{\zeta}}_{M_2} + \underbrace{\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}}\boldsymbol{B}_{\gamma_0}\boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}}\boldsymbol{\zeta}}_{M_3},$$

Then it is obvious that

$$\mathrm{Var}[\Delta(\gamma_0)|\boldsymbol{Z},\boldsymbol{\varepsilon}] = \mathbb{E}\left[\Delta^2(\gamma_0)|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] - \tilde{\Delta}_*^2(\gamma_0), \tag{A.45}$$

where by (2.35)

$$\tilde{\Delta}_*^2(\gamma_0) = \left(\sum_{k=1}^{p}\beta_k^2\boldsymbol{z}_k^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{z}_k + \mathrm{trace}\left(\boldsymbol{\Lambda}_{\varepsilon}^2\boldsymbol{B}_{\gamma_0}\right)\right)^2 \tag{A.46}$$

and we can have

$$\begin{aligned}\mathbb{E}\left[\Delta^2(\gamma_0)|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] &= \mathbb{E}\left[M_1^2|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] + \mathbb{E}\left[M_2^2|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] + \mathbb{E}\left[M_3^2|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] \\ &\quad + 2\,\mathbb{E}\left[M_1M_2|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] + 2\,\mathbb{E}\left[M_1M_3|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] + 2\,\mathbb{E}\left[M_2M_2|\boldsymbol{Z},\boldsymbol{\varepsilon}\right].\end{aligned} \tag{A.47}$$

Define

$$\tilde{g}_{k,j,l,m} = \beta_k\beta_j\beta_l\beta_m\xi_k\xi_j\xi_l\xi_m(\boldsymbol{z}_k^\top \boldsymbol{B}_{\gamma_0}\boldsymbol{z}_j)(\boldsymbol{z}_l^\top \boldsymbol{B}_{\gamma_0}\boldsymbol{z}_m).$$

Since ξ_i's are i.i.d. Rademacher random variables , if (k, j, l, m) has an odd multiplicity, we have $\mathbb{E}\left[\xi_k\xi_j\xi_l\xi_m\right] = 0$. This implies that

$$\begin{aligned}\mathbb{E}\left[M_1^2|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] &= \mathbb{E}\left[M_1^2|\boldsymbol{Z}\right]\\ &= \sum_{k=1}^{p}\mathbb{E}\left[\tilde{g}_{k,k,k,k}|\boldsymbol{Z}\right] + \sum_{j\neq i}\mathbb{E}\left[\tilde{g}_{i,i,j,j}|\boldsymbol{Z}\right] + \sum_{j\neq i}\mathbb{E}\left[\tilde{g}_{i,j,i,j}|\boldsymbol{Z}\right] + \sum_{j\neq i}\mathbb{E}\left[\tilde{g}_{i,j,j,i}|\boldsymbol{Z}\right],\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[M_2^2|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] &= \mathbb{E}\left[4\boldsymbol{\xi}^\top\boldsymbol{\Lambda}_{\boldsymbol{\beta}}\boldsymbol{Z}^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}}\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}}\boldsymbol{B}_{\gamma_0}\boldsymbol{Z}\boldsymbol{\Lambda}_{\boldsymbol{\beta}}\boldsymbol{\xi}|\boldsymbol{Z},\boldsymbol{\varepsilon}\right]\\ &= 4\sum_{k=1}^{p}\beta_k^2\boldsymbol{z}_k^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{\Lambda}_{\varepsilon}^2\boldsymbol{B}_{\gamma_0}\boldsymbol{z}_k,\end{aligned}$$

$$\mathbb{E}\left[M_3^2|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] = \mathbb{E}\left[\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}}\boldsymbol{B}_{\gamma_0}\boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}}\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}}\boldsymbol{B}_{\gamma_0}\boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}}\boldsymbol{\zeta}|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] = \operatorname{trace}\left(\left(\boldsymbol{\Lambda}_{\varepsilon}^2\boldsymbol{B}_{\gamma_0}\right)^2\right),$$

$$\mathbb{E}\left[M_1M_2|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] = \mathbb{E}\left[2\boldsymbol{\xi}^\top\boldsymbol{\Lambda}_{\boldsymbol{\beta}}\boldsymbol{Z}^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{Z}\boldsymbol{\Lambda}_{\boldsymbol{\beta}}\boldsymbol{\xi}\boldsymbol{\xi}^\top\boldsymbol{\Lambda}_{\boldsymbol{\beta}}\boldsymbol{Z}^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}}\boldsymbol{\zeta}|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] = 0,$$

$$\mathbb{E}\left[M_1M_3|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] = \mathbb{E}\left[M_1|\boldsymbol{Z}\right]\mathbb{E}\left[M_3|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] = \sum_{k=1}^{p}\beta_k^2\boldsymbol{z}_k^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{z}_k\cdot\sigma_0^2\operatorname{trace}\left(\boldsymbol{\Lambda}_{\varepsilon}^2\boldsymbol{B}_{\gamma_0}\right),$$

$$\mathbb{E}\left[M_2M_3|\boldsymbol{Z},\boldsymbol{\varepsilon}\right] = \mathbb{E}\left[2\boldsymbol{\xi}^\top\boldsymbol{\Lambda}_{\boldsymbol{\beta}}\boldsymbol{Z}^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}}\boldsymbol{\zeta}\boldsymbol{\zeta}^\top\boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}}\boldsymbol{B}_{\gamma_0}\boldsymbol{\Lambda}_{\boldsymbol{\varepsilon}}\boldsymbol{\zeta}\right] = 0.$$

Combining these results with (A.47), we can have

$$\mathbb{E}\left[\Delta^2(\gamma_0)|\boldsymbol{Z},\boldsymbol{\varepsilon}\right]$$

$$= \sum_{k=1}^{p}\mathbb{E}\left[\tilde{g}_{k,k,k,k}|\boldsymbol{Z}\right] + \sum_{j\neq i}\mathbb{E}\left[\tilde{g}_{i,i,j,j}|\boldsymbol{Z}\right] + \sum_{j\neq i}\mathbb{E}\left[\tilde{g}_{i,j,i,j}|\boldsymbol{Z}\right] + \sum_{j\neq i}\mathbb{E}\left[\tilde{g}_{i,j,j,i}|\boldsymbol{Z}\right]$$

(A.48)

$$+ 4\sum_{k=1}^{p}\beta_k^2\boldsymbol{z}_k^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{B}_{\gamma_0}\boldsymbol{z}_k + \operatorname{trace}\left(\left(\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{B}_{\gamma_0}\right)^2\right) + 2\sum_{k=1}^{p}\beta_k^2\boldsymbol{z}_k^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{z}_k\cdot\sigma_0^2\operatorname{trace}\left(\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{B}_{\gamma_0}\right).$$

Thus by (A.45), (A.47) and (A.48) we can have

$$\operatorname{Var}\left[\sqrt{n}(\Delta(\gamma_0))|\boldsymbol{Z},\boldsymbol{\varepsilon}\right]$$

$$= 2n\sum_{j\neq i}\beta_i^2\beta_j^2\left(\boldsymbol{z}_i^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{z}_j\right)^2 + 4n\sum_{k=1}^{p}\beta_k^2\boldsymbol{z}_k^\top\boldsymbol{B}_{\gamma_0}\boldsymbol{\Lambda}_\varepsilon^2\boldsymbol{B}_{\gamma_0}\boldsymbol{z}_k + 2n\sum_{j\neq i}\varepsilon_i^2\varepsilon_j^2(\boldsymbol{B}_{\gamma_0})_{ij}^2.$$

A.2.10. Proof of Lemma 2.4.1.10. Recall that $\Delta_{**}(\gamma)$ is of the form (2.38), so we need to study the asymptotics of $\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)$, $\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\right)$, $\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}ZZ^\top\right)$ and $\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}ZZ^\top\right)$. Denoting by λ_k the eigenvalues of $p^{-1}\boldsymbol{Z}\boldsymbol{Z}^\top$, by Corollary A.1.1 and the fact that $\gamma_0 = \|\beta\|^2/\sigma_0^2, p^{-1}\boldsymbol{Z}\boldsymbol{Z}^\top = \gamma^{-1}\left(\boldsymbol{V}_\gamma - \boldsymbol{I}_n\right)$, we have

$$\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{1+\gamma\lambda_k} \xrightarrow{a.s.} \int_{b_-(\tau)}^{b_+(\tau)}\frac{f_\tau(x)}{1+\gamma x}\mathrm{d}x + \delta_0 = h_1(\gamma,\tau),$$

$$\frac{1}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\right) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{(1+\gamma\lambda_k)^2} \xrightarrow{a.s.} \int_{b_-(\tau)}^{b_+(\tau)}\frac{f_\tau(x)}{(1+\gamma x)^2}\mathrm{d}x + \delta_0 = h_2(\gamma,\tau),$$

$$\frac{\sigma_0^2\gamma_0}{np}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}ZZ^\top\right) = \frac{\sigma_0^2\gamma_0}{n\gamma}\left(\operatorname{trace}\left(\boldsymbol{I}_n\right) - \operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right)\right) \xrightarrow{a.s.} \frac{\gamma_0\sigma_0^2}{\gamma}(1-h_1(\gamma,\tau)),$$

$$\frac{\sigma_0^2\gamma_0}{np}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}ZZ^\top\right) = \frac{\sigma_0^2\gamma_0}{n\gamma}\left(\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\right) - \operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\right)\right) \xrightarrow{a.s.} \frac{\gamma_0\sigma_0^2}{\gamma}(h_1(\gamma,\tau)-h_2(\gamma,\tau)),$$

where

$$\delta_0 = \begin{cases} 0, & \tau \le 1, \\ 1-\dfrac{1}{\tau}, & \tau > 1. \end{cases}$$

Then, there holds

$$\begin{aligned}
&\Delta_{**}(\gamma) \\
&= \frac{\sigma_0^2}{n}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-1}\left(\boldsymbol{I}_n + \frac{\gamma_0}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right) - \frac{\sigma_0^2}{\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})}\operatorname{trace}\left(\boldsymbol{V}_\gamma^{-2}\left(\boldsymbol{I}_n + \frac{\gamma_0}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right)\right) \\
&\xrightarrow{a.s.} \left(\frac{\gamma_0\sigma_0^2}{\gamma}(1-h_1(\gamma,\tau)) + \sigma_0^2 h_1(\gamma,\tau)\right) - \frac{1}{h_1(\gamma)}\left(\frac{\gamma_0\sigma_0^2}{\gamma}(h_1(\gamma,\tau)-h_2(\gamma,\tau)) + \sigma_0^2 h_2(\gamma,\tau)\right) \\
&= \sigma_0^2\left(\frac{\gamma_0}{\gamma}-1\right)\left(\frac{h_2(\gamma,\tau)-h_1^2(\gamma,\tau)}{h_1(\gamma,\tau)}\right).
\end{aligned}$$

When $\tau \le 1$,

$$\frac{h_2(\gamma,\tau)-h_1^2(\gamma,\tau)}{h_1(\gamma,\tau)} = \frac{\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{(1+\gamma x)^2}\mathrm{d}x - \left(\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{1+\gamma x}\mathrm{d}x\right)^2}{\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{1+\gamma x}\mathrm{d}x}. \tag{A.49}$$

Since on $[b_-(\tau), b_+(\tau)]$, $f_\tau(x) > 0$ for both $\tau \le 1$ and $\tau > 1$, $(1+\gamma x)^{-1}$ are strictly decreasing ($\gamma > 0$), we have, by monotone function inequalities [[23], pages 148-149],

$$\left(\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{1+\gamma x}\mathrm{d}x\right)^2 < \left(\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{(1+\gamma x)^2}\mathrm{d}x\right)\left(\int_{b_-(\tau)}^{b_+(\tau)} f_\tau(x)\mathrm{d}x\right) = \int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{(1+\gamma x)^2}\mathrm{d}x,$$

which implies

$$\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{(1+\gamma x)^2}\mathrm{d}x - \left(\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{1+\gamma x}\mathrm{d}x\right)^2 > 0.$$

Similarly, when $\tau > 1$, since

$$\int_{b_-(\tau)}^{b_+(\tau)} f_\tau(x)\mathrm{d}x = \frac{1}{\tau},$$

the inequality above becomes

$$\left(\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{1+\gamma x}\mathrm{d}x\right)^2 < \left(\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{(1+\gamma x)^2}\mathrm{d}x\right)\left(\int_{b_-(\tau)}^{b_+(\tau)} f_\tau(x)\mathrm{d}x\right) = \frac{1}{\tau}\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{(1+\gamma x)^2}\mathrm{d}x.$$

Then

$$
\begin{aligned}
&\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{(1+\gamma x)^2}\mathrm{d}x + \left(1-\frac{1}{\tau}\right) - \left(\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{1+\gamma x}\mathrm{d}x + \left(1-\frac{1}{\tau}\right)\right)^2 \\
&= \int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{(1+\gamma x)^2}\mathrm{d}x - \left(\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{1+\gamma x}\mathrm{d}x\right)^2 \\
&+ \left(1-\frac{1}{\tau}\right)\left(1 - 2\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{1+\gamma x}\mathrm{d}x - \left(1-\frac{1}{\tau}\right)\right) \\
&> \left(1-\frac{1}{\tau}\right)\left(\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{(1+\gamma x)^2}\mathrm{d}x - 2\int_{b_-(\tau)}^{b_+(\tau)} \frac{f_\tau(x)}{1+\gamma x}\mathrm{d}x + \frac{1}{\tau}\right) \\
&= \left(1-\frac{1}{\tau}\right)\left(\int_{b_-(\tau)}^{b_+(\tau)} \frac{\gamma^2 x^2}{(1+\gamma x)^2} f_\tau(x)\mathrm{d}x\right) \\
&> 0.
\end{aligned}
$$

Also, for both $\tau > 1$ and $\tau \leq 1$, the denominator $h_1(\gamma, \tau)$ is positive obviously. Thus

$$
\frac{h_2(\gamma,\tau) - h_1^2(\gamma,\tau)}{h_1(\gamma,\tau)} > 0.
$$

Then it is shown that for both $\tau \leq 1$ and $\tau > 1$, the limit of $\Delta_{**}(\gamma)$ is $c_\gamma = \sigma_0^2\left(\frac{\gamma_0}{\gamma} - 1\right) d_{\gamma,\tau}$, which is $> 0, = 0$ or < 0 depending on whether γ is $< \gamma_0, = \gamma_0$ or $> \gamma_0$.

A.2.11. Proof of Lemma 2.4.1.11. Recall that $\Delta(\gamma) = \boldsymbol{y}^\top \boldsymbol{B}_\gamma \boldsymbol{y}$ and

$$
\boldsymbol{B}_\gamma = \frac{\boldsymbol{V}_\gamma^{-1}}{n} - \frac{\boldsymbol{V}_\gamma^{-2}}{\mathrm{trace}(\boldsymbol{V}_\gamma^{-1})}.
$$

Since for $l = 1, 2, \cdots$,

$$
\frac{\mathrm{d}}{\mathrm{d}\gamma}\boldsymbol{V}_\gamma^{-l} = -l\boldsymbol{V}_\gamma^{-(l+1)}\left(\frac{1}{p}\boldsymbol{Z}\boldsymbol{Z}^\top\right) = -\frac{l}{\gamma}\left(\boldsymbol{V}_\gamma^{-l} - \boldsymbol{V}_\gamma^{-(l+1)}\right),
$$

and

$$
\frac{\mathrm{d}}{\mathrm{d}\gamma}\mathrm{trace}(\boldsymbol{V}_\gamma^{-1}) = -\frac{1}{\gamma}\mathrm{trace}\left(\boldsymbol{V}_\gamma^{-1} - \boldsymbol{V}_\gamma^{-2}\right),
$$

we can have that

$$\frac{\mathrm{d}}{\mathrm{d}\gamma}\boldsymbol{B}_\gamma = -\frac{1}{\gamma}\left(\frac{\boldsymbol{V}_\gamma^{-1} - \boldsymbol{V}_\gamma^{-2}}{n} + \frac{-2\boldsymbol{V}_\gamma^{-2} + 2\boldsymbol{V}_\gamma^{-3}}{\mathrm{trace}(\boldsymbol{V}_\gamma^{-1})} + \frac{\boldsymbol{V}_\gamma^{-2}\,\mathrm{trace}\left(\boldsymbol{V}_\gamma^{-1} - \boldsymbol{V}_\gamma^{-2}\right)}{\left(\mathrm{trace}(\boldsymbol{V}_\gamma^{-1})\right)^2}\right). \tag{A.50}$$

By a similar argument from the proof of Theorem 3.2.4, it can be checked that for every fixed γ,

$$\Delta'(\gamma) = \boldsymbol{y}^\top \frac{\mathrm{d}}{\mathrm{d}\gamma}\boldsymbol{B}_\gamma \boldsymbol{y}$$

converges in probability to $\Delta'_\infty(\gamma)$, where $\Delta'_\infty(\gamma)$ is some constant only depends on γ and τ, i.e. $\Delta'(\gamma_0) = \Delta'_\infty(\gamma_0) + o_P(1)$. More specifically, similar to the argument in Section 2.4, we have that for any $l = 1, 2, \ldots$

$$\left|\frac{1}{n}\boldsymbol{y}^\top \boldsymbol{V}_\gamma^{-l}\boldsymbol{y} - \frac{1}{n}\sigma_0^2\,\mathrm{trace}\left(\boldsymbol{V}_\gamma^{-l}\boldsymbol{V}_{\gamma_0}\right)\right| \xrightarrow{P} 0, \tag{A.51}$$

and

$$\begin{aligned}\frac{1}{n}\sigma_0^2\,\mathrm{trace}\left(\boldsymbol{V}_\gamma^{-l}\boldsymbol{V}_{\gamma_0}\right) &= \frac{1}{n}\sigma_0^2\frac{\gamma_0}{\gamma}\,\mathrm{trace}\left(\boldsymbol{V}_\gamma^{-(l-1)}\right) - \sigma_0^2\left(\frac{\gamma_0}{\gamma} - 1\right)\\ &\xrightarrow{a.s.} \sigma_0^2\frac{\gamma_0}{\gamma}h_{l-1}(\gamma,\tau) - \sigma_0^2\left(\frac{\gamma_0}{\gamma} - 1\right)h_l(\gamma,\tau).\end{aligned} \tag{A.52}$$

Combining (A.51) and (A.52), we have when $\gamma = \gamma_0$

$$\frac{1}{n}\boldsymbol{y}^\top \boldsymbol{V}_{\gamma_0}^{-l}\boldsymbol{y} \xrightarrow{P} \sigma_0^2 h_{l-1}(\gamma_0,\tau). \tag{A.53}$$

Therefore, by (A.50) and (A.53), as $n \to \infty$,

$$\boldsymbol{y}^\top \frac{\mathrm{d}}{\mathrm{d}\gamma}\boldsymbol{B}_{\gamma_0}\boldsymbol{y} \xrightarrow{P} \frac{\sigma_0^2}{\gamma_0}\frac{h_1^2(\gamma_0,\tau) - h_2(\gamma_0,\tau)}{h_1(\gamma_0,\tau)},$$

which means

$$\Delta'_\infty(\gamma_0) = \frac{\sigma_0^2}{\gamma_0}\frac{h_1^2(\gamma_0,\tau) - h_2(\gamma_0,\tau)}{h_1(\gamma_0,\tau)}. \tag{A.54}$$

By the Taylor series expansion, we have

$$\Delta(\hat{\gamma}) = \Delta(\gamma_0) + \Delta'(\gamma_0)(\hat{\gamma} - \gamma_0) + \frac{1}{2}(\hat{\gamma} - \gamma_0)^2 \Delta''(\gamma_\delta)$$

where γ_δ is a number between γ_0 and $\hat{\gamma}$. Since $\Delta(\hat{\gamma}) = 0$, we can rewrite this as

$$\sqrt{n}(\hat{\gamma} - \gamma_0) = -\frac{\sqrt{n}\Delta(\gamma_0)}{\Delta'(\gamma_0) + \frac{1}{2}(\hat{\gamma} - \gamma_0)\Delta''(\gamma_\delta)}. \tag{A.55}$$

Since we have already shown that $\Delta'(\gamma_0) = \Delta'_\infty(\gamma_0) + o_P(1)$ and $\hat{\gamma} - \gamma_0 = o_P(1)$, once we establish that

$$\Delta''(\gamma_\delta) = O_P(1), \tag{A.56}$$

it follows that

$$\Delta'(\gamma_0) + \frac{1}{2}(\hat{\gamma} - \gamma_0)^2 \Delta''(\gamma_\delta) = \Delta'_\infty(\gamma_0) + o_P(1). \tag{A.57}$$

Since $\hat{\gamma} - \gamma_0 = o_P(1)$, we can have $\gamma_\delta - \gamma_0 = o_P(1)$, then by (A.51) and (A.52)

$$\frac{1}{n}\boldsymbol{y}^\top \boldsymbol{V}_{\gamma_\delta}^{-l}\boldsymbol{y} \xrightarrow{P} \sigma_0^2 \frac{\gamma_0}{\gamma_\delta} h_{l-1}(\gamma_\delta, \tau). \tag{A.58}$$

Then by (A.58) and some algebra, we can have

$$\begin{aligned} \Delta''(\gamma_\delta) &= \boldsymbol{y}^\top \frac{\mathrm{d}^2}{\mathrm{d}\gamma^2}\boldsymbol{B}_{\gamma_\delta}\boldsymbol{y} \\ &\xrightarrow{P} 2\sigma_0^2 \frac{\gamma_0}{\gamma_\delta^3}\left(-2h_1(\gamma_\delta) + h_2(\gamma_\delta) + 2\frac{h_2(\gamma_\delta) - h_3(\gamma_\delta)}{h_1(\gamma_\delta)} + \frac{h_2^2(\gamma_\delta)}{h_1^2(\gamma_\delta)}\right). \end{aligned} \tag{A.59}$$

with $h_l(\gamma_\delta) = h_l(\gamma_\delta, \tau)$. Thus $\Delta''(\hat{\gamma}_\delta) = O_P(1)$, (A.57) is proved. Then by (A.55) we can have

$$\sqrt{n}(\hat{\gamma} - \gamma_0) = -\frac{\sqrt{n}\Delta(\gamma_0)}{\Delta'_\infty(\gamma_0)} + o_P(1), \tag{A.60}$$

where $\Delta'_\infty(\gamma_0)$ is defined in (A.54).

A.2.12. Proof of Lemma 2.4.1.13. Recall that

$$(\boldsymbol{B}_\gamma)_{ii} = \frac{1}{n}(\boldsymbol{V}_\gamma^{-1})_{ii} - \frac{\frac{1}{n}(\boldsymbol{V}_\gamma^{-2})_{ii}}{\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})},$$

and for $l = 1, 2$,

$$(\boldsymbol{V}_\gamma^{-l})_{ii} = 1 - \frac{\gamma}{p}\sum_{k=1}^{l} \tilde{\boldsymbol{z}}_i^\top \widetilde{\boldsymbol{V}}_\gamma^{-k} \tilde{\boldsymbol{z}}_i.$$

Then

$$\begin{aligned}
&\sum_{i=1}^{n}(\varepsilon_i^2 - \sigma_0^2)\left((\boldsymbol{B}_\gamma)_{ii} - \frac{1}{n}\operatorname{trace}(\boldsymbol{B}_\gamma)\right) \\
&= -\gamma \sum_{i=1}^{n} \frac{\varepsilon_i^2 - \sigma_0^2}{n}\frac{1}{p}\left(\tilde{\boldsymbol{z}}_i^\top \widetilde{\boldsymbol{V}}_\gamma^{-1}\tilde{\boldsymbol{z}}_i - \operatorname{trace}\left(\widetilde{\boldsymbol{V}}_\gamma^{-1}\frac{1}{n}\boldsymbol{Z}^\top\boldsymbol{Z}\right)\right) \\
&\quad + \frac{\gamma}{\frac{1}{n}\operatorname{trace}(\boldsymbol{V}_\gamma^{-1})}\sum_{l=1}^{2}\sum_{i=1}^{n}\frac{\varepsilon_i^2 - \sigma_0^2}{n}\frac{1}{p}\left(\tilde{\boldsymbol{z}}_i^\top \widetilde{\boldsymbol{V}}_\gamma^{-l}\tilde{\boldsymbol{z}}_i - \operatorname{trace}\left(\widetilde{\boldsymbol{V}}_\gamma^{-l}\frac{1}{n}\boldsymbol{Z}^\top\boldsymbol{Z}\right)\right).
\end{aligned} \tag{A.61}$$

Since by Lemma 2.4.0.1

$$\sum_{i=1}^{n}\left(\frac{\varepsilon_i^2 - \sigma_0^2}{n}\right)^2 \le \frac{1}{n}\max_{i\in[n]}\varepsilon_i^4 = o_P(1),$$

similar to (2.63), we can have for $l = 1, 2$

$$\sum_{i=1}^{n}\frac{\varepsilon_i^2 - \sigma_0^2}{n}\frac{1}{p}\left(\tilde{\boldsymbol{z}}_i^\top \widetilde{\boldsymbol{V}}_\gamma^{-1}\tilde{\boldsymbol{z}}_i - \operatorname{trace}\left(\widetilde{\boldsymbol{V}}_\gamma^{-1}\frac{1}{n}\boldsymbol{Z}^\top\boldsymbol{Z}\right)\right) = o_P\left(\frac{1}{\sqrt{n}}\right),$$

which implies (2.65) by (A.61).

APPENDIX B

Supporting Proofs of Chapter 3

In the beginning of this appendix, we list a useful preliminary result from [7].

LEMMA B.0.0.1 (Proposition S1. in [7]). *Suppose that $\boldsymbol{X}$ is an $n \times p$ matrix with iid entris $x_{ij} \sim \mathcal{N}(0,1)$, then $\boldsymbol{W} = \boldsymbol{X}^\top \boldsymbol{X}$ is a Wishart$(n, \boldsymbol{I}_p)$ random matrix. Let $\boldsymbol{\beta} \in \mathbb{R}^p$, then we have*

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\beta}^\top \boldsymbol{W} \boldsymbol{\beta}\right] &= n\|\boldsymbol{\beta}\|^2 \\
\mathbb{E}\left[\mathrm{tr}(\boldsymbol{W})\boldsymbol{\beta}^\top \boldsymbol{W} \boldsymbol{\beta}\right] &= (pn^2 + 2n)\|\boldsymbol{\beta}\|^2 \\
\mathbb{E}\left[\boldsymbol{\beta}^\top \boldsymbol{W}^2 \boldsymbol{\beta}\right] &= (pn + n^2 + 1)\|\boldsymbol{\beta}\|^2 \\
\mathbb{E}\left[\boldsymbol{\beta}^\top \boldsymbol{W}^3 \boldsymbol{\beta}\right] &= (p^2 n + 3pn^2 + 2pn + n^3 + 3n^2 + 4n)\|\boldsymbol{\beta}\|^2
\end{aligned}$$

B.1. Proof of Lemma 3.4.0.1

Let S_k denote the symmetric group on k elements. Then each permutation $\pi \in S_k$ can be uniquely expressed as a product of disjoint cycles $\pi = C_1 \cdots C_{m(\pi)}$, where $C_j = (c_{1j} \cdots c_{k_j j})$, $k_1 + \cdots + k_{m(\pi)} = k$, and all of the $c_{ij} \in \{1, \ldots, k\}$ are distinct.

Let $H_1, \ldots, H_k$ be $d \times d$ symmetric matrices and define the polynomial

$$r_\pi(\Sigma)(H_1, \ldots, H_k) = \prod_{j=1}^{m(\pi)} \mathrm{tr}\left(\prod_{i=1}^{k_j} \Sigma H_{c_{ij}}\right). \tag{B.1}$$

For a Wishart$(n, \boldsymbol{\Sigma})$ random matrix $\boldsymbol{W} = \boldsymbol{X}^\top \boldsymbol{X}$, Theorem 1 in [27] and Proposition 1 in [14] give the following formula:

$$\mathbb{E}\left\{\operatorname{tr}(WH_1)\cdots\operatorname{tr}(WH_k)\right\} = \sum_{\pi\in S_k} 2^{k-m(\pi)} n^{m(\pi)} r_\pi(\Sigma)(H_1,\ldots,H_k). \tag{B.2}$$

in oue case $\Sigma = \boldsymbol{I}_p$, and we can define corresponding

$$r_\pi(H_1,\ldots,H_k) = \prod_{j=1}^{m(\pi)} \operatorname{tr}\left(\prod_{i=1}^{k_j} H_{c_{ij}}\right) \tag{B.3}$$

Using this formula, we can have

$$\begin{aligned}
&\mathbb{E}\left[\boldsymbol{\alpha}^\top \boldsymbol{W}\boldsymbol{\alpha}\boldsymbol{\beta}^\top \boldsymbol{W}\boldsymbol{\beta}\right] \\
&= \mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{\alpha}\boldsymbol{\alpha}^\top\right)\cdot\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{\beta}\boldsymbol{\beta}^\top\right)\right] \\
&= 2n\operatorname{tr}\left(\boldsymbol{\alpha}\boldsymbol{\alpha}^\top\boldsymbol{\beta}\boldsymbol{\beta}^\top\right) + n^2\operatorname{tr}\left(\boldsymbol{\alpha}\boldsymbol{\alpha}^\top\right)\operatorname{tr}\left(\boldsymbol{\beta}\boldsymbol{\beta}^\top\right) \\
&= 2n(\boldsymbol{\alpha}^\top\boldsymbol{\beta}_j)^2 + n^2\|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2.
\end{aligned}$$

Now let $\boldsymbol{u}_1,\cdots,\boldsymbol{u}_p \in \mathbb{R}^p$ be an orthonormal basis of $\mathbb{R}^p$.Then define the $p\times p$ symmetric matrices

$$\boldsymbol{H}_{\alpha i} = \frac{1}{2}\left(\boldsymbol{\alpha}\boldsymbol{u}_i^\top + \boldsymbol{u}_i\boldsymbol{\alpha}^\top\right), \quad \boldsymbol{H}_{\beta i} = \frac{1}{2}\left(\boldsymbol{\beta}\boldsymbol{u}_i^\top + \boldsymbol{u}_i\boldsymbol{\beta}^\top\right),$$

and $\boldsymbol{H}_{\alpha 0} = \boldsymbol{\alpha}\boldsymbol{\alpha}^\top$. Since $\boldsymbol{\beta}^\top\boldsymbol{W}^2\boldsymbol{\beta} = \sum_{i=1}^p(\boldsymbol{\beta}^\top\boldsymbol{W}\boldsymbol{u}_i)^2$, then

$$\begin{aligned}
&\mathbb{E}\left[\boldsymbol{\alpha}^\top \boldsymbol{W}\boldsymbol{\alpha}\boldsymbol{\beta}^\top \boldsymbol{W}^2\boldsymbol{\beta}\right] \\
&= \mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{H}_{\alpha 0}\right)\sum_{i=1}^p \operatorname{tr}^2\left(\boldsymbol{W}\boldsymbol{H}_{\beta i}\right)\right] \\
&= \sum_{i=1}^p \mathbb{E}\left[tr\left(\boldsymbol{W}\boldsymbol{H}_{\alpha 0}\right)\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{H}_{\beta i}\right)\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{H}_{\beta i}\right)\right],
\end{aligned}$$

where by (B.3)

$$\begin{aligned}
&\mathbb{E}\left[tr\left(\boldsymbol{W}\boldsymbol{H}_{\alpha 0}\right)\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{H}_{\beta i}\right)\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{H}_{\beta i}\right)\right] \\
&= \sum_{\pi \in S_3} 2^{3-m(\pi)} n^{m(\pi)} r_\pi(\boldsymbol{H}_{\alpha 0}, \boldsymbol{H}_{\beta i}, \boldsymbol{H}_{\beta i}) \\
&= 8nr_{(123)} + 4n^2 r_{(12)(3)} + 2n^2 r_{(1)(23)} + n^3 r_{(1)(2)(3)} \\
&= 8n \operatorname{tr}\left(\boldsymbol{H}_{\alpha 0}\boldsymbol{H}_{\beta i}\boldsymbol{H}_{\beta i}\right) + 4n^2 \operatorname{tr}\left(\boldsymbol{H}_{\alpha 0}\boldsymbol{H}_{\beta i}\right)\operatorname{tr}\left(\boldsymbol{H}_{\beta i}\right) + 2n^2 \operatorname{tr}\left(\boldsymbol{H}_{\alpha 0}\right)\operatorname{tr}\left(\boldsymbol{H}_{\beta i}\boldsymbol{H}_{\beta i}\right) \\
&+ n^3 \operatorname{tr}\left(\boldsymbol{H}_{\alpha 0}\right)\operatorname{tr}^2\left(\boldsymbol{H}_{\beta i}\right).
\end{aligned}$$

Then by the definition of $\boldsymbol{H}_{\alpha 0}$ and $\boldsymbol{H}_{\beta i}$, we can have

$$\begin{aligned}
\sum_{i=1}^{p} 8n \operatorname{tr}\left(\boldsymbol{H}_{\alpha 0}\boldsymbol{H}_{\beta i}\boldsymbol{H}_{\beta i}\right) &= 2n(p+1)(\boldsymbol{\alpha}^\top\boldsymbol{\beta})^2 + 4n\|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2 \\
\sum_{i=1}^{p} 4n^2 \operatorname{tr}\left(\boldsymbol{H}_{\alpha 0}\boldsymbol{H}_{\beta i}\right)\operatorname{tr}\left(\boldsymbol{H}_{\beta i}\right) &= 4n^2(\boldsymbol{\alpha}^\top\boldsymbol{\beta})^2 \\
\sum_{i=1}^{p} 2n^2 \operatorname{tr}\left(\boldsymbol{H}_{\alpha 0}\right)\operatorname{tr}\left(\boldsymbol{H}_{\beta i}\boldsymbol{H}_{\beta i}\right) &= n^2(1+p)\|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2 \\
\sum_{i=1}^{p} 2n^2 n^3 \operatorname{tr}\left(\boldsymbol{H}_{\alpha 0}\right)\operatorname{tr}^2\left(\boldsymbol{H}_{\beta i}\right) &= n^3\|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2.
\end{aligned}$$

From all the equalities above, it follows that

$$\mathbb{E}\left[\boldsymbol{\alpha}^\top\boldsymbol{W}\boldsymbol{\alpha}\boldsymbol{\beta}^\top\boldsymbol{W}^2\boldsymbol{\beta}\right] = (2n + 2np + 4n^2)(\boldsymbol{\alpha}^\top\boldsymbol{\beta})^2 + (4n + n^2 + n^2 p + n^3)\|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2.$$

Similarly,

$$\mathbb{E}\left[\boldsymbol{\alpha}^\top\boldsymbol{W}^2\boldsymbol{\alpha}\boldsymbol{\beta}^\top\boldsymbol{W}^2\boldsymbol{\beta}\right] = \sum_{i=1}^{p}\sum_{j=1}^{p} \mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{H}_{\alpha i}\right)\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{H}_{\alpha i}\right)\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{H}_{\beta j}\right)\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{H}_{\beta j}\right)\right],$$

where by (B.3)

$$\mathbb{E}\left[\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{H}_{\alpha i}\right)\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{H}_{\alpha i}\right)\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{H}_{\beta j}\right)\operatorname{tr}\left(\boldsymbol{W}\boldsymbol{H}_{\beta j}\right)\right]$$
$$= 32nr_{(1234)} + 16nr_{(1324)} + 32n^2r_{(1)(234)} + 8n^2r_{(13)(24)} + 4n^2r_{(12)(34)} + 4n^3r_{(12)(3)(4)}$$
$$+ 8n^3r_{(13)(2)(4)} + n^4r_{(1)(2)(3)(4)},$$

and

$$\begin{aligned}
r_{(1234)} &= \operatorname{tr}\left(\boldsymbol{H}_{\alpha i}^2\boldsymbol{H}_{\beta j}^2\right) \\
r_{(1324)} &= \operatorname{tr}\left(\boldsymbol{H}_{\alpha i}\boldsymbol{H}_{\beta j}\boldsymbol{H}_{\alpha i}\boldsymbol{H}_{\beta j}\right) \\
r_{(1)(234)} &= \operatorname{tr}\left(\boldsymbol{H}_{\alpha i}\right)\operatorname{tr}\left(\boldsymbol{H}_{\alpha i}\boldsymbol{H}_{\beta j}\boldsymbol{H}_{\beta j}\right) \\
r_{(13)(24)} &= \operatorname{tr}^2\left(\boldsymbol{H}_{\alpha i}\boldsymbol{H}_{\beta j}\right) \\
r_{(12)(34)} &= \operatorname{tr}\left(\boldsymbol{H}_{\alpha i}^2\right)\operatorname{tr}\left(\boldsymbol{H}_{\beta j}^2\right) \\
r_{(12)(3)(4)} &= \operatorname{tr}\left(\boldsymbol{H}_{\alpha i}^2\right)\operatorname{tr}^2\left(\boldsymbol{H}_{\beta j}\right) \\
r_{(13)(2)(4)} &= \operatorname{tr}\left(\boldsymbol{H}_{\alpha i}\boldsymbol{H}_{\beta j}\right)\operatorname{tr}\left(\boldsymbol{H}_{\alpha i}\right)\operatorname{tr}\left(\boldsymbol{H}_{\beta j}\right) \\
r_{(1)(2)(3)(4)} &= \operatorname{tr}^2\left(\boldsymbol{H}_{\alpha i}\right)\operatorname{tr}^2\left(\boldsymbol{H}_{\beta j}\right).
\end{aligned}$$

Then by the definition of $\boldsymbol{H}_{\alpha 0}$ and $\boldsymbol{H}_{\beta i}$, we can have

$$\sum_{i=1}^{p} \operatorname{tr}\left(\boldsymbol{H}_{\alpha i}^2 \boldsymbol{H}_{\beta j}^2\right) = \frac{1}{16}\left((4+3p+p^2)(\boldsymbol{\alpha}^\top \boldsymbol{\beta})^2 + (4+4p)\|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2\right)$$

$$\sum_{i=1}^{p} \operatorname{tr}\left(\boldsymbol{H}_{\alpha i}^2 \boldsymbol{H}_{\beta j}^2\right) = \frac{1}{8}\left((6+p)(\boldsymbol{\alpha}^\top \boldsymbol{\beta})^2 + \|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2\right)$$

$$\sum_{i=1}^{p} \operatorname{tr}\left(\boldsymbol{H}_{\alpha i}\right) \operatorname{tr}\left(\boldsymbol{H}_{\alpha i}\boldsymbol{H}_{\beta j}\boldsymbol{H}_{\beta j}\right) = \frac{1}{4}\left(p(\boldsymbol{\alpha}^\top \boldsymbol{\beta})^2 + 3\|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2\right)$$

$$\sum_{i=1}^{p} \operatorname{tr}^2\left(\boldsymbol{H}_{\alpha i}\boldsymbol{H}_{\beta j}\right) = \frac{1}{4}\left((p+2)(\boldsymbol{\alpha}^\top \boldsymbol{\beta})^2 + \|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2\right)$$

$$\sum_{i=1}^{p} \operatorname{tr}\left(\boldsymbol{H}_{\alpha i}^2\right) \operatorname{tr}\left(\boldsymbol{H}_{\beta j}^2\right) = \frac{1}{4}(1+2p+p^2)\|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2$$

$$\sum_{i=1}^{p} \operatorname{tr}\left(\boldsymbol{H}_{\alpha i}^2\right) \operatorname{tr}^2\left(\boldsymbol{H}_{\beta j}\right) = \frac{1}{2}(1+p)\|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2$$

$$\sum_{i=1}^{p} \operatorname{tr}\left(\boldsymbol{H}_{\alpha i}\boldsymbol{H}_{\beta j}\right) \operatorname{tr}\left(\boldsymbol{H}_{\alpha i}\right) \operatorname{tr}\left(\boldsymbol{H}_{\beta j}\right) = (\boldsymbol{\alpha}^\top \boldsymbol{\beta})^2$$

$$\sum_{i=1}^{p} \operatorname{tr}^2\left(\boldsymbol{H}_{\alpha i}\right) \operatorname{tr}^2\left(\boldsymbol{H}_{\beta j}\right) = \|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\left[\boldsymbol{\alpha}^\top \boldsymbol{W}^2 \boldsymbol{\alpha} \boldsymbol{\beta}^\top \boldsymbol{W}^2 \boldsymbol{\beta}\right] \\
&= (2np^2 + 10n^2p + 8n^3 + 8np + 4n^2 + 20n)(\boldsymbol{\alpha}^\top \boldsymbol{\beta})^2 \\
&\ + (n^2p^2 + n^4 + 2n^3p + 2n^2p + 2n^3 + 27n^2 + 8np + 10n)\|\boldsymbol{\alpha}\|^2\|\boldsymbol{\beta}\|^2.
\end{aligned}$$

B.2. Proof of Lemma 3.4.0.2

Recall that

$$\boldsymbol{Y}^\top \boldsymbol{Y} = \boldsymbol{E}^\top \boldsymbol{E} + \boldsymbol{B}^\top \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{B} + \boldsymbol{E}^\top \boldsymbol{X} \boldsymbol{B} + \boldsymbol{B}^\top \boldsymbol{X}^\top \boldsymbol{E}$$

and

$$\boldsymbol{Y}^\top \boldsymbol{X} \boldsymbol{X}^\top \boldsymbol{Y} = \boldsymbol{E}^\top \boldsymbol{X} \boldsymbol{X}^\top \boldsymbol{E} + \boldsymbol{B}^\top \left(\boldsymbol{X}^\top \boldsymbol{X}\right)^2 \boldsymbol{B} + \boldsymbol{E}^\top \boldsymbol{X} \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{B} + \boldsymbol{B}^\top \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{X}^\top \boldsymbol{E}.$$

Then,

$$\mathbb{E}\left[\frac{1}{n}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{Y})\right] = \|\boldsymbol{B}\|_F^2 + \operatorname{tr}(\boldsymbol{\Sigma}_e),$$
$$\text{and}\quad \mathbb{E}\left[\frac{1}{n^2}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y})\right] = \frac{p+n+1}{n}\|\boldsymbol{B}\|_F^2 + \frac{p}{n}\operatorname{tr}(\boldsymbol{\Sigma}_e).$$

Also, we can have

$$\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{Y}) = \sum_{i=1}^q \boldsymbol{\beta}_i^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}_i + 2\sum_{i=1}^q \boldsymbol{\beta}_i^\top\boldsymbol{X}^\top\tilde{\boldsymbol{e}}_i + \sum_{i=1}^q \tilde{\boldsymbol{e}}_i^\top\tilde{\boldsymbol{e}}_i,$$
$$\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y}) = \sum_{i=1}^q \boldsymbol{\beta}_i^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}_i + 2\sum_{i=1}^q \boldsymbol{\beta}_i^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{X}^\top\tilde{\boldsymbol{e}}_i + \sum_{i=1}^q \tilde{\boldsymbol{e}}_i^\top\boldsymbol{X}\boldsymbol{X}^\top\tilde{\boldsymbol{e}}_i.$$

B.2.1. Proof of (3.19). We have

$$\begin{aligned}
&\operatorname{tr}^2(\boldsymbol{Y}^\top\boldsymbol{Y})\\
&= \left(\sum_i \boldsymbol{\beta}_i^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}_i\right)\left(\sum_j \boldsymbol{\beta}_j^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}_j\right) + 2\left(\sum_i \boldsymbol{\beta}_i^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}_i\right)\left(\sum_j \tilde{\boldsymbol{e}}_j^\top\tilde{\boldsymbol{e}}_j\right)\\
&+ 4\left(\sum_i \boldsymbol{\beta}_i^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}_i\right)\left(\sum_j \boldsymbol{\beta}_j^\top\boldsymbol{X}^\top\tilde{\boldsymbol{e}}_j\right) + 4\left(\sum_i \boldsymbol{\beta}_i^\top\boldsymbol{X}^\top\tilde{\boldsymbol{e}}_i\right)\left(\sum_j \tilde{\boldsymbol{e}}_j^\top\tilde{\boldsymbol{e}}_j\right)\\
&+ 4\left(\sum_i \boldsymbol{\beta}_i^\top\boldsymbol{X}^\top\tilde{\boldsymbol{e}}_i\right)\left(\sum_j \boldsymbol{\beta}_j^\top\boldsymbol{X}^\top\tilde{\boldsymbol{e}}_j\right) + \left(\sum_i \tilde{\boldsymbol{e}}_i^\top\tilde{\boldsymbol{e}}_i\right)\left(\sum_j \tilde{\boldsymbol{e}}_j^\top\tilde{\boldsymbol{e}}_j\right),
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{tr}^2(\boldsymbol{Y}^\top\boldsymbol{Y})\right]\\
&= \mathbb{E}\left[\sum_{i,j} \boldsymbol{\beta}_i^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}_i \cdot \boldsymbol{\beta}_j^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}_j\right] + 2\mathbb{E}\left[\sum_i \boldsymbol{\beta}_i^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta}_i\right]\mathbb{E}\left[\sum_j \tilde{\boldsymbol{e}}_j^\top\tilde{\boldsymbol{e}}_j\right]\\
&+ 4\mathbb{E}\left[\sum_i (\boldsymbol{\beta}_i^\top\boldsymbol{X}^\top\tilde{\boldsymbol{e}}_i)^2\right] + \mathbb{E}\left[\left(\sum_i \tilde{\boldsymbol{e}}_i^\top\tilde{\boldsymbol{e}}_i\right)^2\right].
\end{aligned}$$

Since $\text{vec}(\boldsymbol{E}) \sim \mathcal{N}(0, \boldsymbol{\Sigma}_e \otimes \boldsymbol{I}_n)$,

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_i \tilde{\boldsymbol{e}}_i^\top \tilde{\boldsymbol{e}}_i\right)^2\right] \\
&= \mathbb{E}\left[\left(\text{vec}^\top(\boldsymbol{E})\text{vec}(\boldsymbol{E})\right)^2\right] \\
&= \text{Var}\left(\text{vec}^\top(\boldsymbol{E})\text{vec}(\boldsymbol{E})\right) + \mathbb{E}^2\left[\text{vec}^\top(\boldsymbol{E})\text{vec}(\boldsymbol{E})\right] \\
&= 2n\,\text{tr}(\boldsymbol{\Sigma}_e^2) + n^2\,\text{tr}^2(\boldsymbol{\Sigma}_e).
\end{aligned}$$

Let $\boldsymbol{W} = \boldsymbol{X}^\top \boldsymbol{X}$, since $\mathbb{E}\left[\boldsymbol{\beta}^\top \boldsymbol{W} \boldsymbol{\beta}\right] = n\|\boldsymbol{\beta}\|^2$ by Lemma B.0.0.1,

$$\begin{aligned}
&\mathbb{E}\left[\sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{\beta}_i\right] \mathbb{E}\left[\sum_j \tilde{\boldsymbol{e}}_j^\top \tilde{\boldsymbol{e}}_j\right] \\
&= \sum_{i=1}^q \mathbb{E}\left[\boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{\beta}_i\right] \mathbb{E}\left[\text{vec}^\top(\boldsymbol{E})\text{vec}(\boldsymbol{E})\right] \\
&= n \sum_{i=1}^q \|\boldsymbol{\beta}_i\|^2 n\,\text{tr}(\boldsymbol{\Sigma}_e) = n^2 \|\boldsymbol{B}\|_F^2\,\text{tr}(\boldsymbol{\Sigma}_e).
\end{aligned}$$

Then

$$\mathbb{E}\left[\sum_i (\boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i)^2\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_i (\boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i)^2 | \boldsymbol{X}\right]\right],$$

where

$$
\begin{aligned}
&\mathbb{E}\left[\sum_i (\boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i)^2 \mid \boldsymbol{X}\right] \\
&= \mathbb{E}\left[\sum_i \tilde{\boldsymbol{e}}_i^\top \boldsymbol{X}\boldsymbol{\beta}_i\boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i \mid \boldsymbol{X}\right] \\
&= \mathbb{E}\left[\operatorname{vec}(\boldsymbol{E})^\top \begin{bmatrix} \boldsymbol{X}\boldsymbol{\beta}_1\boldsymbol{\beta}_1^\top \boldsymbol{X}^\top & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{X}\boldsymbol{\beta}_q\boldsymbol{\beta}_q^\top \boldsymbol{X}^\top \end{bmatrix} \operatorname{vec}(\boldsymbol{E})\right] \\
&= \operatorname{tr}\left(\begin{bmatrix} \boldsymbol{X}\boldsymbol{\beta}_1\boldsymbol{\beta}_1^\top \boldsymbol{X}^\top & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{X}\boldsymbol{\beta}_q\boldsymbol{\beta}_q^\top \boldsymbol{X}^\top \end{bmatrix} \cdot (\boldsymbol{\Sigma}_e \otimes \boldsymbol{I}_n)\right) \\
&= \sum_{i=1}^q \sigma_{ii}^2 \operatorname{tr}\left(\boldsymbol{X}\boldsymbol{\beta}_i\boldsymbol{\beta}_i^\top \boldsymbol{X}^\top\right) = \sum_{i=1}^q \sigma_{ii}^2 \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i.
\end{aligned}
$$

It follows that

$$
\mathbb{E}\left[\sum_i (\boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i)^2\right] = \sum_{i=1}^q \sigma_{ii}^2 \mathbb{E}\left[\boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i\right] = n\sum_{i=1}^q \sigma_{ii}^2 \|\boldsymbol{\beta}_i\|^2.
$$

By Lemma 3.4.0.1, we can have

$$
\mathbb{E}\left[\boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i \cdot \boldsymbol{\beta}_j^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_j\right] = 2n(\boldsymbol{\beta}_i^\top \boldsymbol{\beta}_j)^2 + n^2\|\boldsymbol{\beta}_i\|^2\|\boldsymbol{\beta}_j\|^2.
$$

Therefore,

$$
\begin{aligned}
\mathbb{E}\left[\sum_{i,j} \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i \cdot \boldsymbol{\beta}_j^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_j\right] &= \sum_{i,j}\left(2n(\boldsymbol{\beta}_i^\top \boldsymbol{\beta}_j)^2 + n^2\|\boldsymbol{\beta}_i\|^2\|\boldsymbol{\beta}_j\|^2\right) \\
&= n^2\|\boldsymbol{B}\|_F^4 + 2n\|\boldsymbol{B}^\top \boldsymbol{B}\|_F^2.
\end{aligned}
$$

From all the equalities above, we can have

$$\mathbb{E}\left[\operatorname{tr}^2(\boldsymbol{Y}^\top\boldsymbol{Y})\right] = n^2\|\boldsymbol{B}\|_F^4 + 2n\|\boldsymbol{B}^\top\boldsymbol{B}\|_F^2 + 2n^2\|\boldsymbol{B}\|_F^2\operatorname{tr}(\boldsymbol{\Sigma}_e) + 4n\sum_{i=1}^q \sigma_{ii}^2\|\boldsymbol{\beta}_i\|^2 + 2n\operatorname{tr}(\boldsymbol{\Sigma}_e^2) + n^2\operatorname{tr}^2(\boldsymbol{\Sigma}_e).$$

Therefore,

$$\begin{aligned}
&\operatorname{Var}\left(\frac{1}{n}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{Y})\right)\\
&= \frac{1}{n^2}\mathbb{E}\left[\operatorname{tr}^2(\boldsymbol{Y}^\top\boldsymbol{Y})\right] - \mathbb{E}^2\left[\frac{1}{n}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{Y})\right]\\
&= \|\boldsymbol{B}\|_F^4 + \frac{2}{n}\|\boldsymbol{B}^\top\boldsymbol{B}\|_F^2 + 2\|\boldsymbol{B}\|_F^2\operatorname{tr}(\boldsymbol{\Sigma}_e) + \frac{4}{n}\sum_{i=1}^q \sigma_{ii}^2\|\boldsymbol{\beta}_i\|^2 + \operatorname{tr}^2(\boldsymbol{\Sigma}_e) + \frac{2}{n}\operatorname{tr}(\boldsymbol{\Sigma}_e^2)\\
&\quad - (\|\boldsymbol{B}\|_F^2 + \operatorname{tr}(\boldsymbol{\Sigma}_e))^2\\
&= \frac{2}{n}\|\boldsymbol{B}^\top\boldsymbol{B}\|_F^2 + \frac{4}{n}\sum_{i=1}^q \sigma_{ii}^2\|\boldsymbol{\beta}_i\|^2 + \frac{2}{n}\operatorname{tr}(\boldsymbol{\Sigma}_e^2).
\end{aligned}$$

B.2.2. Proof of (3.20). Now let's focus on

$$\begin{aligned}
&\operatorname{Cov}\left(\frac{1}{n}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{Y}), \frac{1}{n^2}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y})\right)\\
&= \frac{1}{n^3}\mathbb{E}\left[\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{Y})\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y})\right] - \mathbb{E}\left[\frac{1}{n}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{Y})\right]\mathbb{E}\left[\frac{1}{n^2}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y})\right].
\end{aligned}$$

Since

$$
\begin{aligned}
&\operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{Y}) \operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y}) \\
&= \left(\sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i\right)\left(\sum_j \boldsymbol{\beta}_j^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_j\right) \\
&+ \left(\sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i\right)\left(\sum_j \tilde{\boldsymbol{e}}_j^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_j\right) + \left(\sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i\right)\left(\sum_j \tilde{\boldsymbol{e}}_j^\top \tilde{\boldsymbol{e}}_j\right) \\
&+ 2\left(\sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i\right)\left(\sum_j \boldsymbol{\beta}_j^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_j\right) + 2\left(\sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i\right)\left(\sum_j \boldsymbol{\beta}_j^\top \boldsymbol{X}^\top \tilde{\boldsymbol{e}}_j\right) \\
&+ 2\left(\sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i\right)\left(\sum_j \tilde{\boldsymbol{e}}_j^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_j\right) + 2\left(\sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i\right)\left(\sum_j \tilde{\boldsymbol{e}}_j^\top \tilde{\boldsymbol{e}}_j\right) \\
&+ 4\left(\sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i\right)\left(\sum_j \boldsymbol{\beta}_j^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_j\right) \\
&+ \left(\sum_i \tilde{\boldsymbol{e}}_i^\top \tilde{\boldsymbol{e}}_i\right)\left(\sum_j \tilde{\boldsymbol{e}}_j^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_j\right),
\end{aligned}
$$

we can have

$$
\begin{aligned}
&\mathbb{E}\left[\operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{Y}) \operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y})\right] \\
&= \mathbb{E}\left[\sum_{i,j} \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i \boldsymbol{\beta}_j^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_j\right] + \mathbb{E}\left[\sum_{i,j} \tilde{\boldsymbol{e}}_i^\top \tilde{\boldsymbol{e}}_i \tilde{\boldsymbol{e}}_j^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_j\right] \\
&+ \mathbb{E}\left[\sum_{i,j} \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i \tilde{\boldsymbol{e}}_j^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_j\right] + \mathbb{E}\left[\sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i\right] \mathbb{E}\left[\sum_j \tilde{\boldsymbol{e}}_j^\top \tilde{\boldsymbol{e}}_j\right] \\
&+ 4\mathbb{E}\left[\sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i\right].
\end{aligned}
$$

Since

$$
\mathrm{E}[\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x}\boldsymbol{x}^\top \boldsymbol{B}\boldsymbol{x}] = 2\operatorname{tr}(\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\Sigma}) + 4\boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{B}\boldsymbol{\mu} + (\operatorname{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^\top \boldsymbol{A}\boldsymbol{\mu})(\operatorname{tr}(\boldsymbol{B}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^\top \boldsymbol{B}\boldsymbol{\mu}),
$$

and $\operatorname{vec}(\boldsymbol{E}) \sim \mathcal{N}(0, \boldsymbol{\Sigma}_e \otimes \boldsymbol{I}_n)$, then we can have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i,j} \tilde{\boldsymbol{e}}_i^\top \tilde{\boldsymbol{e}}_i \tilde{\boldsymbol{e}}_j^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_j | \boldsymbol{X}\right] \\
&= \mathbb{E}\left[\operatorname{vec}^\top(\boldsymbol{E}) \operatorname{vec}(\boldsymbol{E}) \operatorname{vec}^\top(\boldsymbol{E})(\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top) \operatorname{vec}(\boldsymbol{E}) | \boldsymbol{X}\right] \\
&= 2 \operatorname{tr}\left((\boldsymbol{\Sigma}_e \otimes \boldsymbol{I}_n)(\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top)(\boldsymbol{\Sigma}_e \otimes \boldsymbol{I}_n)\right) + \operatorname{tr}(\boldsymbol{\Sigma}_e \otimes \boldsymbol{I}_n) \operatorname{tr}\left((\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top)(\boldsymbol{\Sigma}_e \otimes \boldsymbol{I}_n)\right) \\
&= 2 \operatorname{tr}\left(\boldsymbol{\Sigma}_e^2 \otimes \boldsymbol{X}\boldsymbol{X}^\top\right) + n \operatorname{tr}(\boldsymbol{\Sigma}_e) \operatorname{tr}\left(\boldsymbol{\Sigma}_e \otimes \boldsymbol{X}\boldsymbol{X}^\top\right) \\
&= 2 \operatorname{tr}\left(\boldsymbol{\Sigma}_e^2\right) \operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top\right) + n \operatorname{tr}^2(\boldsymbol{\Sigma}_e) \operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\left[\sum_{i,j} \tilde{\boldsymbol{e}}_i^\top \tilde{\boldsymbol{e}}_i \tilde{\boldsymbol{e}}_j^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_j\right] &= \mathbb{E}\left[2 \operatorname{tr}\left(\boldsymbol{\Sigma}_e^2\right) \operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top\right) + n \operatorname{tr}^2(\boldsymbol{\Sigma}_e) \operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top\right)\right] \\
&= 2np \operatorname{tr}\left(\boldsymbol{\Sigma}_e^2\right) + pn^2 \operatorname{tr}^2(\boldsymbol{\Sigma}_e).
\end{aligned}$$

Then

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i,j} \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{\beta}_i \tilde{\boldsymbol{e}}_j^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_j | \boldsymbol{X}\right] \\
&= \sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{\beta}_i \mathbb{E}\left[\sum_j \tilde{\boldsymbol{e}}_j^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_j | \boldsymbol{X}\right] \\
&= \sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{\beta}_i \mathbb{E}\left[\operatorname{vec}^\top(\boldsymbol{E})(\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top) \operatorname{vec}(\boldsymbol{E}) | \boldsymbol{X}\right] \\
&= \sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{\beta}_i \operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top\right) \operatorname{tr}(\boldsymbol{\Sigma}_e).
\end{aligned}$$

By Lemma B.0.0.1 we have $\mathbb{E}\left[\operatorname{tr}(\boldsymbol{W})\boldsymbol{\beta}^\top \boldsymbol{W}\boldsymbol{\beta}\right] = (pn^2+2n)\|\boldsymbol{\beta}\|^2$, then

$$\begin{aligned}
&\mathbb{E}\left[\sum_{i,j} \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i \tilde{\boldsymbol{e}}_j^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_j\right] \\
&= \mathbb{E}\left[\sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i \operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top\right)\right] \operatorname{tr}\left(\boldsymbol{\Sigma}_e\right) \\
&= (pn^2+2n)\operatorname{tr}\left(\boldsymbol{\Sigma}_e\right) \sum_i \|\boldsymbol{\beta}\|_i \\
&= (pn^2+2n)\operatorname{tr}\left(\boldsymbol{\Sigma}_e\right) \|\boldsymbol{B}\|_F^2
\end{aligned}$$

By Lemma B.0.0.1 $\mathbb{E}\left[\boldsymbol{\beta}_i^\top W^2 \boldsymbol{\beta}_i\right] = (pn+n^2+n)\|\boldsymbol{\beta}_i\|^2$, then

$$\mathbb{E}\left[\sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i\right] \mathbb{E}\left[\sum_j \tilde{\boldsymbol{e}}_j^\top \tilde{\boldsymbol{e}}_j\right] = n^2(p+n+1)\|\boldsymbol{B}\|_F^2 \operatorname{tr}\left(\boldsymbol{\Sigma}_e\right).$$

Since

$$\begin{aligned}
&\mathbb{E}\left[\sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \ \tilde{\boldsymbol{e}}_i \mid \boldsymbol{X}\right] \\
&= \mathbb{E}\left[\sum_i \tilde{\boldsymbol{e}}_i^\top \boldsymbol{X}\boldsymbol{\beta}_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i \mid \boldsymbol{X}\right] \\
&= \mathbb{E}\left[\operatorname{vec}(\boldsymbol{E})^\top \begin{bmatrix} \boldsymbol{X}\boldsymbol{\beta}_1\boldsymbol{\beta}_1^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{X}\boldsymbol{\beta}_q\boldsymbol{\beta}_q^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}^\top \end{bmatrix} \operatorname{vec}(\boldsymbol{E})\right] \\
&= \operatorname{tr}\left(\begin{bmatrix} \boldsymbol{X}\boldsymbol{\beta}_1\boldsymbol{\beta}_1^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{X}\boldsymbol{\beta}_q\boldsymbol{\beta}_q^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \end{bmatrix} \cdot \left(\boldsymbol{\Sigma}_e \otimes \boldsymbol{I}_n\right)\right) \\
&= \sum_{i=1}^q \sigma_{ii}^2 \operatorname{tr}\left(\boldsymbol{X}\boldsymbol{\beta}_i\boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top\right) = \sum_{i=1}^q \sigma_{ii}^2 \boldsymbol{\beta}_i^\top (\boldsymbol{X}^\top \boldsymbol{X})^2 \boldsymbol{\beta}_i
\end{aligned}$$

by $\mathbb{E}\left[\boldsymbol{\beta}_i^\top W^2 \boldsymbol{\beta}_i\right] = (pn + n^2 + n)\|\boldsymbol{\beta}_i\|^2$, it follows that

$$\mathbb{E}\left[\sum_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top\ \tilde{\boldsymbol{e}}_i\right] = (pn + n^2 + n)\sum_{i=1}^q \sigma_{ii}^2 \|\boldsymbol{\beta}_i\|^2.$$

By Proposition (3.4.0.1), we can have

$$\mathbb{E}\left[\sum_{i,j} \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i \boldsymbol{\beta}_j^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_j\right]$$
$$= (2n + 2np + 4n^2)\|\boldsymbol{B}^\top \boldsymbol{B}\|_F^2 + (4n + n^2 + n^2 p + n^3)\|\boldsymbol{B}\|_F^4.$$

From all the equalities above,

$$\begin{aligned}
&\operatorname{Cov}\left(\frac{1}{n}\operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{Y}), \frac{1}{n^2}\operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y})\right) \\
&= \frac{1}{n^3}\mathbb{E}\left[\operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{Y})\operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y})\right] - \mathbb{E}\left[\frac{1}{n}\operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{Y})\right]\mathbb{E}\left[\frac{1}{n^2}\operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y})\right] \\
&= \frac{2}{n}\left\{\left(\frac{1}{n} + \frac{p}{n} + 2\right)\|\boldsymbol{B}^\top \boldsymbol{B}\|_F^2 + \frac{2}{n}\|\boldsymbol{B}\|_F^4 + \frac{p}{n}(\operatorname{tr}(\boldsymbol{\Sigma}_e))^2\right. \\
&\left. + \left(\frac{2}{n} + 2 + \frac{2p}{n}\right)\sum_{i=1}^p \sigma_{ii}^2\|\boldsymbol{\beta}_i\|^2 + \frac{1}{n}\|\boldsymbol{B}\|_F^2\operatorname{tr}(\boldsymbol{\Sigma}_e)\right\}.
\end{aligned}$$

B.2.3. Proof of (3.21). Finally, we calculate $\operatorname{Var}\left(\operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y})\right)$. Similar to $\operatorname{Var}\left(\operatorname{tr}(\boldsymbol{Y}^\top \boldsymbol{Y})\right)$, we can have

$$\begin{aligned}
&\mathbb{E}\left[\operatorname{tr}^2(\boldsymbol{Y}^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{Y})\right] \\
&= \mathbb{E}\left[\sum_{i,j} \boldsymbol{\beta}_i^\top (\boldsymbol{X}^\top \boldsymbol{X})^2 \boldsymbol{\beta}_i \cdot \boldsymbol{\beta}_j^\top (\boldsymbol{X}^\top \boldsymbol{X})^2 \boldsymbol{\beta}_j\right] + 2\mathbb{E}\left[\sum_{i,j} \boldsymbol{\beta}_i^\top (\boldsymbol{X}^\top \boldsymbol{X})^2 \boldsymbol{\beta}_i \tilde{\boldsymbol{e}}_j^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_j\right] \\
&+ 4\mathbb{E}\left[\sum_i (\boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i)^2\right] + \mathbb{E}\left[\left(\sum_i \tilde{\boldsymbol{e}}_i^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i\right)^2\right].
\end{aligned}$$

Since

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_i \tilde{\boldsymbol{e}}_i^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i\right)^2 |\boldsymbol{X}\right] \\
&= \mathbb{E}\left[\left(\text{vec}^\top(\boldsymbol{E})(\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top)\text{vec}(\boldsymbol{E})\right)^2\right] \\
&= \text{Var}\left(\text{vec}^\top(\boldsymbol{E})(\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top)\text{vec}(\boldsymbol{E})\right) + \mathbb{E}^2\left[\text{vec}^\top(\boldsymbol{E})(\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top)\text{vec}(\boldsymbol{E})\right] \\
&= 2\,\text{tr}\left(\boldsymbol{\Sigma}_e^2 \otimes (\boldsymbol{X}\boldsymbol{X}^\top)^2\right) + \text{tr}^2\left(\boldsymbol{\Sigma}_e \otimes \boldsymbol{X}\boldsymbol{X}^\top\right) \\
&= 2\,\text{tr}\left(\boldsymbol{\Sigma}_e^2\right)\text{tr}\left((\boldsymbol{X}\boldsymbol{X}^\top)^2\right) + \text{tr}^2\left(\boldsymbol{\Sigma}_e\right)\text{tr}^2\left(\boldsymbol{X}\boldsymbol{X}^\top\right),
\end{aligned}$$

by Lemma B.0.0.1, we can have

$$\mathbb{E}\left[\left(\sum_i \tilde{\boldsymbol{e}}_i^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i\right)^2\right] = 2p(np + n^2 + n)\,\text{tr}\left(\boldsymbol{\Sigma}_e^2\right) + (p^2n^2 + 2pn)\,\text{tr}^2\left(\boldsymbol{\Sigma}_e\right).$$

Then

$$\begin{aligned}
\mathbb{E}\left[\sum_{i,j} \boldsymbol{\beta}_i^\top (\boldsymbol{X}^\top\boldsymbol{X})^2 \boldsymbol{\beta}_i \tilde{\boldsymbol{e}}_j^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_j |\boldsymbol{X}\right] &= \sum_{i=1}^q \boldsymbol{\beta}_i^\top (\boldsymbol{X}^\top\boldsymbol{X})^2 \boldsymbol{\beta}_i \mathbb{E}\left[\text{vec}^\top(\boldsymbol{E})(\boldsymbol{I}_q \otimes \boldsymbol{X}\boldsymbol{X}^\top)\text{vec}(\boldsymbol{E})|\boldsymbol{X}\right] \\
&= \sum_{i=1}^q \boldsymbol{\beta}_i^\top (\boldsymbol{X}^\top\boldsymbol{X})^2 \boldsymbol{\beta}_i\,\text{tr}(\boldsymbol{X}\boldsymbol{X}^\top)\,\text{tr}(\boldsymbol{\Sigma}_e).
\end{aligned}$$

By Lemma B.0.0.1,

$$\mathbb{E}\left[\sum_{i=1}^q \boldsymbol{\beta}_i^\top (\boldsymbol{X}^\top\boldsymbol{X})^2 \boldsymbol{\beta}_i\,\text{tr}(\boldsymbol{X}\boldsymbol{X}^\top)\,\text{tr}(\boldsymbol{\Sigma}_e)\right] = \left(p^2n^2 + pn(n^2 + n + 4) + 4n(n+1)\right)\text{tr}(\boldsymbol{\Sigma}_e)\|\boldsymbol{B}\|_F^2.$$

Since

$$\mathbb{E}\left[\sum_i (\boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i)^2\right] = \mathbb{E}\left[\mathbb{E}\left[\sum_i (\boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i)^2 |\boldsymbol{X}\right]\right]$$

where

$$
\begin{aligned}
&\mathbb{E}\left[\sum_i (\boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i)^2 \mid \boldsymbol{X}\right] \\
&= \mathbb{E}\left[\sum_i \tilde{\boldsymbol{e}}_i^\top \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i \mid \boldsymbol{X}\right] \\
&= \mathbb{E}\left[\operatorname{vec}(\boldsymbol{E})^\top \begin{bmatrix} \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_1 \boldsymbol{\beta}_1^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_q \boldsymbol{\beta}_q^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \end{bmatrix} \operatorname{vec}(\boldsymbol{E})\right] \\
&= \operatorname{tr}\left(\begin{bmatrix} \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_1 \boldsymbol{\beta}_1^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_q \boldsymbol{\beta}_q^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \end{bmatrix} \cdot (\boldsymbol{\Sigma}_e \otimes \boldsymbol{I}_n)\right) \\
&= \sum_{i=1}^q \sigma_{ii}^2 \operatorname{tr}\left(\boldsymbol{X}\boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{\beta}_i \boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top\right) = \sum_{i=1}^q \sigma_{ii}^2 \boldsymbol{\beta}_i^\top (\boldsymbol{X}^\top \boldsymbol{X})^3 \boldsymbol{\beta}_i
\end{aligned}
$$

by Lemma B.0.0.1,

$$
\begin{aligned}
&\mathbb{E}\left[\sum_i (\boldsymbol{\beta}_i^\top \boldsymbol{X}^\top \boldsymbol{X}\boldsymbol{X}^\top \tilde{\boldsymbol{e}}_i)^2\right] \\
&= \sum_{i=1}^q \sigma_{ii}^2 \mathbb{E}\left[\boldsymbol{\beta}_i^\top (\boldsymbol{X}^\top \boldsymbol{X})^3 \boldsymbol{\beta}_i\right] \\
&= \sum_{i=1}^q \sigma_{ii}^2 (p^2 n + 3pn^2 + 2pn + n^3 + 3n^2 + 4n)\|\boldsymbol{\beta}_i\|^2.
\end{aligned}
$$

By Lemma 3.4.0.1,

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i,j} \boldsymbol{\beta}_i^\top (\boldsymbol{X}^\top \boldsymbol{X})^2 \boldsymbol{\beta}_i \cdot \boldsymbol{\beta}_j^\top (\boldsymbol{X}^\top \boldsymbol{X})^2 \boldsymbol{\beta}_j\right] \\
&= (2np^2 + 10n^2 p + 8n^3 + 8np + 4n^2 + 20n)\|\boldsymbol{B}^\top \boldsymbol{B}\|_F^2 \\
&\quad + (n^2 p^2 + n^4 + 2n^3 p + 2n^2 p + 2n^3 + 27n^2 + 8np + 10n)\|\boldsymbol{B}\|_F^4.
\end{aligned}
$$

From all the equalities above, we can have

$$\begin{aligned}
&\operatorname{Var}\left(\frac{1}{n^2}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y})\right) \\
&= \frac{1}{n^4}\mathbb{E}\left[\operatorname{tr}^2(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y})\right] - \mathbb{E}^2\left[\frac{1}{n}\operatorname{tr}(\boldsymbol{Y}^\top\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{Y})\right] \\
&= \frac{2}{n}\left(\left(\frac{p}{n}\right)^2 + \frac{p}{n} + \frac{p}{n^2}\right)\operatorname{tr}(\boldsymbol{\Sigma}_e^2) + \frac{2}{n}\frac{p}{n^2}\operatorname{tr}^2(\boldsymbol{\Sigma}_e) \\
&\quad + \frac{2}{n}\left(4 + \frac{2}{n} + \frac{5p}{n} + \frac{p^2}{n^2} + \frac{4p}{n^2} + \frac{10}{n^2}\right)\|\boldsymbol{B}^\top\boldsymbol{B}\|_F^2 + \frac{2}{n}\left(\frac{13}{n} + \frac{4p}{n^2} + \frac{5}{n^2}\right)\|\boldsymbol{B}\|_F^4 \\
&\quad + \frac{2}{n}\left(2\frac{p^2}{n^2} + \frac{6p}{n} + \frac{6p}{n^2} + 2 + \frac{6}{n} + \frac{8}{n^2}\right)\sum_{i=1}^{p}\sigma_{ii}^2\|\boldsymbol{\beta}_i\|^2 + \frac{2}{n}\left(\frac{4p}{n^2} + \frac{4}{n} + \frac{4}{n^2}\right)\|\boldsymbol{B}\|_F^2\operatorname{tr}(\boldsymbol{\Sigma}_e).
\end{aligned}$$

Bibliography

[1] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, vol. 20, Springer, 2010.

[2] Z. D. Bai, *Methodologies in spectral analysis of large dimensional random matrices, a review*, Advances in Statistics, (1999), pp. 174–240.

[3] M. Bayati, M. A. Erdogdu, and A. Montanari, *Estimating lasso risk and noise level*, Advances in Neural Information Processing Systems, 26 (2013).

[4] S. Chatterjee, *A new method of normal approximation*, The Annals of Probability, 36 (2008), pp. 1584 – 1610.

[5] S. Chatterjee, *Fluctuations of eigenvalues and second order poincaré inequalities*, Probability Theory and Related Fields, 143 (2009), pp. 1–40.

[6] G. De los Campos, D. Sorensen, and D. Gianola, *Genomic heritability: What is it?*, PLOS Genetics, 11 (2015), pp. 1–21.

[7] L. H. Dicker, *Variance estimation in high-dimensional linear models*, Biometrika, 101 (2014), pp. 269–284.

[8] L. H. Dicker and M. A. Erdogdu, *Maximum likelihood for variance estimation in high-dimensional linear models*, in Artificial Intelligence and Statistics, PMLR, 2016, pp. 159–167.

[9] L. H. Dicker and M. A. Erdogdu, *Flexible results for quadratic forms with applications to variance components estimation*, The Annals of Statistics, 45 (2017), pp. 386–414.

[10] E. Dobriban and S. Wager, *High-dimensional asymptotics of prediction: Ridge regression and classification*, Ann. Statist., 46 (2018), pp. 247–279.

[11] D. S. Falconer, *Introduction to quantitative genetics*, Edinburgh: Oliver and Boyd, 1961.

[12] J. Fan, S. Guo, and N. Hao, *Variance estimation using refitted cross-validation in ultrahigh dimensional regression*, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 74 (2012), pp. 37–65.

[13] T. Ge, M. Reuter, A. M. Winkler, A. J. Holmes, P. H. Lee, L. S. Tirrell, J. L. Roffman, R. L. Buckner, J. W. Smoller, and M. R. Sabuncu, *Multidimensional heritability analysis of neuroanatomical shape*, Nature communications, 7 (2016), p. 13291.

[14] P. Graczyk, G. Letac, and H. Massam, *The hyperoctahedral group, symmetric group representations and the moments of the real wishart distribution*, Journal of Theoretical Probability, 18 (2005), pp. 1–42.

[15] A. Gusev, S. H. Lee, G. Trynka, H. Finucane, B. J. Vilhjálmsson, H. Xu, C. Zang, S. Ripke, B. Bulik-Sullivan, E. Stahl, and et. al., *Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases*, 2014.

[16] H. O. Hartley and J. N. K. Rao, *Maximum-likelihood estimation for the mixed analysis of variance model*, Biometrika, 54 (1967), pp. 93–108.

[17] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1990.

[18] X. Hu and X. Li, *Misspecification analysis of high-dimensional random effects models for estimation of signal-to-noise ratios*, arXiv preprint arXiv:2202.06400, (2022).

[19] N. Ignatiadis and P. Lolas, *Group-regularized ridge regression via empirical bayes noise level cross-validation*, arXiv e-prints, (2020), pp. arXiv–2010.

[20] L. Janson, R. F. Barber, and E. Candes, *Eigenprism: inference for high dimensional signal-to-noise ratios*, Journal of the Royal Statistical Society. Series B, Statistical methodology, 79 (2017), p. 1037.

[21] J. Jiang, *Reml estimation: Asymptotic behavior and related topics*, The Annals of Statistics, (1996), pp. 255–286.

[22] J. Jiang, *Linear and generalized linear mixed models and their applications.*, Springer, 257 (2007), pp. 1–28.

[23] J. Jiang., *Large Sample Techniques for Statistics*, Cambridge University Press, 2010.

[24] J. Jiang, C. Li, D. Paul, C. Yang, and H. Zhao, *On high-dimensional misspecified mixed model analysis in genome-wide association study*, The Annals of Statistics, 44 (2016), pp. 2127–2160.

[25] B. Laurent and P. Massart, *Adaptive estimation of a quadratic functional by model selection*, The Annals of Statistics, 28 (2000), pp. 1302 – 1338.

[26] E. L. Lehmann and G. Casella, *Theory of point estimation*, Springer Science & Business Media, 2006.

[27] G. Letac and H. Massam, *All Invariant Moments of the Wishart Distribution*, Scandinavian Journal of Statistics, 31 (2004), pp. 295–318.

[28] Z. Li, X. Hu, and X. Li, *Estimating signal-to-noise ratios for multivariate high-dimensional linear models*, 2025.

[29] R. Ma and L. H. Dicker, *The mahalanobis kernel for heritability estimation in genome-wide association studies: fixed-effects and random-effects methods*, (2019).

[30] V. A. Marčenko and L. A. Pastur, *DISTRIBUTION OF EIGENVALUES FOR SOME SETS OF RANDOM MATRICES*, Mathematics of the USSR-Sbornik, 1 (1967), pp. 457–483.

[31] A. Owen, *Quasi-regression for heritability*, 2012.

[32] P. S. Rao, *Variance components: mixed models, methodologies and applications*, vol. 78, CRC Press, 1997.

[33] M. Rudelson and R. Vershynin, *Hanson-wright inequality and sub-gaussian concentration*, Electronic Communications in Probability, 18 (2013), pp. 1 – 9.

[34] J. Silverstein and S. Choi, *Analysis of the limiting spectral distribution of large dimensional random matrices*, Journal of Multivariate Analysis, 54 (1995), pp. 295–309.

[35] Y. Song, X. Lin, and P. Sur, *Hede: Heritability estimation in high dimensions by ensembling debiased estimators*, arXiv preprint arXiv:2406.11184, (2024).

[36] D. Steinsaltz, A. Dahl, and K. W. Wachter, *Statistical properties of simple random-effects models for genetic heritability*, Electronic journal of statistics, 12 (2018), pp. 321–356.

[37] T. Sun and C.-H. Zhang, *Scaled sparse linear regression*, Biometrika, 99 (2012), pp. 879–898.

[38] J. A. Tropp, *An introduction to matrix concentration inequalities*, 2015.

[39] L. A. P. V. A. Marchenko, *Distribution of eigenvalues for some sets of random matrices*, Mat. Sb. (N.S.), 72(114) (1967), pp. 507–536.

[40] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, arXiv preprint arXiv:1011.3027, (2010).

[41] N. Verzelen and E. Gassiat, *Adaptive estimation of high-dimensional signal-to-noise ratios*, (2018).

[42] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2019.

[43] J. Yang, S. H. Lee, M. Goddard, and P. Visscher, *Gcta: a tool for genome-wide complex trait analysis*, American journal of human genetics, 88 (2011), pp. 76–82.

[44] J. Yang, S. H. Lee, M. E. Goddard, and P. M. Visscher, *Gcta: a tool for genome-wide complex trait analysis*, The American Journal of Human Genetics, 88 (2011), pp. 76–82.

[45] J. Yang, J. Zeng, M. E. Goddard, N. R. Wray, and P. M. Visscher, *Concepts, estimation and interpretation of snp-based heritability*, Nature genetics, 49 (2017), pp. 1304–1310.

[46] H. Zhou, L. Hu, J. Zhou, and K. Lange, *Mm algorithms for variance components models*, Journal of Computational and Graphical Statistics, 28 (2019), pp. 350–361. PMID: 31592195.

[47] J. J. Zhou, M. H. Cho, C. Lange, S. Lutz, E. K. Silverman, and N. M. Laird, *Integrating multiple correlated phenotypes for genetic association analysis by maximizing heritability*, Human heredity, 79 (2015), pp. 93–104.