

Article

Consistency of Restricted Maximum Likelihood Estimators in High-Dimensional Kernel Linear Mixed-Effects Models with Applications in Estimating Genetic Heritability

Xiaoxi Shen ¹  and Qing Lu ^{2,*} ¹ Department of Mathematics, Texas State University, San Marcos, TX 78666, USA; rcd67@txstate.edu² Department of Biostatistics, University of Florida, Gainesville, FL 32611, USA

* Correspondence: lucienq@phhp.ufl.edu

Abstract

Restricted maximum likelihood (REML) estimators are commonly used to obtain unbiased estimators for the variance components in linear mixed models. In modern applications, particularly in genomic studies, the dimension of the design matrix with respect to the random effects can be high. Motivated by this, we first introduce high-dimensional kernel linear mixed models, derive the REML equations, and establish theoretical results on the consistency of REML estimators for several commonly used kernel matrices. The validity of the theories is demonstrated via simulation studies. Our results provide rigorous justification for the consistency of REML estimators in high-dimensional kernel linear mixed models and offer insights into the application of estimating genetic heritability. Finally, we apply the kernel linear mixed models to estimate genetic heritability in a real-world data application.

Keywords: mixed effects models; restricted maximum likelihood; kernel matrix; random matrix theory

MSC: 62F12



Academic Editor: MinJae Lee

Received: 12 June 2025

Revised: 16 July 2025

Accepted: 20 July 2025

Published: 23 July 2025

Citation: Shen, X.; Lu, Q. Consistency of Restricted Maximum Likelihood Estimators in High-Dimensional Kernel Linear Mixed-Effects Models with Applications in Estimating Genetic Heritability. *Mathematics* **2025**, *13*, 2363. <https://doi.org/10.3390/math13152363>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Heritability, particularly narrow-sense heritability, refers to the proportion of phenotypic variance explained by additive genetic variance. For example, it is well-known that the estimated heritability of human height is roughly 0.8 [1,2]. However, early identified genes could only explain a small fraction of height heritability. More recently, large-scale genome-wide association studies (GWASs) involving 5.4 million individuals have identified 12,111 height-related single-nucleotide polymorphisms (SNPs). It has been shown that common SNPs can explain 40–50% of the phenotypic variation in human height [3]. Therefore, the previously observed “missing heritability” can be attributed to the small effect sizes of individual SNPs. Furthermore, because a classical GWAS relies on linear models and multiple testing procedures, a SNP must meet stringent criteria at the whole-genome level (e.g., p -value $< 10^{-8}$) to be considered significant. As a result, many SNPs with small yet meaningful effects remain undetectable using classical GWAS methods.

The idea that heritability is spread across multiple SNPs that contribute to phenotypic traits, also known as polygenicity, has been widely accepted over the past few decades. As a result, rather than estimating the individual effect of each SNP, state-of-the-art models in GWASs estimate the cumulative effect of multiple SNPs in a genetic region, such as all SNPs

in a gene, a pathway, or even a chromosome. For quantitative traits, linear mixed-effects models are commonly used to model the cumulative effect of multiple SNPs. It was shown that, using a linear mixed model, known as genome-wide complex trait analysis (GCTA) [4], 45% of the variance in human height could be explained by all common SNPs [5]. This finding has since been supported by the large-scale GWAS mentioned above.

Linear mixed-effects models (LMMs) provide powerful alternatives in GWASs, and various studies have explored LMMs in the high-dimensional setting. In Li et al. [6], the authors proposed inference methods for linear mixed models using techniques similar to the debiased LASSO method [7,8], achieving an estimator of the fixed effects with the optimal convergence rate and valid inferential procedures. However, their framework assumes that the cluster size of the random effects is large, which is generally not true in genetic studies where linear mixed-effects models are commonly applied. Similarly, Law and Ritov [9] proposed methods for constructing asymptotic confidence intervals and hypothesis tests for the significance of random effects. In both studies, however, high dimensionality resides in the fixed effects, while the dimension of the random effects is assumed to be low. This limits their applicability to genetic studies, where the SNP matrix is often treated as the design matrix for random effects [4,10].

Therefore, it is of greater interest to investigate the properties of variance component estimators when the random effects are high-dimensional. Jiang et al. [11] provided theoretical guarantees for the consistency of the restricted maximum likelihood (REML) estimator under the setting where the random-effects design matrix is high-dimensional and random in the so-called linear regime ($n/p \rightarrow \tau$ for some $0 < \tau \leq 1$). These techniques have since been extended to construct confidence intervals for variance components [12] and to establish the asymptotic normality of heritability estimators based on GWAS summary statistics [13].

Currently, the most commonly used LMMs in GWASs, as studied in Jiang et al. [11], for a polygenetic quantitative trait can be expressed via the following additive model:

$$y_i = \mu_i + \sum_{j \in \mathcal{C}} G_{ij}u_j + \epsilon_i, \quad i = 1, \dots, n, \tag{1}$$

where μ_i is the fixed effect of the i th individual, which is typically of the form $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$, and \mathbf{x}_i are covariates (e.g., age, gender, and race group) of the i th individual; \mathcal{C} is a set of SNPs; and G_{ij} is the genotype of the j th SNP of the i th individual. u_j is the effect of the j th SNP, which typically follows a normal distribution $u_1, \dots, u_{|\mathcal{C}|} \sim$ i.i.d. $\mathcal{N}(0, \sigma_u^2)$ with $|\mathcal{C}|$ being the cardinality of the set \mathcal{C} . ϵ_i is the random error or other possible environmental effects, which are typically modeled by a normal distribution $\epsilon_1, \dots, \epsilon_n \sim$ i.i.d. $\mathcal{N}(0, \sigma_\epsilon^2)$. By writing $g_i = \sum_{j \in \mathcal{C}} G_{ij}u_j$, model (1) can be written as

$$y_i = \mu_i + g_i + \epsilon_i, \quad i = 1, \dots, n,$$

or in a vector form

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{g} + \boldsymbol{\epsilon}, \tag{2}$$

where $\mathbf{y} = [y_1, \dots, y_n]^T$, $\boldsymbol{\mu} = [\mu_1, \dots, \mu_n]^T$, and $\boldsymbol{\epsilon} = [\epsilon_1, \dots, \epsilon_n]^T$. Under the distributional assumptions in Model (1), the distributions of \mathbf{g} and $\boldsymbol{\epsilon}$ are

$$\mathbf{g} \sim \mathcal{N}_n(\mathbf{0}, \sigma_u^2 \mathbf{G} \mathbf{G}^T), \quad \boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_n), \tag{3}$$

where the matrix \mathbf{G} is an $n \times |\mathcal{C}|$ -matrix containing SNPs in region \mathcal{C} .

An underlying assumption in Model (1) or (2) is that the genetic effect associated with the quantitative trait of interest is linear. This may not be a reasonable assumption

in reality, as generally the relationships between genotypes and phenotypes are rather complex. Therefore, the following semi-parametric models could be more appropriate.

$$y_i = \mu_i + h(G_{i1}, \dots, G_{i|C|}) + \epsilon_i, \quad i = 1, \dots, n, \tag{4}$$

where $h \in \mathcal{H}$ and \mathcal{H} is some function space. In particular, let \mathcal{H} be a reproducing kernel Hilbert space (RKHS) associated with a kernel function $K(\cdot, \cdot)$. The representer theorem [14] shows that when $\mu_i = \mathbf{x}_i^T \boldsymbol{\beta}$, the estimating equation for $\boldsymbol{\beta}$ and $\mathbf{h} = [h(G_{11}, \dots, G_{1|C|}), \dots, h(G_{n1}, \dots, G_{n|C|})]^T$ under penalized least squares is equivalent to the Henderson mixed model equation, which is often used to obtain the BLUP in a linear mixed model. In this case, the linear mixed model is given as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{a} + \boldsymbol{\epsilon}, \tag{5}$$

where $\mathbf{a} \sim \mathcal{N}_n(\mathbf{0}, \sigma_a^2 \mathbf{K})$ with \mathbf{K} being a kernel matrix, $\boldsymbol{\epsilon} \in \mathcal{N}_n(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_n)$, \mathbf{a} and $\boldsymbol{\epsilon}$ are independent.

In this paper, we study the consistency of the REML estimator for variance components in a high-dimensional kernel linear mixed-effects model. Such models have also been widely used in spatial statistics [15], as well as in identifying significant genetic variants [10,16] and predicting genetic risk for diseases [17]. The results established in this paper extend the work by Jiang et al. [11] from linear kernels to other commonly used kernels that capture non-linear genetic effects, thereby providing theoretical justification for the successful application of kernel LMMs in genetic studies.

The rest of the paper is organized as follows: Section 2 provides the formulation of the REML equations for kernel linear mixed-effects models and the main consistency results for the REML estimators. The main results rely heavily on random matrix theories (RMTs), and a brief review of the RMTs is also given in Section 2. Simulation studies are given in Section 3 to verify the theoretical results, followed by a real-world data application.

Notations : Throughout this paper, bold alphabetic or Greek letters are used for vectors, while bold capital alphabetic or Greek letters are used for matrices. For an $n \times n$ matrix \mathbf{A}, \mathbf{B} , $\mathbf{A} \geq 0$ means that \mathbf{A} is positive and semi-definite, while $\mathbf{A} \geq \mathbf{B}$ means that $\mathbf{A} - \mathbf{B} \geq 0$. For a vector \mathbf{x} , $\|\mathbf{x}\| = \sqrt{\sum_{i=1}^n x_i^2}$ denotes its Euclidean norm, and we use $\|\mathbf{A}\|$ to denote the operator norm of a matrix \mathbf{A} ; that is, $\|\mathbf{A}\| = \sqrt{\lambda_{\max}(\mathbf{A}^T \mathbf{A})}$. When \mathbf{A} is symmetric and positive semi-definite, this is the same as the largest eigenvalue of \mathbf{A} . $\det(\mathbf{A})$ is used to denote the determinant of a squared matrix \mathbf{A} . For two matrices \mathbf{A}, \mathbf{B} of the same size, $\mathbf{A} \circ \mathbf{B}$ denotes the Hadamard product; i.e., $[\mathbf{A} \circ \mathbf{B}]_{ij} = A_{ij} B_{ij}$. Moreover, if $f : \mathbb{R} \rightarrow \mathbb{R}$ is a univariate function and \mathbf{A} is a matrix, then $f[\mathbf{A}]$ denotes a matrix of the same size as \mathbf{A} and $[f(\mathbf{A})]_{ij} = f(A_{ij})$.

2. Materials and Methods

2.1. The REML Equations for Variance Components

Let q be the rank of the fixed-effects design matrix \mathbf{X} and $\mathbf{A} \in \mathbb{R}^{n \times (n-q)}$ such that $\mathbf{A}^T \mathbf{X} = \mathbf{0}$ and $\mathbf{A}^T \mathbf{A} = \mathbf{I}_{n-q}$. Then, by multiplying \mathbf{A}^T on both sides of Equation (5), we obtain

$$\tilde{\mathbf{y}} = \tilde{\mathbf{a}} + \tilde{\boldsymbol{\epsilon}}, \tag{6}$$

where $\tilde{\mathbf{y}} = \mathbf{A}^T \mathbf{y}$ and

$$\begin{aligned} \tilde{\mathbf{a}} &= \mathbf{A}^T \mathbf{a} \sim \mathcal{N}_n(\mathbf{0}, \sigma_a^2 \mathbf{A}^T \mathbf{K} \mathbf{A}), \\ \tilde{\boldsymbol{\epsilon}} &= \mathbf{A}^T \boldsymbol{\epsilon} \sim \mathcal{N}_n(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_{n-q}). \end{aligned}$$

Let $f(\tilde{\mathbf{y}}|\boldsymbol{\theta})$ be the probability density function of $\tilde{\mathbf{y}}$. Then the restricted log-likelihood of $\boldsymbol{\theta} := [\sigma_a^2, \sigma_\epsilon^2]^T$ is

$$\begin{aligned} \ell_r(\boldsymbol{\theta}) &= \log f(\tilde{\mathbf{y}}|\boldsymbol{\theta}) \propto -\frac{1}{2} \log \det(\sigma_a^2 \mathbf{A}^T \mathbf{K} \mathbf{A} + \sigma_\epsilon^2 \mathbf{I}_{n-q}) - \frac{1}{2} \tilde{\mathbf{y}}^T (\sigma_a^2 \mathbf{A}^T \mathbf{K} \mathbf{A} + \sigma_\epsilon^2 \mathbf{I}_{n-q})^{-1} \tilde{\mathbf{y}} \\ &= -\frac{1}{2} \log \det(\sigma_\epsilon^2 (\gamma \mathbf{A}^T \mathbf{K} \mathbf{A} + \mathbf{I}_{n-q})) - \frac{1}{2\sigma_\epsilon^2} \tilde{\mathbf{y}}^T (\gamma \mathbf{A}^T \mathbf{K} \mathbf{A} + \mathbf{I}_{n-q})^{-1} \tilde{\mathbf{y}} \\ &= -\frac{n-q}{2} \log \sigma_\epsilon^2 - \frac{1}{2} \log \det(\gamma \mathbf{A}^T \mathbf{K} \mathbf{A} + \mathbf{I}_{n-q}) - \frac{1}{2\sigma_\epsilon^2} \tilde{\mathbf{y}}^T (\gamma \mathbf{A}^T \mathbf{K} \mathbf{A} + \mathbf{I}_{n-q})^{-1} \tilde{\mathbf{y}}, \end{aligned}$$

where $\gamma = \sigma_a^2/\sigma_\epsilon^2$ is the ratio between two variance components. With slight abuse of notation, let $\boldsymbol{\theta} = [\gamma, \sigma_\epsilon^2]^T$ in what follows.

A well-known result for REML estimators is that they are independent of the choice of \mathbf{A} , which is based on the following identity [18]

$$\mathbf{A}(\gamma \mathbf{A}^T \mathbf{K} \mathbf{A} + \mathbf{I}_{n-q})^{-1} \mathbf{A}^T = \mathbf{V}_\gamma^{-1} - \mathbf{V}_\gamma^{-1} \mathbf{X}(\mathbf{X}^T \mathbf{V}_\gamma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}_\gamma^{-1},$$

where $\mathbf{V}_\gamma = \mathbf{I}_n + \gamma \mathbf{K}$. Also, let $\boldsymbol{\Sigma}_\gamma = \mathbf{I}_{n-q} + \gamma \mathbf{A}^T \mathbf{K} \mathbf{A}$, then

$$\mathbf{P}_\gamma := \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T = \mathbf{V}_\gamma^{-1} - \mathbf{V}_\gamma^{-1} \mathbf{X}(\mathbf{X}^T \mathbf{V}_\gamma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}_\gamma^{-1}.$$

Therefore, $\tilde{\mathbf{y}}^T (\gamma \mathbf{A}^T \mathbf{K} \mathbf{A} + \mathbf{I}_{n-q})^{-1} \tilde{\mathbf{y}} = \mathbf{y}^T \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{y} = \mathbf{y}^T \mathbf{P}_\gamma \mathbf{y}$, and the restricted log-likelihood function can be written as

$$\ell_r(\boldsymbol{\theta}) \propto -\frac{n-q}{2} \log \sigma_\epsilon^2 - \frac{1}{2} \log \det(\boldsymbol{\Sigma}_\gamma) - \frac{1}{2\sigma_\epsilon^2} \mathbf{y}^T \mathbf{P}_\gamma \mathbf{y}. \tag{7}$$

By taking the derivatives with respect to γ and σ_ϵ^2 and letting them equal to zero, we know that the REML estimators of γ and σ_ϵ^2 satisfy the following REML equations:

$$\begin{cases} \hat{\sigma}_\epsilon^2 = \frac{1}{n-q} \mathbf{y}^T \mathbf{P}_\gamma \mathbf{y} \\ \frac{\mathbf{y}^T \mathbf{P}_\gamma \mathbf{K} \mathbf{P}_\gamma \mathbf{y}}{\text{tr}(\mathbf{P}_\gamma \mathbf{K})} = \frac{\mathbf{y}^T \mathbf{P}_\gamma \mathbf{y}}{n-q} \end{cases} \tag{8}$$

2.2. A Review on Random Matrix Theory

In this section, we briefly review the theory of random matrices. For more details on the application of random matrix theory in statistics, we refer interested readers to Bai and Silverstein [19] and Paul and Aue [20]. Let \mathbf{A} be an $m \times m$ matrix with eigenvalues $\lambda_j, j = 1, 2, \dots, m$. The empirical spectral distribution (ESD) of the matrix \mathbf{A} is defined to be

$$F_m^A(x) = \frac{1}{m} \sum_{j=1}^m \mathbb{I}_{\{\lambda_j \leq x\}}, \quad x \in \mathbb{R}.$$

For a double array of i.i.d. random variables $\{g_{jk} : j, k = 1, 2, \dots\}$ with a mean of 0 and a variance of σ^2 , write $\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_p] = \begin{bmatrix} \mathbf{g}'_1 \\ \vdots \\ \mathbf{g}'_n \end{bmatrix}$ with $\mathbf{g}_j = (g_{1j}, \dots, g_{nj})^T$ and $\mathbf{g}'_j = (g_{j1}, \dots, g_{jp})^T$. It is well known that the ESD of the sample covariance matrix $\mathbf{S} = \frac{1}{p} \mathbf{G} \mathbf{G}^T$ converges almost surely in distribution to the Marchenko–Pastur (M-P) law.

Theorem 1 (Marchenko–Pastur Law). *Suppose that $\{g_{ij}\}$ are i.i.d. random variables with a mean of 0 and a variance of σ^2 . Also assume that $n/p \rightarrow \tau \in (0, \infty)$ as $n, p \rightarrow \infty$. Then, with a probability of 1, F^S tends toward the M-P law having density*

$$\varphi_\tau(x) = \begin{cases} (2\pi\sigma^2\tau x)^{-1}\sqrt{(b_+(\tau) - x)(x - b_-(\tau))} & \text{if } x \in [b_-(\tau), b_+(\tau)] \\ 0 & \text{otherwise.} \end{cases},$$

and has a point mass $1 - 1/\tau$ at the origin if $\tau > 1$, where $b_\pm(\tau) = \sigma^2(1 \pm \sqrt{\tau})^2$.

The following corollary is by Jiang et al. [11], which is a consequence of convergence in distribution and will be frequently used.

Corollary 1. Under the assumptions of Theorem 1, for the integer l ,

$$n^{-1}\text{tr}(\mathbf{S}^l) \rightarrow \int_{b_-(\tau)}^{b_+(\tau)} x^l \varphi_\tau(x) dx, \text{ as } p \rightarrow \infty.$$

In general, the assumptions on i.i.d. can be reduced. The following theorem is Theorem 4.3 in Bai and Silverstein [19].

Theorem 2. Suppose that the entries of $\mathbf{G} \in \mathbb{R}^{n \times p}$ are random variables that are independent of each n and identically distributed for all n and satisfy $\mathbb{E}[|g_{11} - \mathbb{E}[g_{11}]|^2] = 1$. Also assume that $\mathbf{T} = \text{diag}\{\tau_1, \dots, \tau_p\}$, τ_i is real, and that the empirical distribution function of $\{\tau_1, \dots, \tau_p\}$ converges a.s. to a probability distribution function H as $n \rightarrow \infty$. The entries of both \mathbf{G} and \mathbf{T} may depend on n . Set

$$\mathbf{B} = \mathbf{A} + \frac{1}{n}\mathbf{G}\mathbf{T}\mathbf{G}^T,$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is symmetric and satisfies $F^{\mathbf{A}} \rightarrow H$ almost surely, where H is a distribution (possibly defective) on the real line. Assume that \mathbf{G} , \mathbf{T} , and \mathbf{A} are independent. When $p/n \rightarrow y > 0$ as $n \rightarrow \infty$, then the ESD of \mathbf{B} and $F^{\mathbf{B}}$ converges almost surely as $n \rightarrow \infty$ to a nonrandom distribution function F .

The limiting spectral behaviors of $p^{-1}\mathbf{G}\mathbf{G}^T$ have not only been studied, but the limiting spectral behaviors of inner-product kernel matrices $f[p^{-1}\mathbf{G}\mathbf{G}^T]$ have also been studied. Here we quote Theorem 2.1 in El Karoui [21].

Theorem 3 (Spectrum of Inner-Product Kernel Random Matrices). Let us assume that we observe n i.i.d. random vectors $\mathbf{g}'_i \in \mathbb{R}^p$. Consider the kernel matrix \mathbf{K} with entries

$$K_{ij} = f\left(\frac{\mathbf{g}'_i{}^T \mathbf{g}'_j}{p}\right).$$

Assume that

1. $n \asymp p$; that is, n/p and p/n remain bounded as $p \rightarrow \infty$.
2. Σ_p is a positive definite $p \times p$ matrix and $\|\Sigma_p\|$ remains bounded in p ; that is, there exists $C > 0$ such that $\|\Sigma_p\| \leq C$ for all p .
3. Σ_p/p has a finite limit; that is, there exists $l \in \mathbb{R}$ such that $\lim_{p \rightarrow \infty} \text{tr}(\Sigma_p)/p = l$.
4. $\mathbf{g}'_i = \Sigma_p^{1/2}\Gamma_i$.
5. The entries of Γ_i , a p -dimensional random vector, are i.i.d. Also, as denoted by $\Gamma_i(k)$, the k th entry of Γ_i , we assume that $\mathbb{E}[\Gamma_i(k)] = 0$, $\text{Var}[\Gamma_i(k)] = 1$, and $\mathbb{E}[|\Gamma_i(k)|^{4+\eta}] < \infty$ for some $\eta > 0$.
6. f is a C^1 function in the neighborhood of $l = \lim_{p \rightarrow \infty} \text{tr}(\Sigma_p)/p$ and a C^3 function in the neighborhood of 0.

Under these assumptions, the kernel matrix \mathbf{K} can (in probability) be approximated consistently in the operator norm when p and n tend to ∞ by the matrix \mathbf{M} , where

$$\mathbf{M} = \left(f(0) + f''(0) \frac{\text{tr}(\boldsymbol{\Sigma}_p^2)}{2p^2} \right) \mathbf{1}_n \mathbf{1}_n^T + f'(0) \frac{\mathbf{G}\mathbf{G}^T}{p} + v_p \mathbf{I}_n, \tag{9}$$

where

$$v_p = f\left(\frac{\text{tr}(\boldsymbol{\Sigma}_p)}{p}\right) - f(0) - f'(0) \frac{\text{tr}(\boldsymbol{\Sigma}_p)}{p}.$$

In other words,

$$\|\mathbf{K} - \mathbf{M}\| \xrightarrow{p} 0, \text{ when } p \rightarrow \infty.$$

El Karoui [21] also studied the behaviors of the spectrum of Euclidean distance kernel random matrices or, in other words, radial basis kernel matrices. The following result is from Theorem 2.2 from El Karoui [21].

Theorem 4 (Spectrum of Euclidean Distance Kernel Random Matrices). *Consider the $n \times n$ kernel matrix \mathbf{K} with entries*

$$K_{ij} = f\left(\frac{\|\mathbf{g}'_i - \mathbf{g}'_j\|}{p}\right).$$

Let us call

$$\tau = 2 \frac{\text{tr}(\boldsymbol{\Sigma}_p)}{p}.$$

Let us call $\boldsymbol{\psi}$ the vector with the i th entry $\psi_i = \|\mathbf{g}'_i\|_2^2/p - \text{tr}(\boldsymbol{\Sigma}_p)/p$. Suppose that the assumptions of Theorem 3 hold, but that conditions 5 and 6 are replaced by

- (5') The entries of $\boldsymbol{\Gamma}_i$, a p -dimensional random vector, are i.i.d. Also, as denoted by $\boldsymbol{\Gamma}_i(k)$ the k th entry of $\boldsymbol{\Gamma}_i$, we assume that $\mathbb{E}[\boldsymbol{\Gamma}_i(k)] = 0$, $\text{Var}[\boldsymbol{\Gamma}_i(k)] = 1$, and $\mathbb{E}[|\boldsymbol{\Gamma}_i(k)|^{5+\eta}] < \infty$ for some $\eta > 0$.
- (6') f is C^3 in a neighborhood of τ .

Then \mathbf{K} can be approximated consistently in the operator norm (and in probability) by the matrix \mathbf{M} , which is defined by

$$\begin{aligned} \mathbf{M} = & f(\tau) \mathbf{1}_n \mathbf{1}_n^T + f'(\tau) \left[\mathbf{1}_n \boldsymbol{\psi}^T + \boldsymbol{\psi} \mathbf{1}_n^T - 2 \frac{\mathbf{G}\mathbf{G}^T}{p} \right] + \\ & \frac{f''(\tau)}{2} \left[\mathbf{1}_n (\boldsymbol{\psi} \circ \boldsymbol{\psi})^T + (\boldsymbol{\psi} \circ \boldsymbol{\psi}) \mathbf{1}_n^T + 2 \boldsymbol{\psi} \boldsymbol{\psi}^T + 4 \frac{\text{tr}(\boldsymbol{\Sigma}_p^2)}{p^2} \mathbf{1}_n \mathbf{1}_n^T \right] + v_p \mathbf{I}_n, \end{aligned} \tag{10}$$

where $v_p = f(0) + \tau f'(\tau) - f(\tau)$. In other words,

$$\|\mathbf{K} - \mathbf{M}\| \xrightarrow{p} 0, \text{ when } p \rightarrow \infty.$$

The following lemma provides a useful bound for the variance of a quadratic form, and its proof is given in Appendix A.

Lemma 1. *Suppose that $\mathbf{y} = \boldsymbol{\mu} + \mathbf{e}$, with $\boldsymbol{\mu}$ being a fixed vector; $\mathbb{E}[\mathbf{e}] = 0$; $\mathbb{E}[\mathbf{e}\mathbf{e}^T] = \boldsymbol{\Sigma}$; and $\mathbb{E}[|\mathbf{e}|^4] < \infty$. Then*

$$\text{Var}[\mathbf{y}^T \boldsymbol{\Gamma} \mathbf{y}] \leq 2 \text{Var}[\mathbf{e}^T \boldsymbol{\Gamma} \mathbf{e}] + 8 \boldsymbol{\mu}^T \boldsymbol{\Gamma}^T \boldsymbol{\Sigma} \boldsymbol{\Gamma} \boldsymbol{\mu}.$$

2.3. Consistency of REML Estimators in Kernel LMMs

In this section, we will show the main results on the consistency of REML estimators for kernel linear mixed models under three commonly used classes of kernel matrices in genetic data analysis. The first one is the weighted product kernel, which is a natural generalization of the product kernel and is the kernel matrix used in a sequence kernel association test (SKAT) [10] in genetic association studies. The other two classes are the inner-product-based kernel matrices and the Euclidean distance-based kernel matrices. An example of an inner-product-based kernel is the polynomial kernel, and an example of a Euclidean distance-based kernel is a Gaussian kernel. Proofs of all results in this section can be found in Appendix A.

2.3.1. Weighted Product Kernel

The following theorem serves as a natural extension of the result by Jiang et al. [11] to the case of a weighted product kernel. It guarantees the consistency of REML estimators of the variance components in kernel linear mixed models when the entries in \mathbf{G} are i.i.d. and have a finite fourth moment, and the weights in the diagonal matrix \mathbf{W} are upper and lower bounded.

Theorem 5. For $\mathbf{K} = p^{-1}\mathbf{G}\mathbf{W}\mathbf{G}^T$ with entries in $\mathbf{G} \in \mathbb{R}^{n \times p}$ and $\mathbf{W} \in \mathbb{R}^{p \times p}$,

1. G_{ij} are i.i.d. with $\mathbb{E}[G_{ij}] = 0$, $\mathbb{E}[G_{ij}^2] = 1$, and $\mathbb{E}[G_{ij}^4] < \infty$;
2. $\mathbf{W} = \text{diag}\{w_1, \dots, w_p\}$ with

$$\max_{1 \leq i \leq p} |w_i| \leq 1, \quad \min_{1 \leq i \leq p} |w_i| \geq \delta, \tag{11}$$

for some $\delta > 0$.

Then as $n/p \rightarrow \tau \in (0, 1)$ and $n, p \rightarrow \infty$, we have

$$\hat{\gamma} \xrightarrow{p} \gamma_0, \quad \hat{\sigma}_\epsilon^2 \xrightarrow{p} \sigma_\epsilon^2,$$

where γ_0 is the true value for γ .

In addition, the same result still holds when \mathbf{G} is replaced by its column standardized version $\tilde{\mathbf{G}}$ as given in the following corollary. The proofs of the theorem and the following corollary are very similar to the proofs in Jiang et al. [11], and they can be found in Appendix A.

Corollary 2. Let \mathbf{G} be a random matrix whose elements are i.i.d. with $\mathbb{E}[G_{ij}] = 0$, $\mathbb{E}[G_{ij}^2] = 1$, and $\mathbb{E}[G_{ij}^4] < \infty$. Let $\tilde{\mathbf{G}}$ be the matrix that is obtained by column standardizing the matrix \mathbf{G} ; i.e.,

$$\tilde{\mathbf{G}} = (\mathbf{G} - \bar{\mathbf{g}} \otimes \mathbf{1}_n)\mathbf{D}_s^{-1},$$

where $\bar{\mathbf{g}} = [\bar{g}_1, \dots, \bar{g}_p]$ with $\bar{g}_j = n^{-1} \sum_{i=1}^n G_{ij}$, $j = 1, \dots, p$, and $\mathbf{D}_s = \text{diag}\{s_1, \dots, s_p\}$ with $s_j^2 = (n-1)^{-1} \sum_{i=1}^n (G_{ij} - \bar{g}_j)^2$. Let $\mathbf{W} \in \mathbb{R}^{p \times p}$ be the weight matrix satisfying Equation (11). Then for $\tilde{\mathbf{K}} = p^{-1}\tilde{\mathbf{G}}\mathbf{W}\tilde{\mathbf{G}}^T$, as $n/p \rightarrow \tau \in (0, 1)$ and $n, p \rightarrow \infty$, we have

$$\hat{\gamma} \xrightarrow{p} \gamma_0, \quad \hat{\sigma}_\epsilon^2 \xrightarrow{p} \sigma_\epsilon^2,$$

where γ_0 is the true value for γ .

The i.i.d. assumption on the entries of \mathbf{G} in Theorem 5 and Corollary 2 may limit their applicability in genetic studies, where SNPs are often correlated due to linkage disequilibrium (LD). However, it is worth noting that the correlation among SNPs does not necessarily hinder the application of these results. This is because the theoretical guarantees rely on the validity of the Marchenko–Pastur law, which was originally established for random matrices with i.i.d. entries having a mean of zero and unit variance. Importantly, the i.i.d. assumption on the entries can be relaxed: the Marchenko–Pastur law still holds for random matrices with i.i.d. rows, as long as each row has a mean of zero and unit variance and satisfies a light tail condition (e.g., sub-Gaussian) [22]. Therefore, as long as the SNP vectors from each individual are i.i.d., we expect that the results in Theorem 5 and Corollary 2 remain valid.

2.3.2. Inner-Product Kernel Matrices

Inner-product kernels, including the linear kernel and the polynomial kernel, are commonly used in practice. Generally, an inner-product-based kernel matrix has the form $\mathbf{K} = f[p^{-1}\mathbf{G}\mathbf{G}^T]$. The first thing that needs to be addressed is the positive definiteness of \mathbf{K} . In fact, based on Hiai [23], a necessary and sufficient condition for \mathbf{K} to be positive definite is when f is real analytic and $f^{(k)}(0) \geq 0$ for all $k \geq 0$. In this subsection, we will implicitly assume that this condition is satisfied.

Theorem 6. For $\mathbf{K} = f[p^{-1}\mathbf{G}\mathbf{G}^T]$ with entries in $\mathbf{G} \in \mathbb{R}^{n \times p}$ and the function f ,

1. $n \asymp p$.
2. G_{ij} are i.i.d. with $\mathbb{E}[G_{ij}] = 0$, $\mathbb{E}[G_{ij}^2] = 1$, and $\mathbb{E}[|G_{ij}|^{4+\eta}] < \infty$ for some $\eta > 0$.
3. f is a C^1 function in the neighborhood of 1 and a C^3 function in the neighborhood of 0.

Then as $n/p \rightarrow \tau \in (0, 1)$ as $n, p \rightarrow \infty$, we have

$$\hat{\gamma} \xrightarrow{p} \gamma_0, \quad \hat{\sigma}_\epsilon^2 \xrightarrow{p} \sigma_\epsilon^2,$$

where γ_0 is the true value for γ .

2.3.3. Euclidean Distance Kernel Matrices

Another commonly used kernels in practice are the Euclidean distance-based kernels, such as the Gaussian kernel. Similar to the case in the inner-product kernel matrices, the positive definiteness of \mathbf{K} is the first issue to be addressed. According to Wendland [24], f being completely monotone is a necessary and sufficient condition for a Euclidean distance kernel matrix to be positive semi-definite. In other words, the function f needs to satisfy $f \in C^\infty([0, \infty))$ and $(-1)^k f^{(k)}(r) \geq 0$ for all $r > 0$ and $k = 0, 1, 2, \dots$. Throughout Section 2.3.3, it will be implicitly assumed that f is completely monotone.

Theorem 7. Let \mathbf{K} be a kernel matrix with entries $K_{ij} = f(p^{-1}\|g'_i - g'_j\|_2^2)$. Suppose that the entries in $\mathbf{G} \in \mathbb{R}^{n \times p}$ and the function f satisfying

1. $n \asymp p$.
2. G_{ij} are i.i.d. with $\mathbb{E}[G_{ij}] = 0$, $\mathbb{E}[G_{ij}^2] = 1$ and $\mathbb{E}[|G_{ij}|^{5+\eta}] < \infty$ for some $\eta > 0$.
3. f is a C^3 function in a neighborhood of 2.

Then as $n/p \rightarrow \tau \in (0, 1)$ and $n, p \rightarrow \infty$, we have

$$\hat{\gamma} \xrightarrow{p} \gamma_0, \quad \hat{\sigma}_\epsilon^2 \xrightarrow{p} \sigma_\epsilon^2,$$

where γ_0 is the true value for γ .

Remark 1. Both Theorems 6 and 7 rely on the approximation of inner-product kernel matrices or Euclidean distance kernel matrices by linear kernels, identity matrices, and some low-rank matrices. Based on the proofs in El Karoui [21], the approximation rate for inner-product kernel matrices is $o_p(p^{-\delta/2})$ for some $\delta < \frac{1}{2}$, while for a Gaussian kernel with entries with $\exp\{-\|\zeta_i - \zeta_j\|^2/2\}$, the approximation rate is $n \exp\{-p^{1/2+2/m+\delta}\}$ with $p^{1/2+2/m+\delta}$ being the rate that $\text{tr}(\text{Var}[\zeta])$ grows to infinity, and this is an extremely fast rate.

3. Results

3.1. Simulation Studies

3.1.1. Weighted Product Kernel

We follow an idea similar to that by Jiang et al. [11]; we simulated the genotype for each SNP. Specifically, the minor allele frequencies (MAFs) for p SNPs $\{f_1, f_2, \dots, f_p\}$ were generated from the uniform distribution $\text{Unif}[0.05, 0.5]$, where f_j is the allele for the j th SNP. In the simulation, the number of SNPs is set to $p = 2n$. For the weight matrix W , the j th element w_j is defined to be

$$w_j = \frac{-\log_{10} f_j}{\max_{1 \leq j \leq p} -\log_{10} f_j}.$$

The logarithm of MAF, which was defined in Li et al. [25], is one of the commonly used weights to detect effects from common variants and also takes the contributions from rare variants into account. To simulate the genotype matrix $\tilde{G} \in \{0, 1, 2\}^{n \times p}$, the Hardy-Weinberg equilibrium was assumed. Specifically, for the j SNP, the genotype value of each individual was sampled from $\{0, 1, 2\}$ according to the probabilities $(1 - f_j)^2, 2f_j(1 - f_j)$, and f_j^2 , respectively. Given the simulated genotypes, column standardization was applied to each column in \tilde{G} , and the resulting genotype matrix was denoted as G . The weighted kernel matrix is defined as

$$K = p^{-1}GWG^T.$$

The responses are simulated based on the following equation:

$$y = X\beta + a + \epsilon, \tag{12}$$

where the fixed-effect design matrix $X = [1_n, \tilde{X}] \in \mathbb{R}^{n \times 3}$ and the elements in \tilde{X} were generated from a standard normal distribution. We set $\beta = [1, 2, -1]^T$. The random effect $a \sim \mathcal{N}_n(\mathbf{0}, 0.6K)$, and the random noise $\epsilon \sim \mathcal{N}_n(\mathbf{0}, 2I_n)$. The choice of the variance component for the random effects is the same as in Jiang et al. [11] under their dense scenario. On the other hand, genetic data is often noisy, and the signal-to-noise ratio is low. To mimic the real situation, we set the variance of the random error to be two. In the simulation, we selected 100, 200, 400, 600, 800, and 1000 as the sample sizes. A total of 1000 Monte Carlo replications were conducted to evaluate the performance of the REML estimators for the variance components in the simulation model.

Figure 1 provides the boxplots for the REML estimators of the variance components in the simulation model (12). As shown from both panels, the REML estimators converge to the true values, which justifies the theoretical results developed in the previous sections. The mean and standard deviation of the REML estimators of the variance components are given in Table 1.

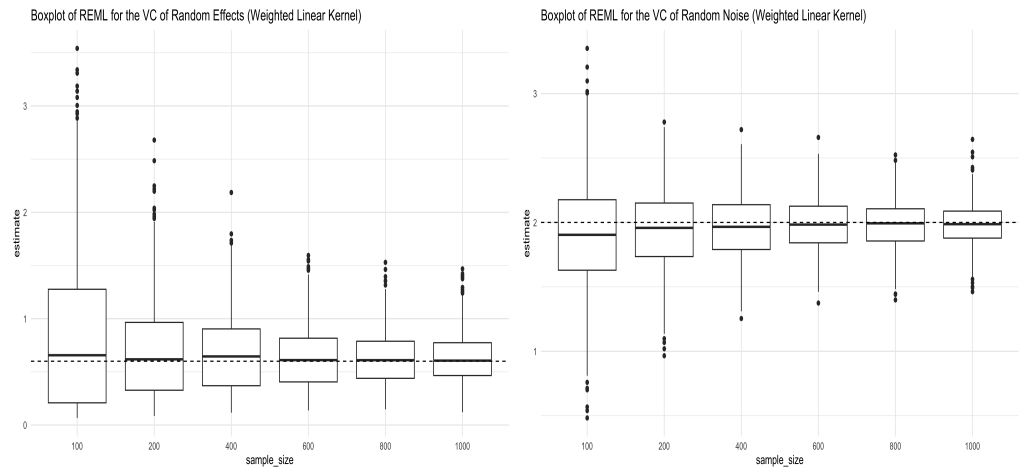


Figure 1. Boxplots for REML estimators of variance components obtained from Simulation (12) under the weighted linear kernel. The left panel shows the boxplots for the REML estimator of σ_a^2 (truth = 0.6), and the right panel shows the boxplots for the REML estimator of σ_ϵ^2 (truth = 2) under different sample sizes.

Table 1. Mean and standard deviations (values in the parenthesis) of REML estimators for σ_a^2 and σ_ϵ^2 based on 1000 Monte Carlo simulations under the weighted linear kernel.

| Sample Size | $\hat{\sigma}_a^2$ | $\hat{\sigma}_\epsilon^2$ |
|-------------|--------------------|---------------------------|
| 100 | 0.837 (0.709) | 1.895 (0.415) |
| 200 | 0.698 (0.451) | 1.941 (0.302) |
| 400 | 0.668 (0.344) | 1.959 (0.237) |
| 600 | 0.634 (0.286) | 1.977 (0.204) |
| 800 | 0.631 (0.247) | 1.979 (0.182) |
| 1000 | 0.631 (0.231) | 1.985 (0.162) |

3.1.2. Inner Product Kernel

We followed the same procedures mentioned in Section 3.1.1 to simulate the genotype matrix G , and Equation (12) was used to generate the response, except that in this case the kernel matrix is $K = (1 + p^{-1}GG^T)^2$, with the polynomial kernel having an order of 2. In the expression, power means the elementwise power. Figure 2 shows the simulation results for the REML estimators of the two variance components in Equation (12). As we can see from both panels in Figure 2, the REML estimators converge to the true values as anticipated. The mean and standard deviation of the REML estimators of the variance components are given in Table 2.

Table 2. Mean and standard deviations (values in the parenthesis) of REML estimators for σ_a^2 and σ_ϵ^2 based on 1000 Monte Carlo simulations under the polynomial kernel with a degree of 2.

| Sample Size | $\hat{\sigma}_a^2$ | $\hat{\sigma}_\epsilon^2$ |
|-------------|--------------------|---------------------------|
| 100 | 0.606 (0.375) | 2.004 (1.021) |
| 200 | 0.618 (0.284) | 1.974 (0.790) |
| 400 | 0.603 (0.206) | 1.985 (0.561) |
| 600 | 0.604 (0.172) | 1.992 (0.469) |
| 800 | 0.603 (0.152) | 1.995 (0.416) |
| 1000 | 0.601 (0.130) | 1.996 (0.364) |

3.1.3. Euclidean Distance Kernel

In this section, we investigated the consistency of REML estimators of the variance components when the kernel used to generate data in Equation (12) is the Gaussian kernel.

When applying the Gaussian kernel $K(u, v) = \exp\{-\frac{\|u-v\|^2}{2\phi}\}$, the tuning parameter ϕ is chosen to be the sample variance of the Euclidean distance between each pair. Figure 3 shows the simulation results for the REML estimators of the two variance components in Equation (12). The REML estimators converge to the true values as shown in both panels in Figure 3. The mean and standard deviation of the REML estimators of the variance components are given in Table 3.

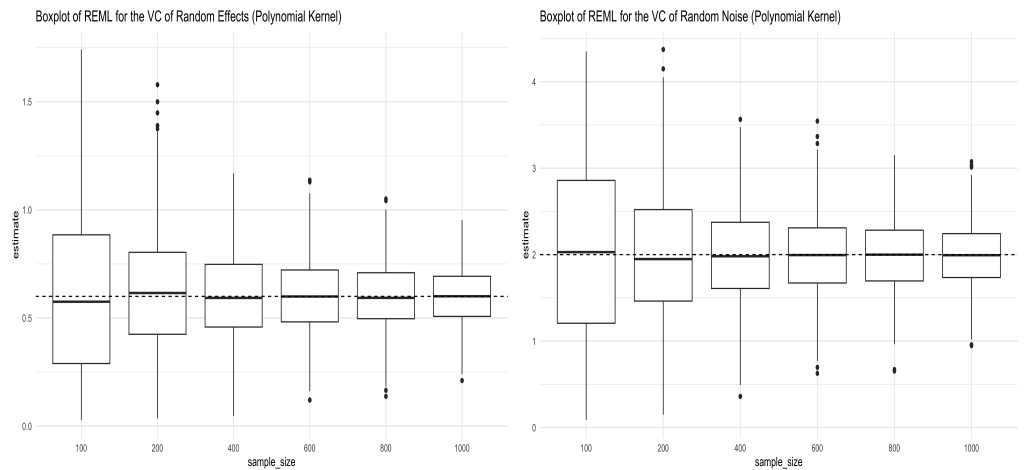


Figure 2. Boxplots for REML estimators of variance components obtained from Simulation (12) under the polynomial kernel with a degree of 2. The left panel shows the boxplots for the REML estimator of σ_a^2 (truth = 0.6), and the right panel shows the boxplots for the REML estimator of σ_ϵ^2 (truth = 2) under different sample sizes.

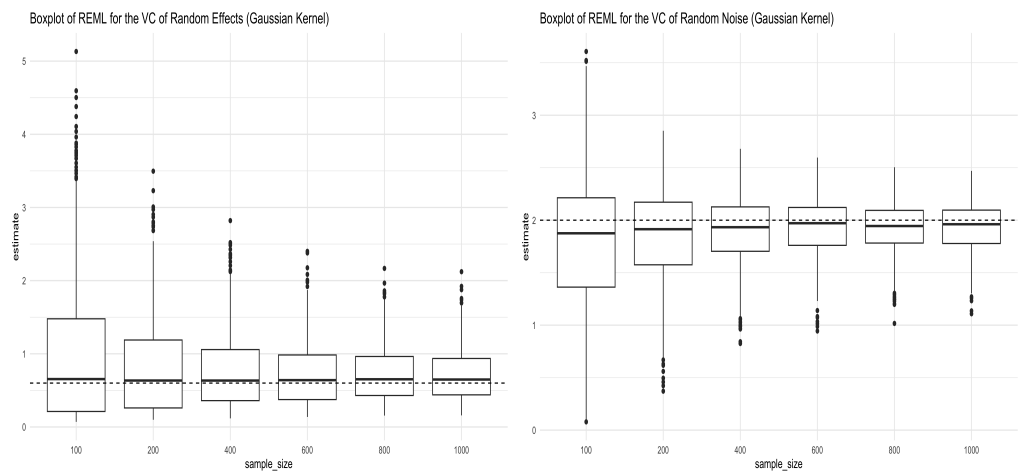


Figure 3. Boxplots for REML estimators of variance components obtained from Simulation (12) under the Gaussian kernel. The left panel shows the boxplots for the REML estimator of σ_a^2 (truth = 0.6), and the right panel shows the boxplots for the REML estimator of σ_ϵ^2 (truth = 2) under different sample sizes.

Table 3. Mean and standard deviations (values in parentheses) of REML estimators for σ_a^2 and σ_ϵ^2 based on 1000 Monte Carlo simulations under the Gaussian kernel.

| Sample Size | $\hat{\sigma}_a^2$ | $\hat{\sigma}_\epsilon^2$ |
|-------------|--------------------|---------------------------|
| 100 | 1.000 (0.955) | 1.774 (0.631) |
| 200 | 0.818 (0.650) | 1.853 (0.432) |
| 400 | 0.755 (0.483) | 1.900 (0.320) |
| 600 | 0.715 (0.406) | 1.934 (0.262) |
| 800 | 0.717 (0.363) | 1.929 (0.229) |
| 1000 | 0.707 (0.335) | 1.938 (0.215) |

3.2. Real Data Analysis

To check the performance of Restricted Maximum Likelihood (REML) estimators in high-dimensional kernel linear mixed models, we conducted a real data analysis similar to that by Jiang et al. [11]. The Study of Addiction: Genetics and Environment (SAGE) (dbGap Study Accession: phs000092.v1.p1) dataset was used to estimate the heritability of body mass index (BMI) based on kernel linear mixed models. To check the consistency of heritability estimation across different traits, we also include height and weight as the phenotypes.

After removing individuals without height or weight measures, a total of $n = 1985$ individuals of European ancestry remain. We also performed a quality control process similar to that by Jiang et al. [11] to avoid bias in the estimation. Specifically, SNPs with a missing rate of $>1\%$, a minor allele frequency (MAF) of $<5\%$, and a p -value of <0.001 from the Hardy–Weinberg equilibrium test were excluded from the analysis. After the quality control process, $p = 4,898,519$ SNPs remained for analysis.

As in Jiang et al. [11], for the fixed-effect design matrix X , besides the intercept, the first 10 principal component scores computed from the product kernels were included. Kernel LMMs with the product kernel $K = p^{-1}GG^T$, a second-order polynomial kernel, and a Gaussian kernel were used to estimate the variance components and hence the heritability. The results are summarized in Table 4. The heritability obtained by the linear kernel is 20.35%, which is very close to the one (19.6%) obtained in Jiang et al. [11]. It can be seen from the table that the heritability estimation of BMI is higher when the polynomial kernel $K = (p^{-1}GG^T)^2$ or the Gaussian kernel $K(u, v) = \exp\{-\frac{\|u-v\|^2}{2\phi}\}$ with $\phi = 1$ is used. The heritability obtained by the polynomial kernel is 50.08%, while the one obtained by the Gaussian kernel is 34.29%.

Table 4. Heritability estimation for BMI, height, and weight under different kernel matrices.

| | Product | Polynomial | Gaussian |
|--------|---------|------------|----------|
| BMI | 20.35% | 50.08% | 34.29% |
| Height | 10.21% | 49.93% | 17.97% |
| Weight | 8.46% | 37.02% | 16.31% |

It is believed that the BMI is highly heritable, and recent twin studies have demonstrated large variation in BMI heritability, ranging from 31% to 90% [26]. Based on the previous results, the heritability estimation obtained using the product kernel is below this range, while the estimations using the Gaussian kernel and the polynomial kernel are within this range. The trend that estimated heritability using the polynomial kernel is the highest among the three choices of kernel and is consistent among the other two phenotypes.

In addition, we also experimented with different values of the constant c in the polynomial kernel, ranging from 0 to 1. Table 5 summarizes the heritability estimates under these choices. The results show that the performance of heritability estimation is sensitive to the choice of kernel hyperparameters. In particular, Table 5 reveals a general pattern: as the value of c increases, the estimated heritability tends to decrease. One explanation is that $(c + \frac{1}{p}GG^T)^2 = c^2J_n + \frac{2c}{p}GG^T + (\frac{1}{p}GG^T)^2$ so that the largest eigenvalue of the $(c + \frac{1}{p}GG^T)^2$ is lower bounded by $\lambda_{\max}(c^2J_n) = nc^2$, which means that when c becomes larger, the rank-one matrix c^2J_n dominates the spectrum of the polynomial kernel matrix. Consequently, the kernel matrix carries less informative structure for estimating the variance component in the random effects, resulting in a smaller estimate of $\hat{\sigma}_a^2$.

Table 5. Heritability estimation for BMI, height, and weight under different kernel matrices.

| | BMI | Height | Weight |
|-----|--------|--------|--------|
| 0 | 50.08% | 49.93% | 37.02% |
| 0.2 | 64.93% | 16.23% | 16.39% |
| 0.4 | 31.26% | 10.54% | 8.82% |
| 0.6 | 21.01% | 7.95% | 6.05% |
| 0.8 | 15.92% | 6.41% | 4.67% |
| 1 | 12.85% | 5.39% | 3.87% |

4. Discussion

Kernel methods have been widely used in machine learning to capture nonlinear relationships between features and responses. In this paper, we demonstrate that when kernels are correctly specified, the REML estimators for the variance components in high-dimensional kernel linear mixed models are consistent across three classes of kernel functions. Simulation studies validate the theoretical results on consistency.

In addition, we would like to highlight a few directions for future work. First, as we have mentioned, the consistency of the REML estimators for variance components in kernel LMMs is established under the assumption that the kernels are correctly specified. In practice, the underlying data-generating process and the appropriate kernel are often unknown. Misspecifying the kernel in a kernel LMM could lead to inconsistent variance component estimators, resulting in biased estimates of heritability. Therefore, it is worth exploring a data-driven approach to identify the appropriate kernel function and evaluate its performance in estimating genetic heritability compared to commonly used kernels.

Second, the theories in this paper are established under the high-dimensional linear regime, where the sample size grows linearly with the number of SNPs. One reason for this assumption is that random matrix theory is widely used to derive theoretical results, and most existing work in random matrix theory focuses on the linear regime. Therefore, it is worthwhile to extend random matrix theory beyond the linear regime. Recently, Ghorbani et al. [27] and Mei et al. [28] developed theories for random kernel matrices under the polynomial regime, where $n \asymp p^\lambda$ with $\lambda < 1$. However, in genetic studies, the number of SNPs is often much larger than the sample size. Thus, it is worthwhile to extend the results in this paper to scenarios where $n \asymp p^\lambda$ and $\lambda > 1$ and to develop the corresponding theories for random kernel matrices under this regime.

Last but not least, as demonstrated in the real data analyses, the performance of kernel LMMs depends heavily on the choice of the kernel matrix. In practice, researchers often lack prior knowledge about which kernel is most appropriate. Therefore, developing a data-driven method for kernel selection would be highly desirable. One widely used strategy is multiple kernel learning [29], in which a single kernel matrix is replaced by a convex combination of several kernel matrices, where the weights of the kernel matrices are also learning by minimizing a loss function (e.g., mean squared error in the regression setting). Extending the current work to accommodate kernel matrices constructed via multiple kernel learning would be a valuable direction for future research.

Author Contributions: Conceptualization, X.S. and Q.L.; methodology and formal analysis, X.S.; writing—original draft preparation, X.S.; writing—review and editing, Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The R code used for implementation can be found at <https://github.com/SxxMichael/KernelLMM> (accessed on 19 July 2025).

Acknowledgments: During the preparation of this manuscript/study, the authors used ChatGPT 4o for the purposes of correcting grammatical mistakes. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Mathematical Proofs of the Main Results

Appendix A.1. Proof of Lemma 1

Proof. Let $Q = \mathbf{y}^T \mathbf{\Gamma} \mathbf{y}$ and note that

$$\begin{aligned} \mathbb{E}[Q] &= \mathbb{E}[\mathbf{y}^T \mathbf{\Gamma} \mathbf{y}] = \mathbb{E}[(\boldsymbol{\mu} + \mathbf{e})^T \mathbf{\Gamma} (\boldsymbol{\mu} + \mathbf{e})] \\ &= \boldsymbol{\mu}^T \mathbf{A} \boldsymbol{\mu} + \mathbb{E}[\mathbf{e}^T \mathbf{\Gamma} \mathbf{e}], \end{aligned}$$

we can obtain

$$\begin{aligned} Q - \mathbb{E}[Q] &= (\boldsymbol{\mu} + \mathbf{e})^T \mathbf{\Gamma} (\boldsymbol{\mu} + \mathbf{e}) - \boldsymbol{\mu}^T \mathbf{\Gamma} \boldsymbol{\mu} - \mathbb{E}[\mathbf{e}^T \mathbf{\Gamma} \mathbf{e}] \\ &= \mathbf{e}^T \mathbf{\Gamma} \mathbf{e} - \mathbb{E}[\mathbf{e}^T \mathbf{\Gamma} \mathbf{e}] + 2\mathbf{e}^T \mathbf{\Gamma} \boldsymbol{\mu}. \end{aligned}$$

Therefore,

$$\begin{aligned} \text{Var}[Q] &= \mathbb{E}[(Q - \mathbb{E}[Q])^2] \\ &= \mathbb{E}\left[\left(\mathbf{e}^T \mathbf{\Gamma} \mathbf{e} - \mathbb{E}[\mathbf{e}^T \mathbf{\Gamma} \mathbf{e}] + 2\mathbf{e}^T \mathbf{\Gamma} \boldsymbol{\mu}\right)^2\right] \\ &\leq \mathbb{E}\left[2(\mathbf{e}^T \mathbf{\Gamma} \mathbf{e} - \mathbb{E}[\mathbf{e}^T \mathbf{\Gamma} \mathbf{e}])^2 + 8\boldsymbol{\mu}^T \mathbf{\Gamma}^T \mathbf{e} \mathbf{e}^T \mathbf{\Gamma} \boldsymbol{\mu}\right] \\ &= 2\text{Var}[\mathbf{e}^T \mathbf{\Gamma} \mathbf{e}] + 8\boldsymbol{\mu}^T \mathbf{\Gamma}^T \boldsymbol{\Sigma} \mathbf{\Gamma} \boldsymbol{\mu} \end{aligned}$$

□

Appendix A.2. Proof of Theorem 5

Proof. Let $\boldsymbol{\Sigma}_\gamma = \mathbf{A}^T \mathbf{V}_\gamma \mathbf{A}$. Then according to the well-known identity (see for example, Searle et al. [18]), we have

$$\begin{aligned} \mathbf{P}_\gamma &= \mathbf{V}_\gamma^{-1} - \mathbf{V}_\gamma^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}_\gamma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}_\gamma^{-1} \\ &= \mathbf{A} (\mathbf{A}^T \mathbf{V}_\gamma \mathbf{A})^{-1} \mathbf{A}^T \\ &= \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T. \end{aligned}$$

Therefore,

$$\begin{aligned} |\text{tr}(\mathbf{P}_\gamma \mathbf{K})| &= \left| \text{tr}(\boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A}) \right| \\ &= \left| \text{tr}(\boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T (p^{-1} \mathbf{G} \mathbf{W} \mathbf{G}^T \mathbf{A})) \right| \\ &\leq \lambda_{\max}(\boldsymbol{\Sigma}_\gamma^{-1}) p^{-1} \text{tr}(\mathbf{A}^T \mathbf{G} \mathbf{W} \mathbf{G}^T \mathbf{A}). \end{aligned}$$

Under the assumptions of \mathbf{W} , we have $\mathbf{I}_p - \mathbf{W} \geq 0$, and hence

$$\mathbf{A}^T \mathbf{G} \mathbf{G}^T \mathbf{A} \geq \mathbf{A}^T \mathbf{G} \mathbf{W} \mathbf{G}^T \mathbf{A},$$

which further implies that

$$\text{tr}(\mathbf{A}^T \mathbf{G} \mathbf{G}^T \mathbf{A}) \geq \text{tr}(\mathbf{A}^T \mathbf{G} \mathbf{W} \mathbf{G}^T \mathbf{A}).$$

As a consequence,

$$\begin{aligned} |\text{tr}(\mathbf{P}_\gamma \mathbf{K})| &\leq \lambda_{\max}(\boldsymbol{\Sigma}_\gamma^{-1}) p^{-1} \text{tr}(\mathbf{A}^T \mathbf{G} \mathbf{G}^T \mathbf{A}) \\ &= \lambda_{\min}^{-1}(\boldsymbol{\Sigma}_\gamma) p^{-1} \text{tr}(\mathbf{A}^T \mathbf{G} \mathbf{G}^T \mathbf{A}) \\ &\leq p^{-1} \text{tr}(\mathbf{A}^T \mathbf{G} \mathbf{G}^T \mathbf{A}) = \mathcal{O}_p(n), \end{aligned}$$

where the last equality follows from Corollary 1.

Now we define

$$\begin{aligned} \Delta &= \Delta(\gamma) = \mathbf{y}^T \mathbf{B}(\gamma) \mathbf{y} \\ \mathbf{B} &= \mathbf{B}(\gamma) = \frac{\mathbf{P}_\gamma \mathbf{K} \mathbf{P}_\gamma}{\text{tr}(\mathbf{P}_\gamma \mathbf{K})} - \frac{\mathbf{P}_\gamma}{n - q} \\ &= \frac{\mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T}{\text{tr}(\mathbf{P}_\gamma \mathbf{K})} - \frac{\mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T}{n - q} \\ &= \mathbf{A}(\mathbf{C}_1 - \mathbf{C}_2) \mathbf{A}^T, \end{aligned}$$

with

$$\begin{aligned} \mathbf{C}_1 &= \boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} / \text{tr}(\mathbf{P}_\gamma \mathbf{K}) \\ \mathbf{C}_2 &= \boldsymbol{\Sigma}_\gamma^{-1} / (n - q). \end{aligned}$$

Next, we write $\Delta = \mathbb{E}[\Delta | \mathbf{G}] + \Delta - \mathbb{E}[\Delta | \mathbf{G}]$, and we are going to show that $\mathbb{E}[\Delta | \mathbf{G}]$ converges to some constant limit, and $\Delta - \mathbb{E}[\Delta | \mathbf{G}] = o_p(1)$. We first show that $\Delta - \mathbb{E}[\Delta | \mathbf{G}] = o_p(1)$. Based on Lemma 1,

$$\begin{aligned} \text{Var}[\Delta | \mathbf{G}] &= \text{Var}[\mathbf{y}^T \mathbf{A}(\mathbf{C}_1 - \mathbf{C}_2) \mathbf{A}^T \mathbf{y} | \mathbf{G}] \\ &= \text{Var}[\tilde{\mathbf{y}}^T (\mathbf{C}_1 - \mathbf{C}_2) \tilde{\mathbf{y}} | \mathbf{G}] \\ &\leq 2\sigma_\epsilon^2 \text{tr}((\mathbf{C}_1 - \mathbf{C}_2) \boldsymbol{\Sigma}_0 (\mathbf{C}_1 - \mathbf{C}_2) \boldsymbol{\Sigma}_0) \\ &= 2\sigma_\epsilon^2 [\text{tr}((\mathbf{C}_1 \boldsymbol{\Sigma}_0)^2) + \text{tr}((\mathbf{C}_2 \boldsymbol{\Sigma}_0)^2) - 2\text{tr}(\mathbf{C}_1 \boldsymbol{\Sigma}_0 \mathbf{C}_2 \boldsymbol{\Sigma}_0)] \end{aligned} \tag{A1}$$

where $\tilde{\mathbf{y}} = \mathbf{A}^T \mathbf{y} \sim \mathcal{N}_{n-q}(\mathbf{0}, \sigma_\epsilon^2 \mathbf{A}^T \mathbf{V}_0 \mathbf{A}) = \mathcal{N}_{n-q}(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{\Sigma}_0)$ and \mathbf{V}_0 and $\boldsymbol{\Sigma}_0$ are the corresponding matrices of \mathbf{V}_γ and $\boldsymbol{\Sigma}_\gamma$ with the truth γ_0 . We now consider each term in the right hand side of Equation (A1). First note that

$$\text{tr}((\mathbf{C}_1 \boldsymbol{\Sigma}_0)^2) = [\text{tr}(\mathbf{P}_\gamma \mathbf{K})]^{-2} \text{tr}((\boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\Sigma}_0)^2).$$

Since

$$\begin{aligned} \text{tr}(\mathbf{P}_\gamma \mathbf{K}) &= \text{tr}(\mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K}) \\ &= \text{tr}(\boldsymbol{\Sigma}_\gamma^{-1/2} \mathbf{A}^T \mathbf{K} \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1/2}) \\ &\geq \lambda_{\min}(\mathbf{K}) \text{tr}(\boldsymbol{\Sigma}_\gamma^{-1}), \end{aligned}$$

and it follows from the assumption on \mathbf{W} and Corollary 1,

$$\begin{aligned} \lambda_{\min}(\mathbf{K}) &= p^{-1}\lambda_{\min}(\mathbf{G}\mathbf{W}\mathbf{G}^T) \\ &\geq \lambda_{\min}(\mathbf{W})\lambda_{\min}(p^{-1}\mathbf{G}\mathbf{G}^T) \\ &\geq \delta\lambda_{\min}(p^{-1}\mathbf{G}\mathbf{G}^T) \\ &\rightarrow \delta b_-(\tau) > 0 \text{ a.s.,} \end{aligned}$$

$$\begin{aligned} \text{tr}(\boldsymbol{\Sigma}_\gamma^{-1}) &= \text{tr}((\mathbf{I}_{n-q} + \gamma\mathbf{A}^T\mathbf{K}\mathbf{A})^{-1}) \\ &= \sum_{i=1}^{n-q} \frac{1}{1 + \lambda_i(\gamma\mathbf{A}^T\mathbf{K}\mathbf{A})} \\ &\geq \frac{n-q}{1 + \gamma\lambda_{\max}(\mathbf{A}^T\mathbf{K}\mathbf{A})} = \mathcal{O}_p(n), \end{aligned}$$

we know that $\text{tr}(\mathbf{P}_\gamma\mathbf{K}) \asymp_p n$. On the other hand,

$$\text{tr}\left(\left(\boldsymbol{\Sigma}_\gamma^{-1}\mathbf{A}^T\mathbf{K}\mathbf{A}\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_0\right)^2\right) = \text{tr}\left(\left(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_\gamma^{-1}\mathbf{A}^T\mathbf{K}\mathbf{A}\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_0^{1/2}\right)^2\right)$$

and since $\mathbf{I}_p \geq \mathbf{W}$, it follows that

$$\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_\gamma^{-1}\mathbf{A}^T\mathbf{K}\mathbf{A}\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_0^{1/2} \leq \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_\gamma^{-1}\mathbf{A}^T p^{-1}\mathbf{G}\mathbf{G}^T\mathbf{A}\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_0^{1/2}$$

According to Corollary 1

$$\text{tr}\left(\left(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_\gamma^{-1}\mathbf{A}^T\mathbf{K}\mathbf{A}\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_0^{1/2}\right)^2\right) \leq \text{tr}\left(\left(\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\Sigma}_\gamma^{-1}\mathbf{A}^T p^{-1}\mathbf{G}\mathbf{G}^T\mathbf{A}\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_0^{1/2}\right)^2\right) = \mathcal{O}_p(n),$$

which implies that $\text{tr}((\mathbf{C}_1\boldsymbol{\Sigma}_0)^2) = \mathcal{O}_p(n^{-1})$.

For the second term, it is clear that

$$\text{tr}\left((\mathbf{C}_2\boldsymbol{\Sigma}_0)^2\right) = (n-q)^{-2}\text{tr}\left((\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_0)^2\right) \leq (n-q)^{-1}\lambda_{\max}^2\left(\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_0\right).$$

Since

$$\begin{aligned} \lambda_{\max}\left(\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_0\right) &\leq \lambda_{\max}(\boldsymbol{\Sigma}_\gamma^{-1})\lambda_{\max}(\boldsymbol{\Sigma}_0) \\ &= [\lambda_{\min}(\boldsymbol{\Sigma}_\gamma)]^{-1}\lambda_{\max}(\mathbf{I}_{n-q} + \gamma_0\mathbf{A}^T\mathbf{K}\mathbf{A}) \\ &= \frac{1 + \gamma_0\lambda_{\max}(\mathbf{A}^T\mathbf{K}\mathbf{A})}{1 + \gamma\lambda_{\max}(\mathbf{A}^T\mathbf{K}\mathbf{A})} \\ &\leq \frac{\gamma_0}{\gamma} \vee 1, \end{aligned}$$

we can know that $\text{tr}((\mathbf{C}_2\boldsymbol{\Sigma}_0)^2) = \mathcal{O}_p(n^{-1})$.

Similarly for the third term, we have

$$\begin{aligned} \text{tr}(\mathbf{C}_1\boldsymbol{\Sigma}_0\mathbf{C}_2\boldsymbol{\Sigma}_0) &= [(n-q)\text{tr}(\mathbf{P}_\gamma\mathbf{K})]^{-1}\text{tr}\left(\boldsymbol{\Sigma}_\gamma^{-1}\mathbf{A}^T\mathbf{K}\mathbf{A}\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_0\right) \\ &\leq \lambda_{\max}\left(\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_0\right)[(n-q)\text{tr}(\mathbf{P}_\gamma\mathbf{K})]^{-1}\text{tr}\left(\boldsymbol{\Sigma}_\gamma^{-1}\mathbf{A}^T\mathbf{K}\mathbf{A}\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_0\right) \\ &\leq \lambda_{\max}\left(\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_0\right)[(n-q)\text{tr}(\mathbf{P}_\gamma\mathbf{K})]^{-1}\text{tr}\left(\boldsymbol{\Sigma}_\gamma^{-1}\mathbf{A}^T p^{-1}\mathbf{G}\mathbf{G}^T\mathbf{A}\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\Sigma}_0\right) \\ &= \mathcal{O}_p(n^{-2} \cdot n) = \mathcal{O}_p(n^{-1}). \end{aligned}$$

Therefore, $\text{Var}[\Delta|\mathbf{G}] = \mathcal{O}_p(n^{-1})$ and by Chebyshev’s inequality, for any $t > 0$,

$$\mathbb{P}(|\Delta - \mathbb{E}[\Delta|\mathbf{G}]| > t|\mathbf{G}) \leq \frac{\text{Var}[\Delta|\mathbf{G}]}{t^2} \xrightarrow{p} 0, \text{ as } n \rightarrow \infty.$$

It then follows from the Dominated Convergence Theorem that

$$\mathbb{P}(|\Delta - \mathbb{E}[\Delta|\mathbf{G}]| > t) \rightarrow 0, \quad \forall t > 0,$$

which implies that $\Delta - \mathbb{E}[\Delta|\mathbf{G}] = o_p(1)$.

Next, let us focus on $\mathbb{E}[\Delta|\mathbf{G}]$. It is easy to see that

$$\begin{aligned} \mathbb{E}[\Delta|\mathbf{G}] &= \mathbb{E}[\mathbf{y}^T \mathbf{A}(\mathbf{C}_1 - \mathbf{C}_2)\mathbf{A}^T \mathbf{y}|\mathbf{G}] \\ &= \sigma_\epsilon^2 \text{tr}((\mathbf{C}_1 - \mathbf{C}_2)\boldsymbol{\Sigma}_0) \\ &= \sigma_\epsilon^2 \left[\frac{\text{tr}(\boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\Sigma}_0)}{\text{tr}(\mathbf{P}_\gamma \mathbf{K})} - \frac{\text{tr}(\boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\Sigma}_0)}{n - q} \right]. \end{aligned} \tag{A2}$$

Since

$$\begin{aligned} \frac{\text{tr}(\boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\Sigma}_0)}{\text{tr}(\mathbf{P}_\gamma \mathbf{K})} &\leq \frac{\lambda_{\max}(\boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\Sigma}_0) \text{tr}(\boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A})}{\text{tr}(\mathbf{P}_\gamma \mathbf{K})} \\ &= \lambda_{\max}(\boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\Sigma}_0) \leq \frac{\gamma_0}{\gamma} \vee 1, \end{aligned}$$

it follows from Bounded Convergence Theorem that

$$\mathbb{E} \left[\frac{\text{tr}(\boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\Sigma}_0)}{\text{tr}(\mathbf{P}_\gamma \mathbf{K})} \right] \rightarrow \mathbb{E} \left[\lim_{n \rightarrow \infty} \frac{\text{tr}(\boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\Sigma}_0)}{\text{tr}(\mathbf{P}_\gamma \mathbf{K})} \right], \text{ as } n \rightarrow \infty.$$

Moreover, note that

$$\mathbb{E} \left[\text{tr}(\boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\Sigma}_0) \right] = \mathbb{E} \left[\text{tr}(\boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1}) \right] + \gamma_0 \mathbb{E} \left[\text{tr}(\boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A}) \right],$$

and according to Theorem 2, the ESD of $\frac{p}{n} \mathbf{A}^T \mathbf{K} \mathbf{A}$ converges in distribution almost surely to a nonrandom distribution F . Let $\lambda_1, \dots, \lambda_{n-q}$ be the eigenvalues of $(p/n) \mathbf{A}^T \mathbf{K} \mathbf{A}$, then

$$\begin{aligned} \frac{\text{tr}(\boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1})}{n - q} &= \frac{1}{n - q} \sum_{i=1}^{n-q} \frac{(n/p)\lambda_i}{(1 + (n/p)\gamma\lambda_i)^2} \rightarrow \int \frac{\tau x}{(1 + \tau\gamma x)^2} dF(x) \text{ a.s.} \\ \frac{\text{tr}(\mathbf{P}_\gamma \mathbf{K})}{n - q} &= \frac{1}{n - q} \sum_{i=1}^{n-q} \frac{(n/p)\lambda_i}{1 + (n/p)\gamma\lambda_i} \rightarrow \int \frac{\tau x}{1 + \tau\gamma x} dF(x) \text{ a.s.} \\ \frac{\text{tr}((\boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A})^2)}{n - q} &= \frac{1}{n - q} \sum_{i=1}^{n-q} \frac{(n/p)^2 \lambda_i^2}{(1 + \gamma(n/p)\lambda_i)^2} \rightarrow \int \frac{\tau^2 x^2}{(1 + \tau\gamma x)^2} dF(x) \text{ a.s..} \end{aligned}$$

For notation simplicity, we define

$$\begin{aligned} \phi_1 &= \int \frac{\tau x}{(1 + \tau\gamma x)^2} dF(x) \\ \phi_2 &= \int \frac{\tau^2 x^2}{(1 + \tau\gamma x)^2} dF(x). \end{aligned}$$

Similarly, since $(n - q)^{-1} \text{tr}(\Sigma_\gamma^{-1} \Sigma_0) \leq \lambda_{\max}(\Sigma_\gamma^{-1} \Sigma_0) \leq \gamma_0 / \gamma \vee 1$, it follows from Bounded Convergence Theorem that

$$\mathbb{E} \left[\frac{\text{tr}(\Sigma_\gamma^{-1} \Sigma_0)}{n - q} \right] \rightarrow \mathbb{E} \left[\lim_{n \rightarrow \infty} \frac{\text{tr}(\Sigma_\gamma^{-1} \Sigma_0)}{n - q} \right], \quad \text{as } n \rightarrow \infty.$$

Moreover, note that

$$\text{tr}(\Sigma_\gamma^{-1} \Sigma_0) = \text{tr}(\Sigma_\gamma^{-1}) + \gamma_0 \text{tr}(\Sigma_\gamma^{-1} A^T K A),$$

and by Theorem 2,

$$\frac{\text{tr}(\Sigma_\gamma^{-1})}{n - q} = \frac{1}{n - q} \sum_{i=1}^{n-q} \frac{1}{1 + (n/p)\gamma\lambda_i} \rightarrow \int \frac{1}{1 + \tau\gamma x} dF(x) \text{ a.s..}$$

Let

$$\begin{aligned} \psi_1 &= \int \frac{1}{1 + \tau\gamma x} dF(x) \\ \psi_2 &= \int \frac{\tau x}{1 + \tau\gamma x} dF(x). \end{aligned}$$

Then we have

$$\mathbb{E}[\Delta | \mathbf{G}] = \sigma_\epsilon^2 \left(\frac{\phi_1}{\psi_2} + \gamma_0 \frac{\phi_2}{\psi_2} - \psi_1 - \gamma_0 \psi_2 \right) = \sigma_\epsilon^2 \left(\frac{\phi_1}{\psi_2} - \psi_1 \right) \left(1 + \gamma_0 \frac{\phi_2 / \psi_2 - \psi_2}{\phi_1 / \psi_2 - \psi_1} \right)$$

Using Fubini Theorem and some tedious algebra, it follows that

$$\mathbb{E}[\Delta | \mathbf{G}] = \sigma_\epsilon^2 \left(\frac{\gamma_0}{\gamma} - 1 \right) \left(\psi_1 - \frac{\phi_1}{\psi_2} \right).$$

Note that since $f(x) = (1 + \tau x)^{-1}$ is non-increasing and $g(x) = (1 + \tau x)^{-1} \tau x$ is non-decreasing, we have for any $x, y \in \mathbb{R}$

$$(f(x) - f(y))(g(x) - g(y)) \leq 0,$$

and hence

$$\begin{aligned} 0 &\geq \iint (f(x) - f(y))(g(x) - g(y)) dF(x) dF(y) \\ &= 2 \int f(x)g(x) dF(x) - 2 \int f(x) dF(x) \int g(x) dF(x), \end{aligned}$$

which implies that

$$\int \frac{1}{1 + \tau x} dF(x) \int \frac{\tau x}{1 + \tau x} dF(x) \geq \int \frac{\tau x}{(1 + \tau x)^2} dF(x),$$

that is, $\psi_1 \psi_2 \geq \phi_1$. Therefore, $\Delta \xrightarrow{p} \sigma_\epsilon^2 \left(\frac{\gamma_0}{\gamma} - 1 \right) \left(\psi_1 - \frac{\phi_1}{\psi_2} \right)$, which is a constant limit, and the limit is > 0 , $= 0$, and < 0 when $\gamma < \gamma_0$, $\gamma = \gamma_0$, and $\gamma > \gamma_0$, respectively. This proves the desired result as $\hat{\gamma}$ is the solution to $\Delta = 0$, and hence $\hat{\gamma}_n \rightarrow \gamma_0$.

Next, we prove the second part of this theorem. Let

$$\begin{aligned} \hat{\sigma}_\epsilon^2 &= \frac{1}{n-q} \mathbf{y}^T \mathbf{P}_{\hat{\gamma}_n} \mathbf{y} \\ &= \mathbb{E} \left[\frac{1}{n-q} \mathbf{y}^T \mathbf{P}_{\gamma_0} \mathbf{y} \right] + \left(\frac{1}{n-q} \mathbf{y}^T \mathbf{P}_{\hat{\gamma}_n} \mathbf{y} - \frac{1}{n-q} \mathbf{y}^T \mathbf{P}_{\gamma_0} \mathbf{y} \right) + \left(\frac{1}{n-q} \mathbf{y}^T \mathbf{P}_{\gamma_0} \mathbf{y} - \mathbb{E} \left[\frac{1}{n-q} \mathbf{y}^T \mathbf{P}_{\gamma_0} \mathbf{y} \right] \right). \end{aligned}$$

First note that $\mathbf{X}^T \mathbf{P}_{\gamma_0} \mathbf{X} = \mathbf{X}^T \mathbf{V}_{\gamma_0}^{-1} \mathbf{X} - \mathbf{X}^T \mathbf{V}_{\gamma_0}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}_{\gamma_0}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}_{\gamma_0}^{-1} \mathbf{X} = \mathbf{0}$, so we have

$$\begin{aligned} \mathbb{E} \left[\frac{1}{n-q} \mathbf{y}^T \mathbf{P}_{\gamma_0} \mathbf{y} \right] &= \frac{1}{n-q} \left[\text{tr}(\mathbf{P}_{\gamma_0} \sigma_\epsilon^2 \mathbf{V}_{\gamma_0}) + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{P}_{\gamma_0} \mathbf{X} \boldsymbol{\beta} \right] \\ &= \frac{1}{n-q} \left[\sigma_\epsilon^2 \text{tr}(\mathbf{I}_n - \mathbf{V}_{\gamma_0}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}_{\gamma_0} \mathbf{X})^{-1} \mathbf{X}^T) \right] \\ &= \frac{1}{n-q} (n-q) \sigma_\epsilon^2 = \sigma_\epsilon^2. \end{aligned}$$

Next, since $\hat{\gamma}_n \xrightarrow{p} \gamma_0$, by continuous mapping theorem, we have

$$\left\| \frac{1}{n-q} \mathbf{P}_{\hat{\gamma}_n} - \frac{1}{n-q} \mathbf{P}_{\gamma_0} \right\| \leq \frac{1}{n-q} \cdot n \max_{1 \leq i \leq n} (\mathbf{P}_{\hat{\gamma}_n, ii} - \mathbf{P}_{\gamma_0, ii}) = o_p(1),$$

which implies that $\frac{1}{n-q} \mathbf{y}^T \mathbf{P}_{\hat{\gamma}_n} \mathbf{y} - \frac{1}{n-q} \mathbf{y}^T \mathbf{P}_{\gamma_0} \mathbf{y} = o_p(1)$. Finally, since $\mathbf{y} = \sigma_\epsilon \mathbf{V}_{\gamma_0}^{1/2} \mathbf{z}$ where $\mathbf{z} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{I}_n)$, by the famous Hanson Wright inequality, for any $t > 0$,

$$\begin{aligned} &\mathbb{P} \left(\left| \frac{1}{n-q} \mathbf{y}^T \mathbf{P}_{\gamma_0} \mathbf{y} - \mathbb{E} \left[\frac{1}{n-q} \mathbf{y}^T \mathbf{P}_{\gamma_0} \mathbf{y} \right] \right| > t \right) \\ &= \mathbb{P} \left(\left| \mathbf{z}^T \mathbf{V}_{\gamma_0}^{1/2} \mathbf{P}_{\gamma_0} \mathbf{V}_{\gamma_0}^{1/2} \mathbf{z} - \mathbb{E} \left[\mathbf{z}^T \mathbf{V}_{\gamma_0}^{1/2} \mathbf{P}_{\gamma_0} \mathbf{V}_{\gamma_0}^{1/2} \mathbf{z} \right] \right| > (n-q)t / \sigma_\epsilon^2 \right) \\ &\lesssim \exp \left\{ -c \min \left(\frac{(n-q)^2 t^2}{\sigma_\epsilon^4 \left\| \mathbf{V}_{\gamma_0}^{1/2} \mathbf{P}_{\gamma_0} \mathbf{V}_{\gamma_0}^{1/2} \right\|_F^2}, \frac{(n-q)t}{\sigma_\epsilon^2 \left\| \mathbf{V}_{\gamma_0}^{1/2} \mathbf{P}_{\gamma_0} \mathbf{V}_{\gamma_0}^{1/2} \right\|} \right) \right\}. \end{aligned}$$

Note that

$$\begin{aligned} \left\| \mathbf{V}_{\gamma_0}^{1/2} \mathbf{P}_{\gamma_0} \mathbf{V}_{\gamma_0}^{1/2} \right\| &\leq \left\| \mathbf{I}_n - \mathbf{V}_{\gamma_0}^{-1/2} \mathbf{X} (\mathbf{X}^T \mathbf{V}_{\gamma_0}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}_{\gamma_0}^{-1/2} \right\| \leq 1, \\ \left\| \mathbf{V}_{\gamma_0}^{1/2} \mathbf{P}_{\gamma_0} \mathbf{V}_{\gamma_0}^{1/2} \right\|_F^2 &\leq n \left\| \mathbf{V}_{\gamma_0}^{1/2} \mathbf{P}_{\gamma_0} \mathbf{V}_{\gamma_0}^{1/2} \right\| \leq n, \end{aligned}$$

so we obtain

$$\mathbb{P} \left(\left| \frac{1}{n-q} \mathbf{y}^T \mathbf{P}_{\gamma_0} \mathbf{y} - \mathbb{E} \left[\frac{1}{n-q} \mathbf{y}^T \mathbf{P}_{\gamma_0} \mathbf{y} \right] \right| > t \right) \lesssim \exp \{ -\mathcal{O}(n) \} \rightarrow 0,$$

which implies that $\hat{\sigma}_\epsilon^2 = \sigma_\epsilon^2 + o_p(1)$. \square

Appendix A.3. Proof of Corollary 2

Proof. For simplicity, let $\mathbf{A} = (\mathbf{G} - \bar{\mathbf{g}} \otimes \mathbf{1}_n) \mathbf{W}^{1/2}$ and $\mathbf{B} = \tilde{\mathbf{G}} \mathbf{W}^{1/2}$. Then $\tilde{\mathbf{K}} = p^{-1} \mathbf{B} \mathbf{B}^T$. Let $\mathbf{K}' = p^{-1} \mathbf{A} \mathbf{A}^T$. Now note that

$$\begin{aligned} \|\mathbf{A}\|_F^2 &= \text{tr}(\mathbf{A} \mathbf{A}^T) \\ &= \text{tr} \left((\mathbf{G} - \bar{\mathbf{g}} \otimes \mathbf{1}_n) \mathbf{W} (\mathbf{G} - \bar{\mathbf{g}} \otimes \mathbf{1}_n)^T \right) \\ &= \text{tr}(\mathbf{G} \mathbf{W} \mathbf{G}^T) - 2 \text{tr} \left(\mathbf{G} \mathbf{W} (\bar{\mathbf{g}} \otimes \mathbf{1}_n)^T \right) + \text{tr} \left((\bar{\mathbf{g}} \otimes \mathbf{1}_n) \mathbf{W} (\bar{\mathbf{g}} \otimes \mathbf{1}_n)^T \right). \end{aligned}$$

Since $I_p \geq W$, we have $p^{-1}GG^T \geq p^{-1}GWG^T$ and

$$\begin{aligned}
 p^{-1}\text{tr}(GWG^T) &\leq p^{-1}\text{tr}(GG^T) = \mathcal{O}_p(n). \\
 p^{-1}\text{tr}\left(GW(\bar{g} \otimes \mathbf{1}_n)^T\right) &= p^{-1}\text{tr}\left(GW\left(\left(\frac{1}{n}\mathbf{1}_n^T G\right) \otimes \mathbf{1}_n\right)^T\right) \\
 &= p^{-1}\text{tr}\left(\left(\frac{1}{n}GWG^T\mathbf{1}_n\right) \otimes \mathbf{1}_n^T\right) \\
 &= n^{-1}\mathbf{1}_n^T p^{-1}GWG^T\mathbf{1}_n \\
 &\leq \lambda_{\max}\left(p^{-1}GWG^T\right) \\
 &\leq \lambda_{\max}\left(p^{-1}GG^T\right) = \mathcal{O}_p(1). \\
 p^{-1}\text{tr}\left((\bar{g} \otimes \mathbf{1}_n)W(\bar{g} \otimes \mathbf{1}_n)^T\right) &= p^{-1}\text{tr}\left[\left(\left(\frac{1}{n}\mathbf{1}_n^T G\right) \otimes \mathbf{1}_n\right)W\left(\left(\frac{1}{n}\mathbf{1}_n^T G\right) \otimes \mathbf{1}_n\right)^T\right] \\
 &= p^{-1}\text{tr}\left[\left(\frac{1}{n^2}\mathbf{1}_n^T GWG^T\mathbf{1}_n\right) \otimes (\mathbf{1}_n\mathbf{1}_n^T)\right] \\
 &= n^{-1}\mathbf{1}_n^T p^{-1}GWG^T\mathbf{1}_n \\
 &\leq \lambda_{\max}\left(p^{-1}GWG^T\right) = \mathcal{O}_p(1).
 \end{aligned}$$

Combining all of these yields $p^{-1}\|A\|_F^2 = \mathcal{O}_p(n)$. On the other hand,

$$\begin{aligned}
 \|B\|_F^2 &= \text{tr}(BB^T) = \text{tr}(\tilde{G}W\tilde{G}^T) \\
 &= \text{tr}\left((G - \bar{g} \otimes \mathbf{1}_n)D_s^{-1}WD_s^{-1}(G - \bar{g} \otimes \mathbf{1}_n)^T\right) \\
 &\leq \lambda_{\max}(D_s^{-1}WD_s^{-1})\|G - \bar{g} \otimes \mathbf{1}_n\|_F^2 \\
 &\leq \frac{\max_{1 \leq i \leq p} w_i}{\min_{1 \leq j \leq p} s_j^2} \|G - \bar{g} \otimes \mathbf{1}_n\|_F^2.
 \end{aligned}$$

By Lemma 2.6 in Jiang et al. [11],

$$\max_{1 \leq j \leq p} |s_j^2 - 1| \rightarrow 0 \quad \text{a.s.},$$

which implies that $(\min_{1 \leq j \leq p} s_j^2)^{-1} = \mathcal{O}_p(1)$, and hence $\|B\|_F^2 = \mathcal{O}_p(n)$. Finally, since

$$\begin{aligned}
 \|A - B\|_F^2 &= \text{tr}\left((A - B)(A - B)^T\right) \\
 &= \text{tr}\left((G - \bar{g} \otimes \mathbf{1}_n)(I_p - D_s^{-1})W(I_p - D_s^{-1})(G - \bar{g} \otimes \mathbf{1}_n)^T\right) \\
 &\leq \lambda_{\max}\left((I_p - D_s^{-1})W(I_p - D_s^{-1})\right)\|G - \bar{g} \otimes \mathbf{1}_n\|_F^2,
 \end{aligned}$$

and by Corollary 2.3 in Jiang et al. [11],

$$\begin{aligned}
 \lambda_{\max}\left((I_p - D_s^{-1})W(I_p - D_s^{-1})\right) &\leq \left(\max_{1 \leq i \leq p} w_i\right)\lambda_{\max}\left((I_p - D_s^{-1})^2\right) \\
 &\leq \lambda_{\max}\left((I_p - D_s^{-1})^2\right) \\
 &= o_p(1)
 \end{aligned}$$

Therefore, we can bound the Levy’s distance between $F^{K'}$ and $F^{\tilde{K}}$ by Corollary A.42 in Bai and Silverstein [19]:

$$\begin{aligned} L^A(F^{K'}, F^{\tilde{K}}) &\leq 2n^{-2}p^{-2} \left(\|A\|_F^2 + \|B\|_F^2 \right) \|A - B\|_F^2 \\ &= n^{-2} (\mathcal{O}_p(n) + \mathcal{O}_p(n)) \cdot o_p(n) \\ &= o_p(1). \end{aligned}$$

Hence, the ESD of \tilde{K} converges a.s. in distribution to the ESD of K . On the other hand,

$$\begin{aligned} \left\| F^{K'} - F^{GWG^T} \right\| &\leq n^{-1} \text{rank}(A - GW^{1/2}) \\ &= n^{-1} \text{rank}\left((\bar{g} \otimes \mathbf{1}_n) W^{1/2} \right) \\ &\leq n^{-1} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Thus, the ESD of \tilde{K} converges a.s. in distribution to the ESD of $p^{-1}GWG^T$, and the desired results in the corollary follows Theorem 5. \square

Appendix A.4. Proof of Theorem 6

Proof. We follow the same framework as in the proof of Theorem 5. Let

$$\tilde{M} = f'(0) \frac{GG^T}{p} + v_p \mathbf{I}_n.$$

Based on Theorem A.43 in Bai and Silverstein [19], we have

$$\left\| F_n^M - F_n^{\tilde{M}} \right\| \leq \frac{1}{n} \text{rank}(M - \tilde{M}) = \frac{1}{n} \rightarrow 0, \text{ as } n \rightarrow \infty.$$

The matrix \tilde{M} will play a vital part in the remainder of the proof as it is easy to see that the LSD of \tilde{M} converges in distribution a.s. to some nonrandom distribution function by Theorem 2. On the other hand, since the elements in G are i.i.d. with $\mathbb{E}[G_{ij}] = 0$ and $\mathbb{E}[G_{ij}^2] = 1$, it is easy to see that the elements in $A^T G$ are still i.i.d. with a mean of 0 and unit variance. Moreover, since

$$A^T \tilde{M} A = f'(0) \frac{A^T G G^T A}{p} + v_p \mathbf{I}_{n-q},$$

we know that the LSD of $A^T \tilde{M} A$ also converges in distribution a.s. to some nonrandom distribution function, and we denote such distribution by F .

It can be seen from the proof of Theorem 5 that one major part of showing the result is to show that $\text{tr}((C_1 \Sigma_0)^2) = \mathcal{O}_p(n^{-1})$, $\text{tr}((C_2 \Sigma_0)^2) = \mathcal{O}_p(n^{-1})$ and $\text{tr}(C_1 \Sigma_0 C_2 \Sigma_0) = \mathcal{O}_p(n^{-1})$. The proof of $\text{tr}((C_2 \Sigma_0)^2) = \mathcal{O}_p(1)$ is exactly the same arguments as in the proof of Theorem 5, so we focus on the other two quantities.

As a consequence of Theorem A.45 in Bai and Silverstein [19] and Theorem 3, we have

$$L(F_{n-q}^{A^T K A}, F_{n-q}^{A^T M A}) \leq \|M - K\| = o_p(1).$$

Moreover, it follows from Theorem A.43 in Bai and Silverstein [19] that

$$\left\| F_{n-q}^{A^T M A} - F_{n-q}^{A^T \tilde{M} A} \right\| \leq \frac{1}{n-q} \text{rank}(A^T (M - \tilde{M}) A) \leq \frac{1}{n-q},$$

which implies that

$$L(F_{n-q}^{A^T KA}, F_{n-q}^{A^T \bar{M}A}) \leq L(F_{n-q}^{A^T KA}, F_{n-q}^{A^T MA}) + \left\| F_{n-q}^{A^T MA} - F_{n-q}^{A^T \bar{M}A} \right\| = o_p(1).$$

Therefore,

$$\begin{aligned} \frac{1}{n-q} \text{tr}(P_\gamma K) &= \frac{1}{n-q} \text{tr}(\Sigma_\gamma^{-1} A^T KA) = \frac{1}{n-q} \sum_{i=1}^{n-q} \frac{\lambda_i(A^T KA)}{1 + \gamma \lambda_i(A^T KA)} \\ &= \int \frac{x}{1 + \gamma x} dF_{n-q}^{A^T KA}(x) \\ &= \int \frac{x}{1 + \gamma x} d(F_{n-q}^{A^T KA}(x) - F_{n-q}^{A^T \bar{M}A}(x)) + \int \frac{x}{1 + \gamma x} dF_{n-q}^{A^T \bar{M}A}(x) \\ &\rightarrow 0 + \int \frac{x}{1 + \gamma x} dF(x) \text{ a.s.,} \end{aligned} \tag{A3}$$

which implies that $\text{tr}(P_\gamma K) \asymp_p n$. Similarly, note that

$$\begin{aligned} &\frac{1}{n-q} \text{tr}((\Sigma_\gamma^{-1} A^T KA \Sigma_\gamma^{-1} \Sigma_0)^2) \\ &= \frac{1}{n-q} \sum_{i=1}^{n-q} \frac{(1 + \gamma_0 \lambda_i(A^T KA))^2 \lambda_i^2(A^T KA)}{(1 + \lambda_i(A^T KA))^4} \\ &= \int \frac{(1 + \gamma_0 x)^2 x^2}{(1 + \gamma x)^4} dF_{n-q}^{A^T KA}(x) \\ &= \int \frac{(1 + \gamma_0 x)^2 x^2}{(1 + \gamma x)^4} d(F_{n-q}^{A^T KA}(x) - F_{n-q}^{A^T \bar{M}A}(x)) + \int \frac{(1 + \gamma_0 x)^2 x^2}{(1 + \gamma x)^4} dF_{n-q}^{A^T \bar{M}A}(x) \\ &\rightarrow 0 + \int \frac{(1 + \gamma_0 x)^2 x^2}{(1 + \gamma x)^4} dF(x) \text{ a.s.,} \end{aligned}$$

which implies that $\text{tr}((\Sigma_\gamma^{-1} A^T KA \Sigma_\gamma^{-1} \Sigma_0)^2) = \mathcal{O}_p(n)$ and hence

$$\text{tr}((C_1 \Sigma_0)^2) = [\text{tr}(P_\gamma K)]^{-2} \text{tr}((\Sigma_\gamma^{-1} A^T KA \Sigma_\gamma^{-1} \Sigma_0)^2) = \mathcal{O}_p(n^{-1}).$$

For $\text{tr}(C_1 \Sigma_0 C_2 \Sigma_0)$, note that

$$\begin{aligned} &\frac{1}{n-q} \text{tr}(\Sigma_\gamma^{-1} A^T KA \Sigma_\gamma^{-1} \Sigma_0) \\ &= \frac{1}{n-q} \sum_{i=1}^{n-q} \frac{(1 + \gamma_0 \lambda_i(A^T KA)) \lambda_i(A^T KA)}{(1 + \gamma \lambda_i(A^T KA))^2} \\ &= \int \frac{(1 + \gamma_0 x)x}{(1 + \gamma x)^2} dF_{n-q}^{A^T KA}(x) \\ &= \int \frac{(1 + \gamma_0 x)x}{(1 + \gamma x)^2} d(F_{n-q}^{A^T KA}(x) - F_{n-q}^{A^T \bar{M}A}(x)) + \int \frac{(1 + \gamma_0 x)x}{(1 + \gamma x)^2} dF_{n-q}^{A^T \bar{M}A}(x) \\ &\rightarrow 0 + \int \frac{(1 + \gamma_0 x)x}{(1 + \gamma x)^2} dF(x) \text{ a.s.,} \end{aligned} \tag{A4}$$

which implies that $\text{tr}(\Sigma_\gamma^{-1} A^T KA \Sigma_\gamma^{-1} \Sigma_0) = \mathcal{O}_p(n)$, and hence

$$\begin{aligned} \text{tr}(C_1 \Sigma_0 C_2 \Sigma_0) &= [(n-q) \text{tr}(P_\gamma K)]^{-1} \text{tr}(\Sigma_\gamma^{-1} A^T KA \Sigma_\gamma^{-1} \Sigma_0 \Sigma_\gamma^{-1} \Sigma_0) \\ &\leq \lambda_{\max}(\Sigma_\gamma^{-1} \Sigma_0) [(n-q) \text{tr}(P_\gamma K)]^{-1} \text{tr}(\Sigma_\gamma^{-1} A^T KA \Sigma_\gamma^{-1} \Sigma_0) \\ &= \mathcal{O}_p(n^{-2} \cdot n) = \mathcal{O}_p(n^{-1}). \end{aligned}$$

Therefore, $\text{Var}[\Delta|\mathbf{G}] = \mathcal{O}_p(n^{-1})$ and as in the proof of Theorem 5, it follows from Chebyshev’s inequality and Dominated Convergence Theorem that $\Delta - \mathbb{E}[\Delta|\mathbf{G}] = o_p(1)$.

Similar to the proof of Theorem 5, we now focus on $\mathbb{E}[\Delta|\mathbf{G}]$. Recall from Equation (A2) that

$$\begin{aligned} \mathbb{E}[\Delta|\mathbf{G}] &= \sigma_\varepsilon^2 \left[\frac{\text{tr}(\boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\Sigma}_0)}{\text{tr}(\mathbf{P}_\gamma \mathbf{K})} - \frac{\text{tr}(\boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\Sigma}_0)}{n - q} \right] \\ &= \sigma_\varepsilon^2 \left[\frac{\text{tr}(\boldsymbol{\Sigma}_\gamma^{-1} \mathbf{A}^T \mathbf{K} \mathbf{A} \boldsymbol{\Sigma}_\gamma^{-1} \boldsymbol{\Sigma}_0)}{\text{tr}(\mathbf{P}_\gamma \mathbf{K})} - \frac{\text{tr}(\boldsymbol{\Sigma}_\gamma^{-1}) + \gamma_0 \text{tr}(\mathbf{P}_\gamma \mathbf{K})}{n - q} \right]. \end{aligned}$$

Note that

$$\begin{aligned} \frac{1}{n - q} \text{tr}(\boldsymbol{\Sigma}_\gamma^{-1}) &= \frac{1}{n - q} \sum_{i=1}^{n-q} \frac{1}{1 + \gamma \lambda_i(\mathbf{A}^T \mathbf{K} \mathbf{A})} = \int \frac{1}{1 + \gamma x} dF_{n-q}^{\mathbf{A}^T \mathbf{K} \mathbf{A}}(x) \\ &= \int \frac{1}{1 + \gamma x} d(F_{n-q}^{\mathbf{A}^T \mathbf{K} \mathbf{A}}(x) - F_{n-q}^{\mathbf{A}^T \tilde{\mathbf{M}} \mathbf{A}}(x)) + \int \frac{1}{1 + \gamma x} dF_{n-q}^{\mathbf{A}^T \tilde{\mathbf{M}} \mathbf{A}}(x) \\ &\rightarrow 0 + \int \frac{1}{1 + \gamma x} dF(x). \end{aligned}$$

Combined with Equations (A3) and (A4), it is easy to see that $\mathbb{E}[\Delta|\mathcal{G}]$ converges to a constant limit. The remaining part of the proof follows the same statements as in the proof of Theorem 6, so it is omitted. \square

Appendix A.5. Proof of Theorem 7

Proof. Let \mathbf{M} and v_p be as defined in Equation (10). Now let

$$\tilde{\mathbf{M}} = -2f'(\tau) \frac{\mathbf{G} \mathbf{G}^T}{p} + v_p \mathbf{I}_n.$$

By Theorem A.43 in Bai and Silverstein [19] and the triangle inequality for matrix rank, we have

$$\begin{aligned} \|F_n^{\mathbf{M}} - F_n^{\tilde{\mathbf{M}}}\| &\leq \frac{1}{n} \text{rank}(\mathbf{M} - \tilde{\mathbf{M}}) \\ &= \frac{1}{n} \text{rank} \left(f(\tau) \mathbf{1}_n \mathbf{1}_n^T + f'(\tau) [\mathbf{1}_n \boldsymbol{\psi}^T + \boldsymbol{\psi} \mathbf{1}_n^T] + \right. \\ &\quad \left. \frac{f''(\tau)}{2} \left[\mathbf{1}_n (\boldsymbol{\psi} \circ \boldsymbol{\psi})^T + (\boldsymbol{\psi} \circ \boldsymbol{\psi}) \mathbf{1}_n^T + 2\boldsymbol{\psi} \boldsymbol{\psi}^T + 4 \frac{\text{tr}(\boldsymbol{\Sigma}_p^2)}{p^2} \mathbf{1}_n \mathbf{1}_n^T \right] \right) \\ &\leq \frac{7}{n} \rightarrow 0, \quad \text{as } n \rightarrow \infty. \end{aligned}$$

On the other hand, as a consequence of Theorem A.45 in Bai and Silverstein [19] and Theorem 4,

$$L(F_{n-q}^{\mathbf{A}^T \mathbf{K} \mathbf{A}}, F_{n-q}^{\mathbf{A}^T \mathbf{M} \mathbf{A}}) \leq \left\| \mathbf{A}^T (\mathbf{K} - \mathbf{M}) \mathbf{A} \right\| \leq \|\mathbf{K} - \mathbf{M}\| = o_p(1).$$

Moreover, it follows from Theorem A.43 in Bai and Silverstein [19] that

$$\left\| F_{n-q}^{\mathbf{A}^T \mathbf{M} \mathbf{A}} - F_{n-q}^{\mathbf{A}^T \tilde{\mathbf{M}} \mathbf{A}} \right\| \leq \frac{1}{n - q} \text{rank}(\mathbf{A}^T (\mathbf{M} - \tilde{\mathbf{M}}) \mathbf{A}) \leq \frac{1}{n - q} \text{rank}(\mathbf{M} - \tilde{\mathbf{M}}) \leq \frac{7}{n - q},$$

which implies that

$$L(F_{n-q}^{A^T KA}, F_{n-q}^{A^T MA}) \leq L(F_{n-q}^{A^T KA}, F_{n-q}^{A^T MA}) + \left\| F_{n-q}^{A^T MA} - F_{n-q}^{A^T \bar{M}A} \right\| = o_p(1).$$

The remaining of the proof is the same as the proof of Theorem 3. \square

References

- Macgregor, S.; Cornes, B.K.; Martin, N.G.; Visscher, P.M. Bias, precision and heritability of self-reported and clinically measured height in Australian twins. *Hum. Genet.* **2006**, *120*, 571–580. [[CrossRef](#)] [[PubMed](#)]
- Silventoinen, K.; Sammalisto, S.; Perola, M.; Boomsma, D.I.; Cornes, B.K.; Davis, C.; Dunkel, L.; De Lange, M.; Harris, J.R.; Hjelmborg, J.V.; et al. Heritability of adult body height: A comparative study of twin cohorts in eight countries. *Twin Res. Hum. Genet.* **2003**, *6*, 399–408. [[CrossRef](#)] [[PubMed](#)]
- Yengo, L.; Vedantam, S.; Marouli, E.; Sidorenko, J.; Bartell, E.; Sakaue, S.; Graff, M.; Eliassen, A.U.; Jiang, Y.; Raghavan, S.; et al. A saturated map of common genetic variants associated with human height. *Nature* **2022**, *610*, 704–712. [[CrossRef](#)] [[PubMed](#)]
- Yang, J.; Lee, S.H.; Goddard, M.E.; Visscher, P.M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **2011**, *88*, 76–82. [[CrossRef](#)] [[PubMed](#)]
- Yang, J.; Benyamin, B.; McEvoy, B.P.; Gordon, S.; Henders, A.K.; Nyholt, D.R.; Madden, P.A.; Heath, A.C.; Martin, N.G.; Montgomery, G.W.; et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **2010**, *42*, 565–569. [[CrossRef](#)] [[PubMed](#)]
- Li, S.; Cai, T.T.; Li, H. Inference for high-dimensional linear mixed-effects models: A quasi-likelihood approach. *J. Am. Stat. Assoc.* **2022**, *117*, 1835–1846. [[CrossRef](#)] [[PubMed](#)]
- van de Geer, S.; Bühlmann, P.; Ritov, Y.; Dezeure, R. On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Stat.* **2014**, *42*, 1166–1202. [[CrossRef](#)]
- Zhang, C.H.; Zhang, S.S. Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. Stat. Methodol.* **2014**, *76*, 217–242. [[CrossRef](#)]
- Law, M.; Ritov, Y. Inference and estimation for random effects in high-dimensional linear mixed models. *J. Am. Stat. Assoc.* **2023**, *118*, 1682–1691. [[CrossRef](#)]
- Wu, M.C.; Lee, S.; Cai, T.; Li, Y.; Boehnke, M.; Lin, X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.* **2011**, *89*, 82–93. [[CrossRef](#)] [[PubMed](#)]
- Jiang, J.; Li, C.; Paul, D.; Yang, C.; Zhao, H. On high-dimensional misspecified mixed model analysis in genome-wide association study. *Ann. Stat.* **2016**, *44*, 2127–2160. [[CrossRef](#)]
- Dao, C.; Jiang, J.; Paul, D.; Zhao, H. Variance estimation and confidence intervals from genome-wide association studies through high-dimensional misspecified mixed model analysis. *J. Stat. Plan. Inference* **2022**, *220*, 15–23. [[CrossRef](#)] [[PubMed](#)]
- Jiang, J.; Jiang, W.; Paul, D.; Zhang, Y.; Zhao, H. High-dimensional asymptotic behavior of inference based on gwas summary statistic. *Stat. Sin.* **2023**, *33*, 1555–1576. [[CrossRef](#)]
- Liu, D.; Lin, X.; Ghosh, D. Semiparametric regression of multidimensional genetic pathway data: Least-squares kernel machines and linear mixed models. *Biometrics* **2007**, *63*, 1079–1088. [[CrossRef](#)] [[PubMed](#)]
- Banerjee, S.; Carlin, B.P.; Gelfand, A.E. *Hierarchical Modeling and Analysis for Spatial Data*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2003.
- Shen, X.; Wen, Y.; Cui, Y.; Lu, Q. A conditional autoregressive model for genetic association analysis accounting for genetic heterogeneity. *Stat. Med.* **2022**, *41*, 517–542. [[CrossRef](#)] [[PubMed](#)]
- de Los Campos, G.; Vazquez, A.I.; Fernando, R.; Klimentidis, Y.C.; Sorensen, D. Prediction of complex human traits using the genomic best linear unbiased predictor. *PLoS Genet.* **2013**, *9*, e1003608. [[CrossRef](#)] [[PubMed](#)]
- Searle, S.R.; Casella, G.; McCulloch, C.E. *Variance Components*; John Wiley & Sons: Hoboken, NJ, USA, 2009; Volume 391.
- Bai, Z.; Silverstein, J.W. *Spectral Analysis of Large Dimensional Random Matrices*; Springer: Berlin/Heidelberg, Germany, 2010; Volume 20.
- Paul, D.; Aue, A. Random matrix theory in statistics: A review. *J. Stat. Plan. Inference* **2014**, *150*, 1–29. [[CrossRef](#)]
- El Karoui, N. The spectrum of kernel random matrices. *Ann. Stat.* **2010**, *38*, 1–50. [[CrossRef](#)]
- Couillet, R.; Liao, Z. *Random Matrix Methods for Machine Learning*; Cambridge University Press: Cambridge, UK, 2022.
- Hiai, F. Monotonicity for entrywise functions of matrices. *Linear Algebra Its Appl.* **2009**, *431*, 1125–1146. [[CrossRef](#)]
- Wendland, H. *Scattered Data Approximation*; Cambridge University Press: Cambridge, UK, 2004; Volume 17.
- Li, M.; He, Z.; Zhang, M.; Zhan, X.; Wei, C.; Elston, R.C.; Lu, Q. A generalized genetic random field method for the genetic association analysis of sequencing data. *Genet. Epidemiol.* **2014**, *38*, 242–253. [[CrossRef](#)] [[PubMed](#)]
- Schrepft, S.; van Jaarsveld, C.H.; Fisher, A.; Herle, M.; Smith, A.D.; Fildes, A.; Llewellyn, C.H. Variation in the heritability of child body mass index by obesogenic home environment. *JAMA Pediatr.* **2018**, *172*, 1153–1160. [[CrossRef](#)] [[PubMed](#)]

27. Ghorbani, B.; Mei, S.; Misiakiewicz, T.; Montanari, A. When do neural networks outperform kernel methods? *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 14820–14830. [[CrossRef](#)]
28. Mei, S.; Misiakiewicz, T.; Montanari, A. Learning with invariances in random features and kernel models. In Proceedings of the Conference on Learning Theory, PMLR, Boulder, CO, USA, 15–19 August 2021 ; pp. 3351–3418.
29. Gönen, M.; Alpaydm, E. Multiple kernel learning algorithms. *J. Mach. Learn. Res.* **2011**, *12*, 2211–2268.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.