Partitioning Fraction of Variance Explained into Strong Localized Effects and Weak Diffuse Effects

Fang Nan*[1], David Azriel[2], and Armin Schwartzman[1,3]

[1] *Division of Biostatistics, University of California San Diego, La Jolla, CA, USA*
[2] *Faculty of Data and Decision Sciences, The Technion, Haifa, Israel*
[3] *Halıcıoğlu Data Science Institute, University of California San Diego, La Jolla, CA, USA*

Abstract

High-dimensional genetic data present substantial challenges for estimating the fraction of variance explained (FVE) by genome-wide single-nucleotide polymorphisms (SNPs). Standard approaches for SNP heritability estimation, such as GWAS heritability (GWASH) and linkage disequilibrium score (LDSC) regression, typically assume Gaussian distributions for SNP effect sizes. However, empirical evidence indicates that SNP effects are often heavy-tailed, with a small subset of variants exerting disproportionately large influence. Such settings violate the recently established bounded-kurtosis effect (BKE) condition, under which these FVE estimators are consistent. Consequently, widely used methods may yield severely biased estimates when strong effects are present. We introduce a decomposed FVE estimation framework that accommodates heavy-tailed and heterogeneous SNP effect distributions. The proposed approach partitions total heritability into contributions from strong and weak genetic effects, estimating the former using low-dimensional adjusted R^2 and the latter using an extension of FVE estimation methodology that remains valid under BKE compliance. We further develop a test for detecting violations of the BKE condition and compare several high-dimensional screening procedures for identifying strong-effect SNPs when they are not known in advance. Simulation studies show that the proposed decomposition substantially improves estimation accuracy over existing approaches in the presence of heavy-tailed effects. Application to the Adolescent Brain Cognitive Development (ABCD) Study demonstrates the practical utility of the method, yielding more reliable heritability estimates for the PolyVoxel Score, a neuroimaging-based biomarker linked to iron accumulation. Our results highlight the importance of accommodating effect heterogeneity in large-scale genomic studies.

1 Introduction

High-dimensional data, including whole-genome genotyping, exhibits significant complexity, often with dimensionalities that far exceed the number of subjects. As with many complex behavioral

*Correspondence: fnan@ucsd.edu

traits, individual biomarkers typically contribute only marginally to the overall variance in behavioral phenotypes. However, the vast number of potential biomarkers can collectively make a substantial contribution to the fraction of variance explained (FVE).

Genome-wide Association Studies (GWAS) have become the dominant approach in genetic research on complex disorders due to their relative simplicity, decreasing costs, and notable successes [Visscher et al., 2017]. These studies typically involve regressing quantitative trait measures or case-control status against allele counts for millions of single nucleotide polymorphisms (SNPs) using linear or logistic models. The results are summary statistics—regression coefficients, standard errors, test statistics, and p-values that describe the marginal associations between each SNP and the studied outcome.

A key concept in understanding complex disorders is SNP-heritability, h^2_{SNP}, which quantifies FVE attributed to additive relationships among all GWAS SNPs, regardless of their significance [Yang et al., 2010, Lee et al., 2011]. Estimates of h^2_{SNP} suggest that a considerable portion of susceptibility to complex disorders (approximately 10% - 40%) may be captured by linear additive models based on the same SNPs surveyed by GWAS. The coefficients multiplying the SNP counts in such models and representing the association of each SNP with the outcome are often referred to as SNP effects. However, many existing methods for estimating h^2_{SNP} assume a Gaussian distribution for marginal SNP effects, whereas empirical evidence suggests that these effects are not well-characterized by such a distribution [Gusev et al., 2014]. Instead, SNP effects often exhibit a mixture of weak and strong values that do not correspond to a Gaussian distribution [Holland et al., 2020]. This highlights the need for more flexible and realistic methods to model the effect of SNPs. Moreover, current methods often lack robust theoretical foundations and depend on ad hoc simulation studies for validation. Their performance, including bias and error rates, has been shown to be sensitive to the parameters of these simulations [Evans et al., 2018]. To ensure reliable estimates of h^2_{SNP}, it is crucial to establish realistic conditions and develop methods for testing these conditions with real data.

Estimating the FVE requires advanced statistical methods specifically designed for high-dimensional genetic data. Many FVE estimators exist with various properties, such as consistency [Schwartzman et al., 2019, Dicker, 2014], suitability for non-Gaussian effects [Janson et al., 2017] and non-Gaussian predictors [Hou et al., 2019], applicability to summary statistics [Speed and Balding, 2019, Shi et al., 2016], and effectiveness across both small [Schwartzman et al., 2019, Janson et al., 2017] and large datasets [Speed et al., 2012, Shi et al., 2016]. As a benchmark, an FVE estimator for SNP-based heritability (h^2_{SNP}) of continuous traits, termed GWAS heritability (GWASH), under a fixed SNP effects framework [Schwartzman et al., 2019], has been developed. However, both GWASH and other summary statistic-based approaches, such as linkage disequilibrium score (LDSC) regression [Bulik-Sullivan et al., 2015], assume that all SNP effects have similar magnitudes. As a result, these methods typically model SNP effects as drawn from a Gaussian distribution centered around zero. As mentioned above, This assumption often does not reflect real-world scenarios [Gusev et al., 2014], where certain SNPs have significantly stronger effects compared to others. This means

that the distribution of SNP effects could be heavy-tailed, indicated by a large kurtosis. Although Janson et al. [2017] relaxes the Gaussian-effect assumption and is robust to non-Gaussian signals, their approach requires the predictors to be i.i.d., an assumption rarely satisfied in genetic data where linkage disequilibrium induces correlation among SNPs, thus limiting its practical applicability in this context.

A high kurtosis of SNP effects violates an important condition, shown by Azriel et al. [2025] to be necessary and sufficient for consistent FVE estimation using GWASH and LDSC regression. Azriel et al. [2025] formally defined this condition, referred to as bounded-kurtosis effects (BKE), requiring the kurtosis of the distribution of SNP effects to be asymptotically bounded as the number of predictors, i.e., SNPs, increases. In this article, we first extend the GWASH approach by developing a more general framework for FVE estimation that accommodates heavier-tailed, non-Gaussian distributions of SNP effects. Violations of the BKE condition often occur when a small subset of SNPs exerts disproportionately large effects, giving rise to a mixture structure in which strong and weak effects coexist. By explicitly modeling this structure, we account for scenarios where SNP effects are partially concentrated in a small subset of variants while remaining broadly distributed across the genome.

To address the violation of the BKE condition, we propose a decomposed FVE estimator. The essential idea is to quantify the relative contributions of different groups of predictors, a concept commonly referred to as enrichment in the context of GWAS [Finucane et al., 2015]. This approach partitions the total FVE into contributions from strong and weak genetic effects, allowing their heritability to be estimated in sequence and then assembled to obtain the total SNP heritability. After decomposition, we assume that the BKE condition holds within the high-dimensional weak component. The strong-effect component is low-dimensional and therefore does not require the BKE condition. This procedure is conceptually analogous to ANOVA in linear regression, but is adapted here to handle high-dimensional genetic predictors.

In our framework, the number of SNPs with strong effects is assumed to be smaller than the number of subjects, resulting in a low-dimensional setting. Consequently, the heritability attributable to strong effects can be reliably estimated using the adjusted R^2. In contrast, the number of weak-effect SNPs typically exceeds the sample size, creating a high-dimensional estimation problem. For this component, we apply the GWASH method [Schwartzman et al., 2019] and Genome-wide Complex Trait Analysis (GCTA, Yang et al. [2011]) for comparison. LDSC regression with fixed intercept gives values that are numerically very close to GWASH [Schwartzman et al., 2019, Azriel et al., 2025] Therefore, we do not explicitly include LDSC regression with fixed intercept in our study, but the results from GWASH can be understood to apply equally to that estimator. Although GCTA is one of the most widely used methods for estimating SNP heritability, its robustness to heavy-tailed SNP effect distributions has not been rigorously examined in theory. Our simulation studies, however, indicate that GCTA is also sensitive to violations of the BKE condition. Encouragingly, integrating GCTA within our proposed decomposition framework substantially improves its performance under non-Gaussian settings, enhancing estimation accuracy even when the BKE condition is violated.

By modeling the strong and weak effects separately and then aggregating their contributions, our method effectively mitigates the impact of BKE violations and yields a consistent estimate of the total SNP heritability.

Sometimes, SNPs with strong effects are already known from previous studies. However, when such a subset is not known a priori, identifying these strong effects becomes a crucial step in our framework. We first develop a testing procedure to evaluate whether the SNP effects in the sample satisfy the BKE condition, thereby informing the suitability of applying the proposed decomposed FVE estimation. Next, we propose to use high-dimensional screening methods for this purpose. Existing screening methods for high-dimensional data have not been implemented for FVE decomposition before, and they were originally designed for sparse models with many zero coefficients. For instance, the Sure Independence Screening (SIS, [Fan and Lv, 2008]) procedure screens out irrelevant features based on their marginal correlations with the response variable. The Robust Rank Correlation Screening (RRCS, [Li et al., 2012]) method employs Kendall's τ rank correlation to select variables in ultra-high-dimensional data, offering robustness to outliers and invariance to monotonic transformations. The High-dimensional Ordinary Least Squares Projection (HOLP, [Wang and Leng, 2016]) method ensures consistency in variable selection without relying on strong assumptions like those in SIS. Additionally, the Bonferroni correction [Bonferroni, 1936] can be readily applied in our setting to identify SNPs with strong effects. We present a simulation study to evaluate and compare the performance of the four aforementioned methods in connection to FVE estimation.

To demonstrate the practical utility of our proposed framework, we apply it to the Adolescent Brain Cognitive Development (ABCD) Study [Casey et al., 2018], one of the largest longitudinal studies of brain development and child health in the United States. The ABCD dataset includes extensive whole-genome genotyping and rich behavioral and cognitive phenotypes from over 11,000 children who were 9–10 years old at baseline, with their biological and behavioral development longitudinally tracked over subsequent years [Luciana et al., 2018, Casey et al., 2018]. Genetic factors significantly influence mental illnesses [Laville et al., 2019, Polderman et al., 2015], and understanding these genetic contributions is crucial for elucidating the etiology of psychiatric disorders and developing targeted treatments [Sullivan and Geschwind, 2019]. The primary outcome in our analysis is the PolyVoxel Score (PVS), defined by Loughnan et al. [2024], which measures how closely an individual's brain pattern matches the archetypal hemochromatosis brain, marked by regional iron accumulation in motor circuits. In the ABCD dataset, Loughnan et al. [2024] identified potential strong SNP effects contributing to the heritability of the PVS. These findings make the PVS a compelling example for applying our proposed decomposed FVE estimation framework.

2 General Setting

2.1 Polygenic Model and FVE

Suppose a continuous outcome y_i is measured with a panel of m predictors $\vec{\boldsymbol{x}}_i = (x_{i1}, \ldots, x_{im})$ in n independent and identically distributed subjects, $i = 1, \ldots, n$. In the context of genetics, these are SNP allele counts (taking values 0, 1, 2). The poly-additive model, called polygenic model in genetics ([Fisher, 1919, Lynch et al., 1998]), in scalar or vector form, is

$$y_i = \vec{\boldsymbol{x}}_i \boldsymbol{\beta} + \varepsilon_i \quad i = 1, \ldots, n, \qquad \text{or} \qquad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)^\top$, $\boldsymbol{X}$ is the regression matrix with rows $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_n$ representing subjects' SNP panels and columns $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ representing SNPs for all subjects. The errors $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$ are independent of $\boldsymbol{X}$, and ε_i are IID with zero mean and variance σ_ε^2. To simplify the notation and theoretical calculations, we assume $\mathrm{E}(\vec{\boldsymbol{x}}_i) = \boldsymbol{0}$, $\mathrm{E}(y_i) = 0$, $\mathrm{Var}(\vec{\boldsymbol{x}}_{ij}) = 1$ for $j = 1 \ldots, m$, and $\mathrm{Var}(y_i) = 1$. Let $\boldsymbol{\Sigma} = \mathrm{Cov}(\vec{\boldsymbol{x}}_i)$ denote the $m \times m$ covariance matrix among SNP predictors in the underlying population.

The goal is to estimate the SNP heritability or FVE of Model (1). When Model (1) is treated as a fixed-effects model, where the SNP effects $\boldsymbol{\beta}$ are considered fixed, the FVE, denoted by h_β^2, is defined as

$$h_\beta^2 := \frac{\mathrm{Var}(\vec{\boldsymbol{x}}_i\boldsymbol{\beta}|\boldsymbol{\beta})}{\mathrm{Var}(y_i|\boldsymbol{\beta})} = \frac{\mathrm{Var}(y_i|\boldsymbol{\beta})}{\mathrm{Var}(\vec{\boldsymbol{x}}_i\boldsymbol{\beta}|\boldsymbol{\beta}) + \mathrm{Var}(\varepsilon_i|\boldsymbol{\beta})} = \frac{\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta}}{\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta} + \sigma_\varepsilon^2}, \tag{2}$$

also referred to as the "fixed-effects heritability". Here, we define the conditional variance of y_i as $\sigma_{y,\beta}^2 = \mathrm{Var}(y_i|\boldsymbol{\beta}) = \boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta} + \sigma_\varepsilon^2$.

However, it is often more mathematically convenient to treat Model (1) as a random-effects model, where the SNP effects $\boldsymbol{\beta}$ are considered random. In this case, the FVE is referred to as the "random-effects heritability", denoted by h^2. Here we assume that the SNP effects $\beta_1, \ldots, \beta_m$ are i.i.d. with mean zero and a common variance $\mathrm{Var}(\beta_j) = h^2/m$. We also assume that $\boldsymbol{\beta}$ is independent of both $\vec{\boldsymbol{x}}_i$ and ε_i. Under these assumptions, we have $\mathrm{E}(\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta}) = h^2$ and $\mathrm{Var}(\varepsilon_i) = 1 - h^2$. Therefore, the quantity h^2 represents the unconditional FVE under Model (1):

$$\frac{\mathrm{Var}(\vec{\boldsymbol{x}}_i^\top\boldsymbol{\beta})}{\mathrm{Var}(y_i)} = \frac{\mathrm{E}(\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta})}{\mathrm{E}(\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta}) + \mathrm{Var}(\varepsilon_i)} = \frac{h^2}{h^2 + (1 - h^2)} = h^2. \tag{3}$$

To model the SNP effects $\boldsymbol{\beta}$ as non-Gaussian, an important concept in this context is kurtosis. Treating the SNP effects as random, the kurtosis of β_j is defined as

$$\kappa_\beta = \mathrm{Kurt}[\beta_j] = \frac{\mu_{4,\beta}}{\sigma_\beta^4}, \tag{4}$$

where $\mu_{4,\beta}$ is the fourth central moment and σ_β is the standard deviation of β_j. Typically, the quantity $\mathrm{Kurt}[\beta_j] - 3$ is referred to as the excess kurtosis of β_j, using the normal distribution, whose kurtosis is 3, as the reference.

2.2 GWASH Estimator

Considering Model (1) and define $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_m)$ as the vector of correlations between the entries of $\vec{\boldsymbol{x}}$ and y given the SNP effects. Specifically, $\boldsymbol{\eta}$ is given by:

$$\boldsymbol{\eta} = \mathrm{E}(\vec{\boldsymbol{x}}^\top y | \boldsymbol{\beta}) = \boldsymbol{\Sigma}\boldsymbol{\beta}. \tag{5}$$

Let $s^2 = ||\boldsymbol{\eta}||^2$ and μ_2 represent the second spectral moment of the correlation matrix $\boldsymbol{\Sigma}$ of $\vec{\boldsymbol{x}}$, defined as $\mu_2 = \mathrm{tr}(\boldsymbol{\Sigma}^2)/m$. Using these definitions, we have that

$$\frac{s^2}{\mu_2} = \frac{\boldsymbol{\eta}^T\boldsymbol{\eta}}{\mu_2} = \frac{\boldsymbol{\beta}^T\boldsymbol{\Sigma}^2\boldsymbol{\beta}}{\mu_2}. \tag{6}$$

Recall that $\beta_1, \ldots, \beta_m$ are i.i.d. with mean zero and variance $\mathrm{Var}(\beta_i) = h^2/m$, hence,

$$\mathrm{E}\left[\frac{s^2}{\mu_2}\right] = \frac{m}{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\mathrm{E}\left[\boldsymbol{\beta}^T\boldsymbol{\Sigma}^2\boldsymbol{\beta}\right] = \frac{m}{\mathrm{tr}(\boldsymbol{\Sigma}^2)}\mathrm{E}\left[\mathrm{tr}(\boldsymbol{\Sigma}^2\boldsymbol{\beta}\boldsymbol{\beta}^T)\right] = m\frac{h^2}{m} = h^2.$$

When considering standardization in the sample, the predictors $\boldsymbol{X}$ and the outcome $\boldsymbol{y}$ are transformed into their standardized versions as follows:

$$\tilde{\boldsymbol{x}}_j = \frac{\boldsymbol{x}_j - \bar{\boldsymbol{x}}_j}{\hat{\sigma}_{\boldsymbol{x}_j}} \quad j = 1, \ldots, m, \quad \text{and} \quad \tilde{\boldsymbol{y}} = \frac{\boldsymbol{y} - \bar{\boldsymbol{y}}}{\hat{\sigma}_y},$$

where $\bar{\boldsymbol{x}}$ and $\hat{\sigma}_{\boldsymbol{x}_j}$ are the sample mean and standard deviation of $\boldsymbol{x}_j$, $\bar{\boldsymbol{y}}$ and $\hat{\sigma}_y$ are the sample mean and standard deviation of the outcome $\boldsymbol{y}$. Let $\tilde{\boldsymbol{X}} = (\tilde{\boldsymbol{x}}_1, \ldots, \tilde{\boldsymbol{x}}_m)$ represent the standardized version of $\boldsymbol{X}$, obtained by standardizing its columns.

The sample correlation between $\boldsymbol{x}_j$ and $\boldsymbol{y}$ is then computed as:

$$\hat{\eta}_j = \frac{1}{n}\tilde{\boldsymbol{x}}_j^\top\tilde{\boldsymbol{y}} \quad \text{for} \quad j = 1, \ldots, m. \tag{7}$$

which serves as the sample version of $\boldsymbol{\eta}$ defined in (5). The j-th LD-score, for $j = 1, \ldots, m$, as defined by [Bulik-Sullivan et al., 2015], measures the sum of squared sample correlations between $\boldsymbol{x}_j$ and all other predictors. The population and sample versions of the LD score are defined, respectively, as

$$\ell_j = \sum_{p=1}^{m}(\boldsymbol{\Sigma}_{j,p})^2, \quad \hat{\ell}_j = \sum_{i=1}^{n}\left(\frac{1}{n}\tilde{\boldsymbol{x}}_j^\top\tilde{\boldsymbol{x}}_i\right)^2 = \frac{1}{n^2}\tilde{\boldsymbol{x}}_j^\top\tilde{\boldsymbol{X}}\tilde{\boldsymbol{X}}^\top\tilde{\boldsymbol{x}}_j, \tag{8}$$

Notice that since we assume $\mathrm{Var}(X_{ij}) = 1$, it follows that $\boldsymbol{\Sigma}$ is the population correlation matrix of the predictors. The bias-corrected LD-score $\hat{\ell}_j - m/n$ is approximately unbiased for ℓ_j for Gaussian predictors as shown in Equation (5) of [Azriel et al., 2025].

The GWASH estimator can be obtained as a plug-in estimator of Equation (6):

$$\hat{h}^2_{\mathrm{GWASH}} = \frac{\hat{s}^2}{\hat{\mu}_2}, \tag{9}$$

where $\hat{s}^2 = \hat{\boldsymbol{\eta}}^\top \hat{\boldsymbol{\eta}} - m/n$ is approximately an unbiased estimator of s^2 as shown in Equation (6) of [Azriel et al., 2025], and $\hat{\mu}_2$ is an unbiased estimator of the second spectral moment μ_2:

$$\hat{\mu}_2 = \frac{1}{m}\sum_{j=1}^{m} \hat{\ell}_j - \frac{m}{n}. \tag{10}$$

3 Estimation Consistency and the Effect of Kurtosis

3.1 Consistency Conditions of GWASH Estimator

Two important conditions for the consistency of the GWASH estimator and LDSC regression with fixed intercept were discussed by [Azriel et al., 2025]. The first is the weak dependence condition, which requires that the second spectral moment of the LD matrix $\boldsymbol{\Sigma}$ to be bounded. This condition is generally assumed to hold in GWAS as dependence between SNPs is local [Xie et al., 2011]. However, it may be violated in heterogeneous populations, in which case the issue can be mitigated by first regressing the outcome on principal components that capture population structure [Singh Sachan, 2025]. The focus of the current paper is on the second consistency condition, namely, the bounded-kurtosis-of-effects (BKE) condition, which is formally defined as:

$$\frac{1}{m}[\kappa_\beta - 3] \to 0 \quad \text{as} \quad m \to \infty, \tag{11}$$

where κ_β is the kurtosis of β_j as defined in Equation (4).

While a detailed description of the BKE condition can be found in [Azriel et al., 2025], here we briefly discuss the implications of that in our setting. Let

$$\hat{h}^2_{\text{GWASH}} = \frac{N_{\text{GWASH}}}{D_{\text{GWASH}}}, \quad \text{where} \quad N_{\text{GWASH}} = \frac{1}{m}\sum_{j=1}^{m}(\hat{\eta}_j^2 - 1), \quad D_{\text{GWASH}} = \frac{1}{m}\sum_{j=1}^{m}\frac{n}{m}\hat{\ell}_j.$$

Under Model (1), and assuming weak dependence together with additional regularity conditions (e.g., bounded moments of the error term ε) as specified in Equation (12) of Azriel et al. [2025], we have

$$\text{Var}(\hat{h}^2_{\text{GWASH}}|\boldsymbol{X}) = \frac{\text{Var}(N_{\text{GWASH}}|\boldsymbol{X})}{(D_{\text{GWASH}})^2},$$

where

$$\text{Var}(N_{\text{GWASH}}|\boldsymbol{X}) = \frac{1}{m}[\kappa_\beta - 3]\frac{h^4 n^2}{m^2}\frac{1}{m}\sum_{j=1}^{m}\hat{\ell}_j + O\left(\frac{1}{m}\right). \tag{12}$$

The expression above in (12) explains how the BKE condition guarantees the consistency of $\hat{h}^2_{\text{GWASH}}$. Here, we assume that m/n converges to a positive constant as both m and n increase, while the average of $\hat{\ell}_j$'s converges to a constant under the assumption of weak dependence. The expression in the square brackets is the excess kurtosis. If $\boldsymbol{\beta}$ follows a normal distribution, this

excess kurtosis is zero, meaning that the term does not contribute to the variance and the variance goes to zero at a rate of $1/m$ given by the residual term. However, if $\boldsymbol{\beta}$ is not normally distributed, the excess kurtosis may be large, and the variance may not go to zero. We show an example in the next section.

3.2 Kurtosis and Modeling of Coefficient Effects

To more thoroughly investigate and model SNP effect coefficients $\boldsymbol{\beta}$ with varying kurtosis, we present here a Gaussian mixture model as an example, which is implemented in our subsequent simulation studies. However, we note that this is just one possible method for modeling the effects, and alternative approaches may also be considered.

Consider a Gaussian mixture distribution of $\boldsymbol{\beta}$, where a small fraction $p = \kappa/m$ of β's having variance ψ^2 and the rest have variance γ^2, and $p \in [0, 1]$.

$$\boldsymbol{\beta} \sim \begin{cases} N(0, \psi^2) & \text{with probability} \quad p \\ N(0, \gamma^2) & \text{with probability} \quad 1-p \end{cases} \tag{13}$$

The variance of β is $\sigma_\beta^2 = \mathrm{E}(\beta_j^2) = p\psi^2 + (1-p)\gamma^2$. By letting the total variance of $\sigma_\beta^2 = h^2/m$, we have $\sigma_\beta^2 = p\psi^2 + (1-p)\gamma^2 = h^2/m$, with $p\psi^2 = \theta h^2/m$ and $(1-p)\gamma^2 = (1-\theta)h^2/m$, where θ denotes the fraction of variance that the small fraction β's take. Then with restricting $p \in (0, 1)$, we have $\psi^2 = \theta h^2/(mp)$ and $\gamma^2 = (1-\theta)h^2/[m(1-p)]$. In order to derive the kurtosis in terms of p and θ, we first derive the fourth moment of β as

$$\mathrm{E}[\beta_j^4] = \frac{3h^4}{m^2} \times \left[\frac{\theta^2}{p} + \frac{(1-\theta)^2}{1-p}\right].$$

Then, since $\sigma_\beta^4 = (\sigma_\beta^2)^2 = h^4/m^2$, the kurtosis is

$$\kappa_\beta = \frac{\mathrm{E}[\beta_j^4]}{\sigma_\beta^4} = 3\left[\frac{\theta^2}{p} + \frac{(1-\theta)^2}{1-p}\right]. \tag{14}$$

The mixture Gaussian model for $\boldsymbol{\beta}$ provides a flexible framework that accommodates both scenarios where the BKE condition is satisfied and those where it is violated. As shown in Equation (14), when $p = \theta$, the mixture model reduces to a standard Gaussian distribution, which satisfies the BKE condition. Furthermore, each component in the mixture individually satisfies the BKE condition when considered in isolation. However, by manipulating the value of p and θ, for example, setting p to a relatively small value and θ to a large proportion, the resulting distribution exhibits high kurtosis, thereby violating the BKE condition, as demonstrated by the following proposition. The corresponding proof of this proposition is provided in Appendix B.1.

Proposition 3.1. *Under the mixture Gaussian model in (13), suppose $\theta = \theta_0 \in (0, 1)$ is fixed and $p \in [0, 1]$. Then the BKE condition is violated if either $mp \to c < \infty$ or $m(1-p) \to c < \infty$ as $m \to \infty$. The BKE condition is satisfied if $p = \theta_0$, or if $p = 0$ or 1, reducing to a normal Gaussian*

distribution.

3.3 Relationship Between the Kurtosis of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$

In practice, the true coefficients $\boldsymbol{\beta}$ are unobservable; therefore, the exact value of κ_β is not directly available. In such cases, it is important to understand the relationship between κ_β and κ_{η_j}, the latter being estimable from observable data. Here, κ_{η_j} denotes the kurtosis of η_j, defined as

$$\kappa_{\eta_j} = \mathrm{Kurt}[\eta_j] = \frac{\mu_{4,\eta_j}}{\sigma_{\eta_j}^4}, \tag{15}$$

where μ_{4,η_j} is the fourth central moment and σ_{η_j} is the standard deviation of η_j. This relationship between κ_β and κ_{η_j} is established by the following proposition:

Proposition 3.2. *Consider Model (1) and assume the random effects $\beta_j \sim (0, \sigma_\beta^2)$. Then*

$$\kappa_{\eta_j} - 3 = (\kappa_\beta - 3)\frac{\sum_{k=1}^m \boldsymbol{\Sigma}_{jk}^4}{\left(\sum_{k=1}^m \boldsymbol{\Sigma}_{jk}^2\right)^2}. \tag{16}$$

Proposition 3.2 shows that although the kurtosis of β_j is not equal to the kurtosis of η_j, a higher value of κ_{η_j} generally indicates a higher value of κ_β. The proof of this proposition is provided in Appendix B.2. The effect of the dependence in $\boldsymbol{\Sigma}$ on this relationship is demonstrated through simulations in Section 6.2.

4 FVE Decomposition

If the BKE condition is violated, the GWASH estimator and others become inappropriate due to its inconsistency. To address this, we propose an alternative method for estimating FVE, referred to as FVE decomposition, and subsequently introduce the decomposed FVE estimator.

The core idea is to determine the relative contribution of different groups of predictors, by partitioning the total FVE among subsets of predictors. In our framework, SNP effects are partitioned into two components: a low-dimensional subset and a high-dimensional subset, with the latter assumed to satisfy the BKE condition. The FVE attributable to the low-dimensional component can be accurately estimated using the adjusted R^2. For the high-dimensional component, we employ high-dimensional FVE estimators, such as GWASH or GCTA, to quantify its contribution to the total FVE, as the BKE condition is presumed to hold after partitioning. A brief description of GCTA is presented in Appendix A.1.

4.1 FVE Decomposition into Sub-Classes of Predictors

Here, we propose a general FVE decomposition framework that partitions the total FVE into two components of arbitrary dimensionality. In this section, we describe the decomposition at the

population level using theoretical parameters. The estimation procedures corresponding to this decomposition will be detailed in the next section.

In model (1), consider a random row vector $\vec{\boldsymbol{x}} \in \mathbb{R}^m$ and a random variable $y \in \mathbb{R}$, both with mean zero, dropping the subject index i for simplicity. The vector of effects $\boldsymbol{\beta}$ can be characterized as providing the best linear approximation of y by $\vec{\boldsymbol{x}}$:

$$\boldsymbol{\beta} = \arg\min_{b \in \mathbb{R}} \mathrm{E}(y - \vec{\boldsymbol{x}}\boldsymbol{b})^2 = \boldsymbol{\Sigma}^{-1}\mathrm{E}(\vec{\boldsymbol{x}}^\top y) \qquad \text{where} \qquad \boldsymbol{\Sigma} = \mathrm{E}(\vec{\boldsymbol{x}}^\top \vec{\boldsymbol{x}}). \tag{17}$$

Consider a partition of the predictors $\vec{\boldsymbol{x}} = (\vec{\boldsymbol{x}}_1 \ \vec{\boldsymbol{x}}_2)$ into two subsets, where $\vec{\boldsymbol{x}}_1 \in \mathbb{R}^{m_1}$ and $\vec{\boldsymbol{x}}_2 \in \mathbb{R}^{m_2}$, partitioning the effects vector as $\boldsymbol{\beta}^\top = (\boldsymbol{\beta}_1^\top \ \boldsymbol{\beta}_2^\top)$ and the covariance $\boldsymbol{\Sigma}$ into blocks $\boldsymbol{\Sigma}_{11} = \mathrm{E}(\vec{\boldsymbol{x}}_1^\top \vec{\boldsymbol{x}}_1)$, $\boldsymbol{\Sigma}_{12} = \mathrm{E}(\vec{\boldsymbol{x}}_1^\top \vec{\boldsymbol{x}}_2)$, $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}^\top$ and $\boldsymbol{\Sigma}_{22} = \mathrm{E}(\vec{\boldsymbol{x}}_2^\top \vec{\boldsymbol{x}}_2)$. Based on the above partition, we can derive the decomposition of FVE as follows.

Proposition 4.1. *Consider Model (1) and let* $\vec{\boldsymbol{x}}_{\bullet 2} = \vec{\boldsymbol{x}}_2 - \vec{\boldsymbol{x}}_1\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $y_{\bullet 2} = y - \vec{\boldsymbol{x}}_1\boldsymbol{\Sigma}_{11}^{-1}\mathrm{E}(\vec{\boldsymbol{x}}_1^\top y)$, $\boldsymbol{\Sigma}_{\bullet 22} = \mathrm{E}(\vec{\boldsymbol{x}}_{\bullet 2}^\top \vec{\boldsymbol{x}}_{\bullet 2})$. *Then the conditional FVE in (2) can be decomposed as*

$$h^2_\beta = \frac{\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}}{\sigma^2_{y,\beta}} = \frac{\boldsymbol{\beta}_1^{*T} \boldsymbol{\Sigma}_{11} \boldsymbol{\beta}_1^*}{\sigma^2_{y,\beta}} + \frac{\boldsymbol{\beta_2}^\top \boldsymbol{\Sigma_{\bullet 22}} \boldsymbol{\beta_2}}{\sigma^2_{y_{\bullet 2},\beta}} \frac{\sigma^2_{y_{\bullet 2},\beta}}{\sigma^2_{y,\beta}} = h^2_{1,\beta} + \lambda_\beta h^2_{\bullet 2,\beta}, \tag{18}$$

where $\boldsymbol{\beta}_1^* = \boldsymbol{\Sigma}_{11}^{-1}\mathrm{E}(\vec{\boldsymbol{x}}_1^\top y)$, $\boldsymbol{\beta}_2 = \boldsymbol{\Sigma}_{\bullet 22}^{-1}\mathrm{E}(\vec{\boldsymbol{x}}_{\bullet 2}^\top y_{\bullet 2})$, $\lambda_\beta = \sigma^2_{y_{\bullet 2},\beta}/\sigma^2_{y,\beta}$, $\sigma^2_{y,\beta} = \mathrm{Var}(y|\boldsymbol{\beta})$, $\sigma^2_{y_{\bullet 2},\beta} = \mathrm{Var}(y_{\bullet 2}|\boldsymbol{\beta})$.

The first term in the sum (18) relates to the marginal regression of y on $\vec{\boldsymbol{x}}_1$, while the second term corresponds to the regression of y on $\vec{\boldsymbol{x}}_{\bullet 2}$, adjusted for $\vec{\boldsymbol{x}}_1$. Notice that the two terms correspond to orthogonal predictors so that they can be estimated separately and added. The detailed proof of this proposition is provided in Appendix B.3.

4.2 Estimation of FVE Components

In this section, we provide the estimation details for Proposition 4.1. Let $\boldsymbol{X}_1 \in \mathbb{R}^{n \times m_1}$ and $\boldsymbol{X}_2 \in \mathbb{R}^{n \times m_2}$ denote the partitioned SNP matrices, where n is the number of subjects and $m_1 + m_2 = m > n$ is the total number of SNPs. Let $\boldsymbol{y} \in \mathbb{R}^n$ be the trait of interest. We assume that $m_1 < n$, so the estimation of $h^2_{1,\beta}$, as defined in equation (18), corresponds to a low-dimensional setting. Therefore, it can be reliably estimated using the adjusted R^2, which we denote by $\hat{h}^2_1$,

$$\hat{h}^2_1 = \text{Adjusted } R^2(\boldsymbol{X}_1, \boldsymbol{y}). \tag{19}$$

To isolate the contribution of the high-dimensional component $\boldsymbol{X}_2$, we residualize both $\boldsymbol{X}_2$ and $\boldsymbol{y}$ with respect to $\boldsymbol{X}_1$. Specifically, we compute the residualized matrices:

$$\boldsymbol{X}_{\bullet 2} = \boldsymbol{X}_2 - \boldsymbol{X}_1(\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top \boldsymbol{X}_2, \quad \boldsymbol{y}_{\bullet 2} = \boldsymbol{y} - \boldsymbol{X}_1(\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top \boldsymbol{y}.$$

Note that $\boldsymbol{X}_{\bullet 2} \in \mathbb{R}^{n \times m_2}$, and since $m_2 > n$, estimating $h^2_{\bullet 2,\beta}$ is a high-dimensional problem. This

can be addressed using high-dimensional FVE estimation methods, such as GWASH, LDSC, or GCTA. In this case, we apply the GWASH or GCTA estimator to obtain an estimate of the residual heritability, denoted by $\hat{h}^2_{\bullet 2}$,

$$\hat{h}^2_{\bullet 2} = \text{GWASH}(\boldsymbol{X}_{\bullet 2}, \boldsymbol{y}_{\bullet 2}) \quad \text{or} \quad \text{GCTA}(\boldsymbol{X}_{\bullet 2}, \boldsymbol{y}_{\bullet 2}). \tag{20}$$

Moreover, the ratio λ_β in equation (18) can be easily estimated by the sample variance ratio $\hat{\lambda} = \hat{\sigma}^2_{y_{\bullet 2}}/\hat{\sigma}^2_y$, where $\hat{\sigma}^2_{y_{\bullet 2}}$ and $\hat{\sigma}^2_y$ denote the sample variances of $\boldsymbol{y}_{\bullet 2}$ and $\boldsymbol{y}$, respectively. Finally, by substituting the estimates $\hat{h}^2_1$, $\hat{h}^2_{\bullet 2}$, and $\hat{\lambda}$ into equation (18), we obtain the estimated FVE for all SNPs h^2_β, denoted by $\hat{h}^2$. A pseudo-algorithm outlining the estimation procedure for FVE decomposition is provided in Algorithm 1.

Algorithm 1 FVE Decomposition Estimation Procedure

Input: Response vector $\boldsymbol{y} \in \mathbb{R}^n$; SNP matrices $\boldsymbol{X}_1 \in \mathbb{R}^{n \times m_1}$, $\boldsymbol{X}_2 \in \mathbb{R}^{n \times m_2}$.

Output: Estimated total FVE $\hat{h}^2$.

Step 1: Estimation of $h^2_{1,\beta}$

Fit linear regression of $\boldsymbol{y}$ on $\boldsymbol{X}_1$ and compute adjusted R^2, $\hat{h}^2_1 \leftarrow$ adjusted R^2.

Step 2: Residualization

Compute projection matrix: $\boldsymbol{P}_1 = \boldsymbol{X}_1(\boldsymbol{X}_1^\top \boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top$; $\boldsymbol{X}_{\bullet 2} \leftarrow \boldsymbol{X}_2 - \boldsymbol{P}_1\boldsymbol{X}_2, \quad \boldsymbol{y}_{\bullet 2} \leftarrow \boldsymbol{y} - \boldsymbol{P}_1\boldsymbol{y}$.

Step 3: Estimate of $h^2_{\bullet 2,\beta}$

Apply high-dimensional FVE estimation method, e.g., GWASH or GCTA, on $\boldsymbol{X}_{\bullet 2}$ and $\boldsymbol{y}_{\bullet 2}$, $\hat{h}^2_{\bullet 2} \leftarrow \text{GWASH}(\boldsymbol{X}_{\bullet 2}, \boldsymbol{y}_{\bullet 2})$ or $\text{GCTA}(\boldsymbol{X}_{\bullet 2}, \boldsymbol{y}_{\bullet 2})$.

Step 4: Estimation of λ_β

Compute sample variances: $\hat{\sigma}^2_{y_{\bullet 2}} \leftarrow \text{Var}(\boldsymbol{y}_{\bullet 2}), \quad \hat{\sigma}^2_y \leftarrow \text{Var}(\boldsymbol{y})$; $\hat{\lambda} \leftarrow \hat{\sigma}^2_{y_{\bullet 2}}/\hat{\sigma}^2_y$.

Step 5: Estimation of h^2_β

The updated estimate: $\hat{h}^2 \leftarrow \hat{h}^2_1 + \hat{\lambda} \cdot \hat{h}^2_{\bullet 2}$

5 Testing the BKE Condition on Effect Coefficients

We consider now Model (1) and assume that $\beta_1, \ldots, \beta_m$ are IID with mean zero and variance of h^2/m. To more effectively evaluate whether the BKE condition holds, we formulate the following hypothesis test for the effects β_j's:

$$\begin{aligned} H_0 &: \kappa_\beta = 3 \\ H_1 &: \kappa_\beta > 3. \end{aligned} \tag{21}$$

In practice, this hypothesis is not directly testable because the true effects vector $\boldsymbol{\beta}$ is not observable. To address this limitation, we instead base the following proposition on the estimable correlation vector $\boldsymbol{\eta}$, as introduced in Equation (7), to test the hypothesis in terms of κ_β as stated in (21).

Proposition 5.1. *Consider Model (1) and assume the random effects $\boldsymbol{\beta} \sim N(0, h^2\boldsymbol{I}/m)$. Then the*

covariance matrix of the sample correlation vector $\hat{\boldsymbol{\eta}}$ given $\boldsymbol{X}$ is

$$\boldsymbol{V} = \mathrm{Var}[\hat{\boldsymbol{\eta}}|\boldsymbol{X}] = \frac{1-h^2}{n}\hat{\boldsymbol{\Sigma}} + \frac{h^2}{m}\hat{\boldsymbol{\Sigma}}^\top\hat{\boldsymbol{\Sigma}}, \tag{22}$$

where $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X}$ is the sample covariance matrix of $\boldsymbol{X}$. Furthermore, assuming errors $\boldsymbol{\varepsilon} \sim N(0, \sigma_\varepsilon^2\boldsymbol{I})$, the sample correlation vector $\hat{\boldsymbol{\eta}}$ follows a multivariate normal distribution:

$$\hat{\boldsymbol{\eta}}|\boldsymbol{X} \sim N(\boldsymbol{0}, \boldsymbol{V}). \tag{23}$$

The proof of this proposition can be found in Appendix B.4. Let $\boldsymbol{v} = (v_1, \ldots, v_m) = \mathrm{diag}(\boldsymbol{V})$, and $\tilde{\boldsymbol{\eta}} = (\tilde{\eta}_1, \ldots, \tilde{\eta}_m)$ denote the standardized correlation, where

$$\tilde{\eta}_j = \hat{\eta}_j/\sqrt{v_j}, \quad \text{for} \quad j = 1, \ldots, m. \tag{24}$$

While the correlation vector $\boldsymbol{\eta}$ has different variances, the standardized $\tilde{\boldsymbol{\eta}}$ have all variance 1 and are weakly correlated.

This proposition facilitates testing the null hypothesis that the BKE condition holds by examining whether the distribution of the standardized correlation scores $\tilde{\boldsymbol{\eta}}$ conforms to a standard normal distribution. To this end, we employ the Jarque-Bera test, a goodness-of-fit test that jointly evaluates whether the sample skewness and kurtosis match those of the standard normal distribution. Under the null hypothesis of the Jarque-Bera test, $\tilde{\boldsymbol{\eta}}$ follows a standard normal distribution, while the alternative hypothesis asserts that $\tilde{\boldsymbol{\eta}}$ deviates from normality. Combined with the results in Section 3.3, a large kurtosis in $\boldsymbol{\eta}$ generally implies elevated kurtosis in $\boldsymbol{\beta}$. Therefore, testing whether $\tilde{\boldsymbol{\eta}}$ has kurtosis equal to 3 provides indirect evidence regarding departures from the BKE condition, i.e., whether $\kappa_\beta = 3$.

6 Numerical Studies

In this section, we conduct numerical experiments to evaluate our proposed framework, including decomposed FVE estimation, testing for the BKE condition, and screening methods. The structure of this section is as follows. Section 6.1 introduces a numerical approach to control the kurtosis of $\boldsymbol{\beta}$, which we then use in all the consequent simulations. In Section 6.2, we examine the relationship between κ_β and κ_{η_j}, and empirically quantify the findings described in Proposition 3.2. Section 6.3 evaluates the performance of the decomposed FVE estimator. Section 6.4 assesses the performance of the BKE testing procedure. Finally, Section 6.5 investigates the effectiveness of several existing screening methods (detailed in Appendix A.2) and provides a comparative analysis of their performance. A repository containing the code used to run the simulation analyses in this section is available at: `https://github.com/Fangn06/Decomposed-FVE-Estiamtion`.

6.1 Controlling Kurtosis of Effects in Simulation Studies

Due to sampling variability, directly simulating $\boldsymbol{\beta}$ from the Gaussian mixture model in Equation (13) makes it difficult to achieve the target kurtosis. To address this issue, we propose a data generation mechanism that ensures the empirical kurtosis of the simulated $\boldsymbol{\beta}$ aligns with the specified value in the simulation design.

The basic idea of the mechanism is that we decompose $\boldsymbol{\beta}$ into two components $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ as described in Section 3.2. The strong effects $\boldsymbol{\beta}_1 = (\beta_1, \ldots, \beta_{m_1})$ having a common variance ψ^2, and the weak effects $\boldsymbol{\beta}_2 = (\beta_{m_1+1}, \ldots, \beta_m)$ having a common variance γ^2. Then we can find appropriate ψ^2 and γ^2 to ensure the $\boldsymbol{\beta}$ has the desired kurtosis. Based on the derivation of κ_β in (14), we can first obtain $\psi^2 = (1-\theta)h^2/[m(1-p)]$ with the given θ, h^2 and p. Then we generate $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ from the standard normal distribution and then standardize them to ensure these two components have a mean of 0 and variance of 1, respectively. We assume the true Kurt$[\boldsymbol{\beta}]$ is K and the classic formula of kurtosis:

$$K = m\frac{\sum_{j=1}^{m}(\beta_j - \bar{\beta})^4}{(\sum_{j=1}^{m}(\beta_j - \bar{\beta})^2)^2},$$

where $\bar{\beta} = \sum_{j=1}^{m} \beta_j/m$. Since $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ have been centered and scaled, we can solve the following equations to approximate γ^2:

$$K = m\frac{\psi^4 \sum_{j=1}^{m_1} \beta_j^4 + \gamma^4 \sum_{j=m_1+1}^{m_2} \beta_j^4}{(\psi^2 \sum_{j=1}^{m_1} \beta_j^2 + \gamma^2 \sum_{j=m_1+1}^{m_2} \beta_j^2)^2},$$

where $m_1 + m_2 = m$. This equation is a quadratic equation and can be solved either with numeric optimization or the quadratic formula $\gamma^2 = (-b \pm \sqrt{b^2 - 4ac})/2a$, where

$$a = K\left(\sum_{j=m_1+1}^{m_2} \beta_j^2\right)^2 - m\sum_{j=m_1+1}^{m_2} \beta_j^4,$$

$$b = 2K\psi^2 \sum_{j=1}^{m_1} \beta_j^2 \sum_{j=m_1+1}^{m_2} \beta_j^2,$$

$$c = K\psi^4\left(\sum_{j=1}^{m_1} \beta_j^2\right)^2 - m\psi^4 \sum_{j=1}^{m_1} \beta_j^4.$$

Here we should only consider the positive solution of γ^2. Then the generated $\boldsymbol{\beta}$ with specific kurtosis can be expressed as $\boldsymbol{\beta} = (\psi\boldsymbol{\beta}_1, \gamma\boldsymbol{\beta}_2)$.

6.2 Simulation Study I: Relationship Between Kurt$[\boldsymbol{\beta}]$ and Kurt$[\boldsymbol{\eta}]$

To demonstrate the relationship between Kurt$[\boldsymbol{\beta}]$ and Kurt$[\boldsymbol{\eta}]$, we conducted a series of simulations. We varied the AR(1) correlation parameter ρ of $\boldsymbol{\Sigma}$ across the values $\rho = 0, 0.5, 0.9$. The kurtosis of $\boldsymbol{\beta}$ were set to 3, 5, 20, 50, and 100, controlled using the method described in Section 6.1. We set

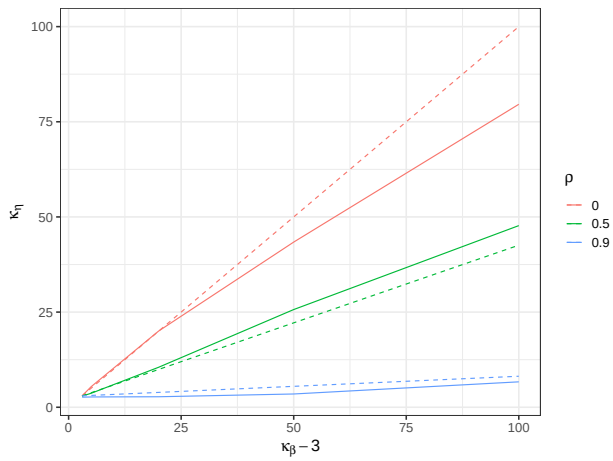

Figure 1: Relationship between κ_β and κ_{η_j}. The solid line depicts the theoretical kurtosis of $\boldsymbol{\eta}$ as a function of the kurtosis of $\boldsymbol{\beta}$, computed from Equation (16). The dashed line shows the empirical kurtosis of $\boldsymbol{\eta}$ based on simulated $\boldsymbol{\beta}$ with the corresponding kurtosis.

$m = 400$, $n = 200$, $\theta = 0.9$, $h^2 = 0.5$, $\sigma_\varepsilon^2 = 0.5$. The indices of strong effects are assigned to be from $j = 201$ to $j = 210$.

In this simulation, for each specified kurtosis level of $\boldsymbol{\beta}$, we first compute the theoretical kurtosis of $\boldsymbol{\eta}$ using Equation (16), which is depicted as the solid line in Figure 1. We then generate $\boldsymbol{\beta}$ with the target kurtosis using the method described in Section 6.1, compute $\boldsymbol{\eta} = \boldsymbol{\Sigma}\boldsymbol{\beta}$ based on the given $\boldsymbol{\Sigma}$ and $\boldsymbol{\beta}$, and compute the sample kurtosis of $\boldsymbol{\eta}$. This sample kurtosis is treated as the empirical kurtosis of $\boldsymbol{\eta}$, shown as the dashed line in Figure 1.

It is worth noting that the theoretical ratio exhibits a boundary effect for $\rho = 0.5$ and 0.9 when the index j is close to 0 or m. In contrast, the ratio remains stable for interior values of j. To avoid this boundary effect, we compute the theoretical kurtosis of η_j at $j = m/2$ using Equation (16). For the empirical vector $\boldsymbol{\eta}$, we extract a subset from $j = 50$ to $j = 350$ and use this subset to calculate its empirical kurtosis.

The final values of κ_{η_j} were averaged over 500 simulation iterationsn for each given κ_β. The results are presented in Figure 1, where the dashed line represents the theoretical κ_{η_j} derived from Proposition 3.1, and the solid lines show the empirically averaged κ_{η_j} from the simulation. We can observe that for different ρ, the empirical results match the trend of the theoretical predictions. These results illustrate the relationship between κ_β and κ_{η_j} under varying correlation levels. When $\rho = 0$ or 0.5, κ_{η_j} closely tracks κ_β, with larger values of κ_{η_j} consistently corresponding to larger values of κ_β. For $\rho = 0.9$, although the increasing trend becomes less pronounced, the two quantities still exhibit a similar overall pattern.

6.3 Simulation study II: Decomposed FVE estimator

Since the overall framework of the proposed decomposed FVE estimation is not limited to high-dimensional settings, we provide results in both low and high-dimensional settings and evaluate

their performance accordingly. In the low-dimension setting, we vary $m = 50, 100, 150$ and fix the ratio between m and n as $n = 5m$. We change the kurtosis such that $\kappa_\beta = 3, 5, 10, 20, 30$ for each m. The strong effects are the first 5 coefficients in this low-dimensional setting, resulting in $p = 5/m$, and we set $\theta = 0.9$ and $h^2 = 0.5$. In the high-dimension setting, we vary $m = 400, 800, 1200$ and fix the ratio between m and n as $n = m/2$. We change the kurtosis such that $\kappa_\beta = 3, 5, 20, 50, 100$ for each m. The strong effects are the first 10 coefficients in this high-dimensional setting, resulting in $p = 10/m$, and we keep $\theta = 0.9$ and $h^2 = 0.5$. For both scenarios, the distribution of $\boldsymbol{X}$ is assumed to follow a multivariate Gaussian with the correlation structure being AR(1) ($\rho = 0, 0.5, 0.9$). The number of simulation instances is 1000 in the following studies.

Due to sampling variability, the realized FVE, denoted as h^2_β in Equation (2), may deviate from the target FVE h^2 specified in the simulation setup. Moreover, fluctuations in the realized FVE may occur across repeated simulation iterations. Although Section 6.1 presents a method for controlling the kurtosis of $\boldsymbol{\beta}$, we find that it is not possible to simultaneously control both the kurtosis of $\boldsymbol{\beta}$ and the realized FVE. To quantify the accuracy and stability of the simulations, we use the relative root mean square error (RMSE) as a performance metric. As in the Equation (2), let

$$h^2_{\beta,i} = \frac{\boldsymbol{\beta}_i^\top \boldsymbol{\Sigma} \boldsymbol{\beta}_i}{\boldsymbol{\beta}_i^\top \boldsymbol{\Sigma} \boldsymbol{\beta}_i + \sigma^2_\varepsilon}$$

represent the realized FVE conditional on $\boldsymbol{\beta}_i$ in the i-th simulation realization, and $\hat{h}^2_i$ denote the estimated FVE in the same realization, for $i = 1, \ldots, M$, where M denotes the total number of simulation realizations. The relative RMSE of the estimators is then calculated as:

$$\text{Relative RMSE}(\hat{h}^2) = \sqrt{\frac{1}{M} \sum_{i=1}^{M} \left(\frac{\hat{h}^2_i - h^2_{\beta,i}}{h^2_{\beta,i}} \right)^2}.$$

Similarly, we quantify the relative bias as

$$\text{Relative BIAS}(\hat{h}^2) = \frac{1}{M} \sum_{i=1}^{M} \left(\frac{\hat{h}^2_i - h^2_{\beta,i}}{h^2_{\beta,i}} \right).$$

In addition to GWASH, we applied Genome-wide Complex Trait Analysis GCTA [Yang et al., 2011] combined with our proposed FVE decomposition method. To implement our FVE decomposition, we estimated the explained variance in the low-dimensional component, $\hat{h}^2_1$ in Equation (19), using the adjusted R^2, while the remaining component, $\hat{h}^2_{\bullet 2}$ in Equation (20), was estimated using both GWASH and GCTA.

The estimation results of the decomposed GWASH estimator compared to the original GWASH estimator are presented in Figure 2 and Figure 14 in Appendix A.5. For both low- and high-dimensional settings, we observe that larger values of m lead to lower relative RMSE, while increasing kurtosis in $\boldsymbol{\beta}$ results in higher RMSE. Across all settings, the decomposed estimator consistently achieves lower RMSE than the original GWASH estimator, with the improvement

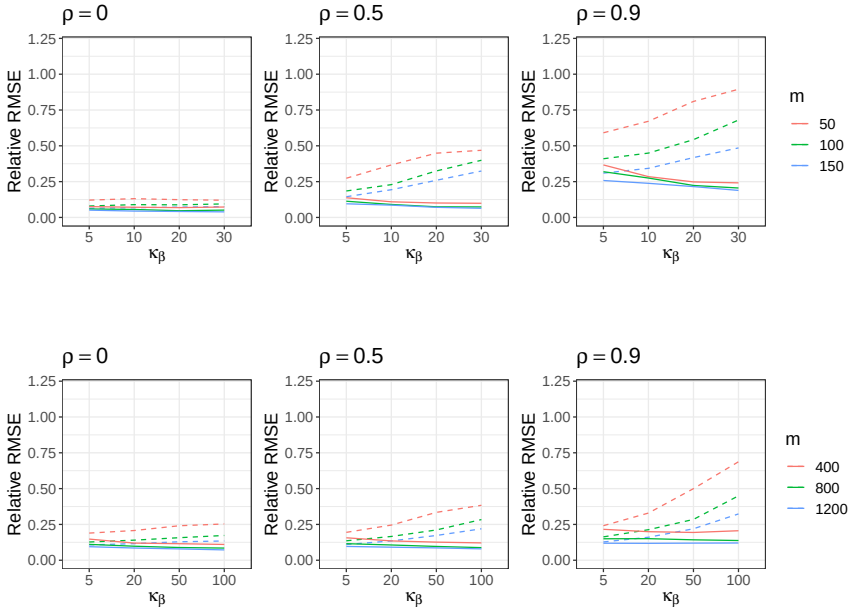

Figure 2: Estimation performance of GWASH without decomposition (dashed line) and decomposed GWASH (solid line) in the low-dimensional setting (top row) and high-dimensional setting (bottom row). The standard error from 1000 simulation instances is about 0.0003 (low-dimension) and 0.0007 (high-dimension).

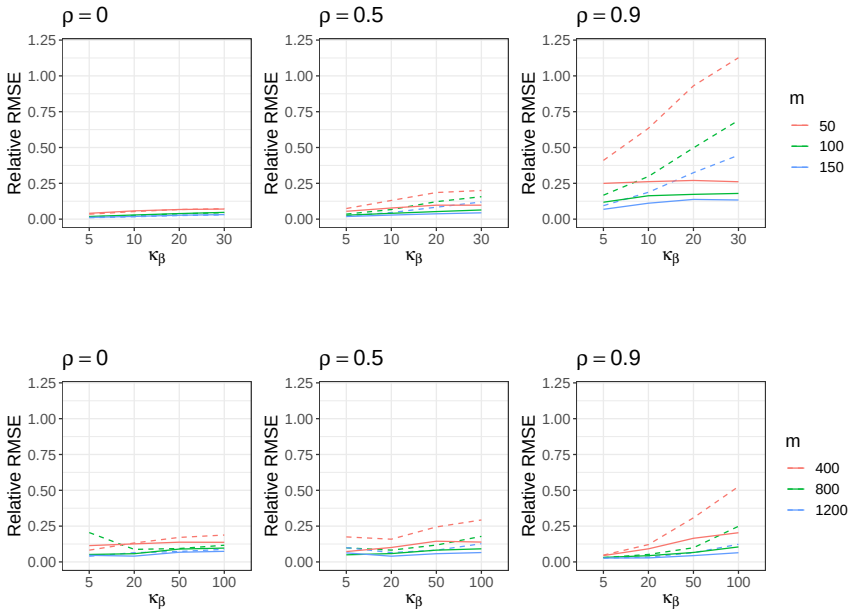

Figure 3: Estimation performance of the GCTA without decomposition (dashed line) and decomposed GCTA (solid line) under the high-dimension setting. The standard error from 1000 simulation instances is about 0.00015 (low-dimension) and 0.0002 (high-dimension).

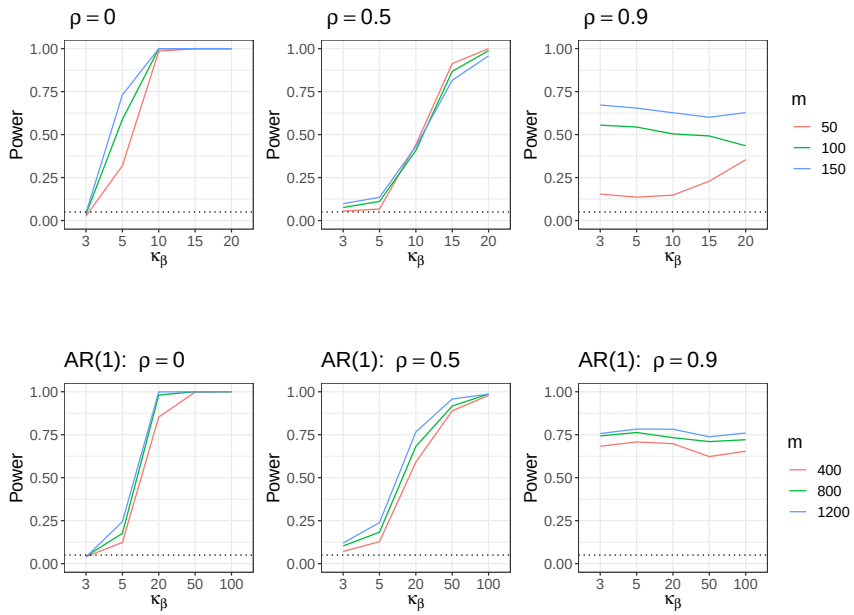

Figure 4: Performance of testing of the BKE condition in the low-dimensional setting (top row) and high-dimensional setting (bottom row). The Type-I error is the value of the power at $\kappa_\beta = 3$.

being more pronounced at larger values of ρ. These results highlight the decomposed estimator's superior ability to mitigate the adverse effects of large kurtosis, which can otherwise violate the BKE assumptions underlying the original GWASH approach.

The estimation results of the decomposed GCTA estimator, in comparison to the original GCTA estimator, are presented in Figure 3 for relative RMSE, and in Figure 15 in Appendix A.5 for relative bias. The performance of the GCTA-based decomposition closely mirrors that of the GWASH-based decomposition. In high-dimensional settings, the decomposed estimator does not exhibit a clear advantage over the original GCTA estimator when the kurtosis of $\boldsymbol{\beta}$ is relatively low ($\kappa_\beta = 5$). However, as the kurtosis increases ($\kappa_\beta \geq 20$), the decomposed estimator achieves a substantially lower RMSE. These findings suggest that GCTA is sensitive to heavy-tailed distributions and that our proposed decomposition framework can help mitigate this issue.

6.4 Simulation Study III: Testing the BKE Condition

We evaluated the performance of the proposed test as formulated in Equation (21) and described in Section 5, under both low- and high-dimensional settings. The results are presented in Figure 4. In both cases, the Type I error rate (under $\kappa_\beta = 3$) remains close to the nominal level of 0.05 across different values of m for $\rho = 0$ and 0.5. The power increases with higher kurtosis, and larger values of m generally lead to greater power. Notably, the test achieves power close to 1 when $\kappa_\beta = 20$ in the low-dimensional setting and $\kappa_\beta = 100$ in the high-dimensional setting.

However, under strong correlation ($\rho = 0.9$), the test fails to adequately control the Type I error,

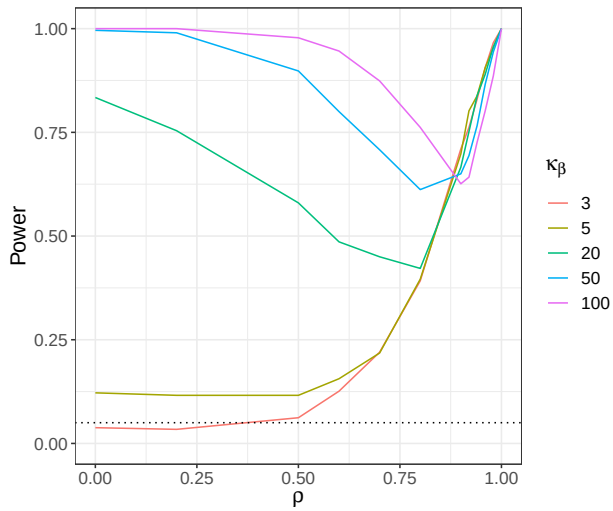

Figure 5: Performance of testing of the BKE condition with varying ρ and κ_β.

and the power exhibits a different pattern compared to the cases with $\rho = 0$ and $\rho = 0.5$. To further examine how the correlation level ρ influences testing performance across varying levels of κ_β, we conduct an additional set of simulation studies. In this experiment, we fix $m = 400$ and $n = 200$, and vary $\rho \in \{0, 0.2, 0.5, 0.6, 0.7, 0.8, 0.9, 0.92, 0.94, 0.96, 0.98, 1\}$ and $\kappa_\beta \in \{3, 5, 20, 50, 100\}$. The power under each scenario is shown in Figure 5.

We observe that under weak correlation ($\rho \leq 0.5$), the Type I error ($\kappa_\beta = 3$) and power remain stable for small κ_β ($\kappa_\beta = 5$), while the power decreases with increasing ρ for larger values of $\kappa_\beta \geq 20$. When $\rho > 0.5$, the Type I error and power for small κ_β increases as ρ increases. In contrast, for large κ_β, the power continues to decline until ρ reaches approximately 0.8–0.9, after which it begins to rise again with further increases in ρ.

6.5 Simulation study IV: Screening Methods for Identifying Strong Localized Effects

In this study, we fix $m = 400$ and the sample size at $n = m/2$. The correlation matrix $\mathbf{\Sigma}$ followed an AR(1) structure with $\rho \in \{0, 0.5, 0.9\}$. We varied the kurtosis to achieve $\kappa_\beta \in \{5, 20, 50, 100\}$ and compared the five screening methods described in Appendix A.2. To implement the screening procedure more appropriately and to avoid overfitting, we first split the data into two parts: 15% of the original sample is used as the screening set to identify strong effects, and the remaining 85% is reserved as the estimation set.

The results of the estimation using different screening methods are presented in Figure 6. For both GWASH and GCTA, the estimation without any screening method shows the largest RMSE, while the oracle has the smallest RMSE, reflecting its superior estimation performance due to the use of true indices for strong effects. For GWASH, when κ_β is relatively small, all screening methods perform similarly. As κ_β increases, some divergence among the screening methods emerges, although

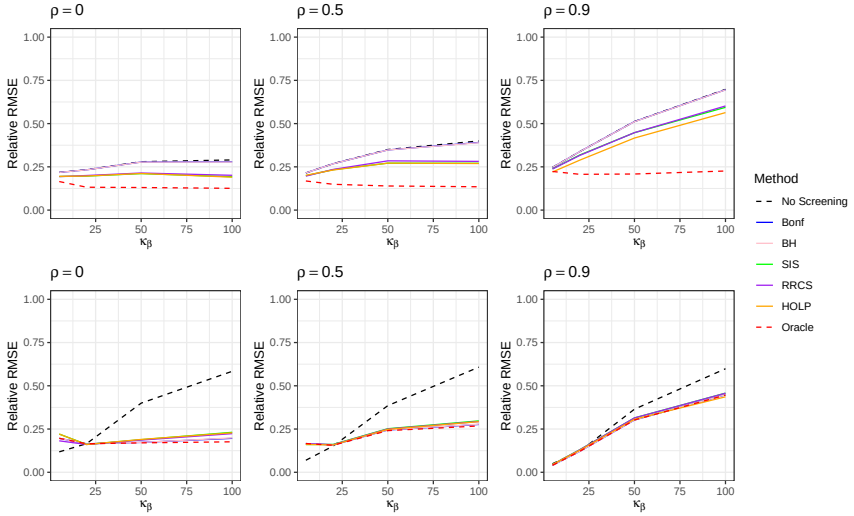

Figure 6: Estimation performance across different screening methods. Results for GWASH are displayed in the top panel, and results for GCTA are displayed in the bottom panel. The standard error from 1000 simulation instances are about 0.001 (GWASH) and 0.0007 (GCTA).

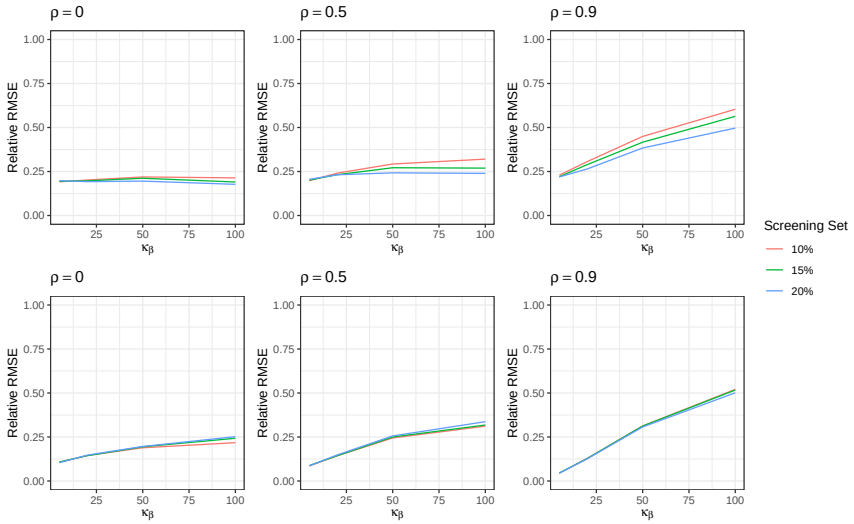

Figure 7: Sensitivity analysis of the HOLP screening procedure using a screening set comprising {10%, 15%, 20%} of the original sample. Results for GWASH are displayed in the top panel, and results for GCTA are displayed in the bottom panel. The standard error from the 1000 simulation instances are about 0.001 for GWASH and GCTA.

the differences remain modest. For GCTA, the differences in RMSE across the various methods are minor.

We further investigate the sensitivity of the screening methods to the choice of screening set size. As an illustrative example, we focus on the HOLP method, which performs best among all screening methods according to Figure 6. We vary the screening set size from 10%, 15%, and 20% of the original sample, and the corresponding estimation performance is shown in Figure 7. Following the recommendation of Wang and Leng [2016], we implemented the default rule for determining the number of selected effects, $d_s = \lfloor n_s / \log n_s \rfloor$, where n_s denotes the sample size of the screening set. Both GWASH and GCTA display sensitivity to the choice of screening set size, with the effect being more pronounced for GWASH, particularly when κ_β is large, whereas GCTA exhibits relatively milder sensitivity.

7 Real Data Analysis

In this section, we apply our proposed FVE decomposition framework to the Adolescent Brain Cognitive Development Study (ABCD; Casey et al. [2018]) to investigate the SNP heritability of our outcome of interest, the PolyVoxel Score (PVS; Loughnan et al. [2022, 2024]), which quantifies the degrees to which an individual's brain resembles the archetypal hemochromatosis brain. We begin by describing the data preprocessing procedures used in our analysis. Next, we present the results of chromosome-level heritability tests, followed by the GWASH and GCTA-based decomposed FVE estimates derived using prior knowledge of strong SNP effects. We then implement a data-driven screening approach to identify strong SNP effects and compute the corresponding decomposed FVE estimates with GWASH and GCTA. Finally, we conduct a sensitivity analysis by varying the size of the screening set used to identify strong effects, using the remaining data for estimation, and compare the estimation results across different sample sizes.

7.1 Data Description and Pre-processing

The ABCD Study provides a comprehensive resource for investigating factors related to child development and mental health. Its genotype data have been widely used in genetic research, including GWAS, with details of the genetic data are available in Fan et al. [2023]. The PolyVoxel Score (PVS) is an imaging-derived measure that quantifies the extent to which an individual's brain resembles the archetypal hemochromatosis Brain, characterized by regional iron accumulation in motor circuits. Developed using T2-weighted MRI scans, the PVS captures heritable variation associated with brain iron homeostasis. Comprehensive information on the development of the PVS and its integration with the ABCD data is provided in Loughnan et al. [2024].

In our analysis, we included several demographic covariates: age; self-reported ethnicity (categorized as White, Black, Asian, Hispanic, or Other); household income (categorized as "less than \$50K," "\$50K–\$100K," and "greater than or equal to \$100K"); sex (male or female); proportions of African, European, East Asian, and American genetic ancestry; and the first 10 genetic principal

components. The principal components were calculated by Fan et al. [2023] and are included in the analysis following the recommendations of Luo et al. [2021]. In the subsequent analysis, we residualized both the outcome and SNPs on these covariates to remove the population structure, i.e., we regressed the PVS and each SNP on these covariates and used the resulting residuals. A more detailed description of the covariate assessment and processing is provided in Section 2.8.4 of Singh Sachan [2025].

A selection procedure for individuals and SNPs was conducted following Singh Sachan [2025] Section 2.8.3. Due to the longitudinal design of the ABCD Study, we restricted the analysis to baseline observations. For participants with multiple observations, only their baseline data were retained, and individuals missing key variables were excluded. For SNPs, we used imputed genotype data provided by the ABCD Study using the TOPMed imputation server [Taliun et al., 2021], with details described in Fan et al. [2023]. SNPs with a minor allele frequency below 5%, multiallelic variants, those failing Hardy–Weinberg equilibrium, non-autosomal SNPs, and SNPs within the MHC region were excluded. Additional quality control details are provided in Singh Sachan [2025]. The final dataset in our analysis included $n = 6{,}623$ individuals and $m = 750{,}032$ SNPs.

7.2 FVE Estimation without Decomposition

We begin by estimating the FVE using GWASH and GCTA without decomposition, applied both to the whole genome and to each of the 22 chromosomes separately. Following the procedure in Section 2.2, the estimated FVE for the k-th chromosome is denoted as $\hat{h}^2_{\text{GWASH},k}$, where $k = 1, \ldots, K = 22$. The standard error of each GWASH estimate can then be computed using the formula from Schwartzman et al. [2019]:

$$\text{SE}_{\text{GWASH},k} = \frac{2}{n}\left(\frac{m_k}{n\hat{\mu}_{2,k}} + 2\frac{\hat{\mu}_{3,k}}{\hat{\mu}^2_{2,k}}\hat{h}^2_k - \hat{h}^4_k\right), \tag{25}$$

where m_k, $\hat{\mu}_{2,k}$, $\hat{\mu}_{3,k}$, and $\hat{h}^2_k$ represent the number of SNPs, estimated second spectral moment μ_2, third spectral moment μ_3, and FVE h^2 for the k-th chromosome, respectively. The GCTA software directly reports the corresponding standard errors in its output.

Following Proposition 1 in Schwartzman et al. [2019], we compute the total SNP heritability estimate with GWASH across the whole genome as

$$\hat{h}^2_{\text{GWASH}} = \sum_{k=1}^{K} \frac{\hat{\mu}_{2,k}}{\hat{\mu}_2}\hat{h}^2_{\text{GWASH},k}. \tag{26}$$

Here, chromosomes are assumed to be independent of each other. Based on this assumption, the genome-wide second and third spectral moments can be estimated as

$$\hat{\mu}_2 = \sum_{k=1}^{K} \frac{m_k}{m}\hat{\mu}_{2,k}, \quad \hat{\mu}_3 = \sum_{k=1}^{K} \frac{m_k}{m}\hat{\mu}_{3,k}, \tag{27}$$

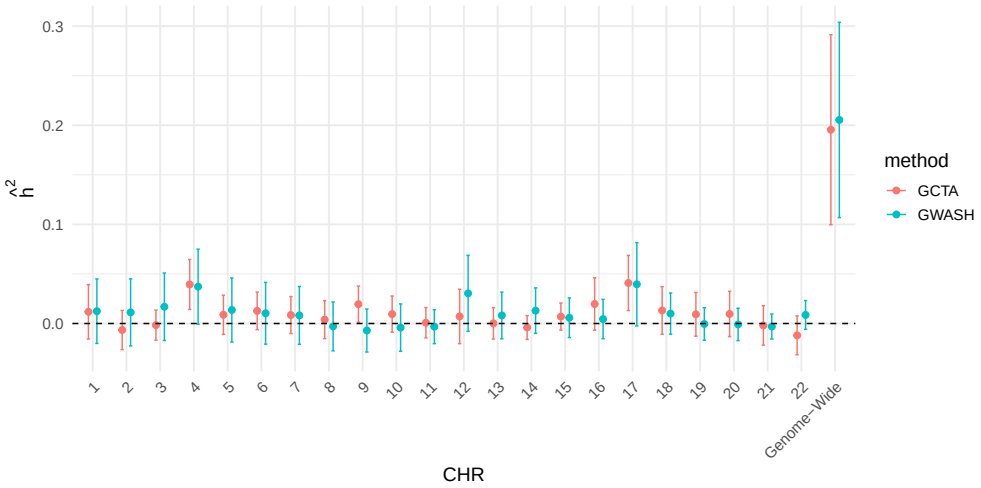

Figure 8: Chromosome-specific point estimates of FVE with 95% confidence intervals, obtained using GWASH and GCTA without decomposition.

where m_k is the number of SNPs on the k-th chromosome and $m = \sum_{k=1}^{K} m_k$ is the total number of SNPs across the genome. The standard error of the genome-wide $\hat{h}^2_{\text{GWASH}}$ is then given by

$$\text{SE}_{\text{GWASH}} = \frac{2}{n}\left(\frac{m}{n\hat{\mu}_2} + 2\frac{\hat{\mu}_3}{\hat{\mu}_2^2}\hat{h}^2 - \hat{h}^4\right). \tag{28}$$

For the estimation of genome-wide FVE using GCTA, the variance can be partitioned across the 22 chromosomes, after which GCTA is applied following the approach outlined in Yang et al. [2011].

The chromosome-specific estimates of $\hat{h}^2_{\text{GWASH},k}$ and the genome-wide FVE estimate $\hat{h}^2_{\text{GWASH}}$, along with their 95% confidence intervals, are summarized in Figure 8. None of the chromosome-specific estimates are statistically distinguishable from zero, with several point estimates even taking negative values. In contrast, the genome-wide estimate is statistically significant. We further applied GCTA to estimate the FVE for each chromosome, as shown in Figure 8. For most chromosomes, as well as the genome-wide estimates, the GCTA results were similar to those from GWASH.

7.3 Testing the BKE Condition on SNP Effects

We applied the proposed BKE condition test to each chromosome, using a significance level of $\alpha = 0.05$ with Bonferroni correction to account for multiple testing across the 22 chromosomes. The results, presented in Table 1, show that at least 7 chromosomes provide evidence against satisfying the BKE condition. We also performed the test on the whole genome, obtaining a p-value of 0.6797. This global result aligns with the chromosome-level findings, as the majority of chromosomes (15 out of 22) individually appear to satisfy the BKE condition. Thus, the chromosome-level analysis indicates that violations of the BKE condition may occur in specific regions of the genome, even when the global test fails to detect them.

Table 1: Bonferroni adjusted p-values for testing the BKE condition on each chromosome.

CHR	p-value	Reject H_0	CHR	p-value	Reject H_0
1	3.34×10^{-3}	True	12	1	False
2	1	False	13	1	False
3	4.92×10^{-3}	True	14	1	False
4	1	False	15	1	False
5	4.83×10^{-9}	True	16	3.07×10^{-2}	True
6	1	False	17	1.00×10^{-1}	False
7	1	False	18	8.20×10^{-1}	False
8	1.59×10^{-1}	False	19	8.68×10^{-2}	False
9	1	False	20	8.71×10^{-6}	True
10	3.67×10^{-4}	True	21	3.52×10^{-6}	True
11	1	False	22	1.29×10^{-1}	False

7.4 Decomposed FVE Estimation Based on Prior Knowledge

Loughnan et al. [2024] has identified significant SNPs associated with the PVS outcome using a subsample of the UK Biobank. These identified SNPs may exhibit similar patterns in the ABCD dataset. In our analysis, we selected the SNPs that were identified in the UK Biobank and also present in the ABCD dataset as strong-effect SNPs contributing to the PVS. In total, there are 39 such SNPs spanning 18 chromosomes (no significant SNPs were identified on chromosomes 7, 15, 16, and 21 in the ABCD data). A detailed list of these SNPs is available in the supplementary materials of Loughnan et al. [2024]. Since we have not yet conducted any data-driven screening procedure at this stage of the analysis, we treat these SNPs as prior knowledge.

We grouped these SNPs selected based on prior knowledge by chromosome. For each chromosome, we first estimated the adjusted R^2 attributable to the identified SNPs. We then regressed out these SNPs from both the remaining SNPs and the PVS outcome, following the procedure outlined in Algorithm 1. Using the residualized SNP matrix and outcome, we estimated $\hat{h}^2_{\bullet 2}$ via GWASH and GCTA, along with the corresponding variance correction factor $\hat{\lambda}$. The final estimated FVE for each chromosome was then computed as $\hat{h}^2 = \hat{h}^2_1 + \hat{\lambda} \cdot \hat{h}^2_{\bullet 2}$. We also constructed 95% confidence intervals for $\hat{h}^2_1$ and $\hat{h}^2_{\bullet 2}$, with standard errors derived following the approaches of Cramer [1987], Yang et al. [2011] and Equation (25).

The chromosome-specific decomposed FVE estimates obtained using GWASH and GCTA are summarized in Figure 9. Based on prior knowledge, strong effects were not identified on chromosomes 7, 15, 16, and 21. For these chromosomes, we treat their $\hat{h}^2_1 = 0$ in the top panel, and their FVE estimates without decomposition from GWASH and GCTA are shown in the middle panel of Figure 9. We also computed the genome-wide estimates for the two components, $\hat{h}^2_1$ and $\hat{\lambda} \times \hat{h}^2_{\bullet 2}$, as well as the total $\hat{h}^2 = \hat{h}^2_1 + \hat{h}^2_{\bullet 2}$. From the top panel, we observe that the 95% confidence interval of the genome-wide estimate of $\hat{h}^2_1$ only slightly crosses zero, suggesting potential statistical significance. In the middle panel, GCTA and GWASH produced broadly consistent results. Comparing the decomposed FVE estimates with those obtained without decomposition, we find that the differences

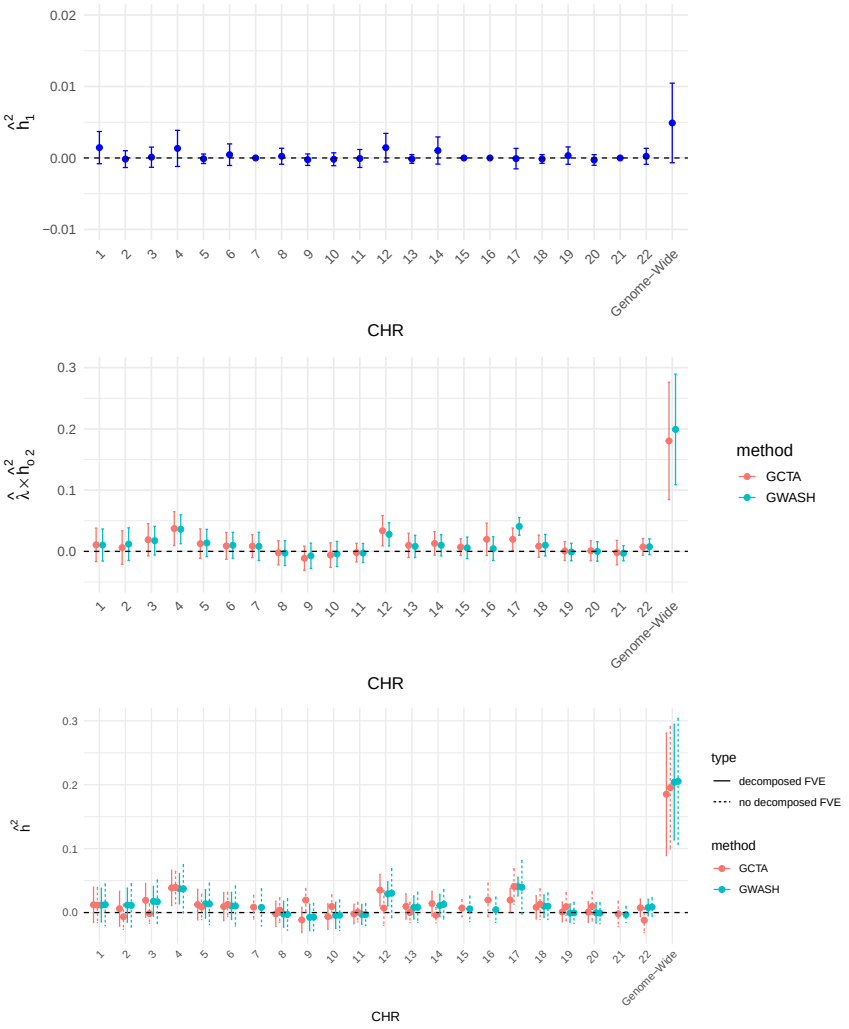

Figure 9: Chromosome-specific decomposed FVE estimates based on prior knowledge of 39 SNPs identified by previous studies as having strong effects. Top panel: FVE attributable to the 39 known SNPs. Middle panel: Remaining high-dimensional FVE after removing the effect of the known 39. Bottom panel: Total FVE estimates without and with decomposition. Note that the y-axis scales differ across the three panels.

are negligible. This outcome is likely because only 39 SNPs out of approximately 750,000 were selected, suggesting that their overall contribution is limited.

7.5 Decomposed FVE Estimation with Screening Methods

Besides leveraging prior knowledge to identify strong SNP effects, we also employed data-driven screening methods. In particular, we first applied the BH-based procedure with $\alpha = 0.2$, using 15% of the original sample as the screening set and the remaining samples for estimation to mitigate overfitting. This approach identified 52 SNPs across 7 chromosomes, as shown in Figure 10. For chromosomes without identified strong effects, the estimated low-dimensional FVE is set to $\hat{h}_1^2 = 0$, while the estimated FVE of the remaining high-dimensional component is taken to be the estimated FVE obtained without decomposition. As before, GCTA and GWASH produced broadly consistent estimates of the FVE in the high-dimensional component. However, when comparing decomposed FVE estimates with those obtained without decomposition, notable discrepancies emerge for chromosomes 8, 9, 14, and 22 under GCTA, whereas for GWASH, the decomposed and non-decomposed estimates differ only trivially.

As an alternative method, we applied the HOLP method to each chromosome in the ABCD dataset to identify strong SNP effects. Again, we partitioned the data into a screening set (15%) and an estimation set (85%) to mitigate potential overfitting. HOLP was applied exclusively to the screening set to select SNPs, and decomposed FVE estimates were subsequently obtained using the selected SNPs in the estimation set. Following the recommendation of Wang and Leng [2016], we implemented the default rule for determining the number of selected SNPs, $d_s = \lfloor n_s / \log n_s \rfloor$, where n_s denotes the screening sample size. Under this rule, with 15% of the original sample used as the screening set, we obtained $d_s = 143$. The estimation results are presented in Figure 11. It can be observed that the differences between GCTA and GWASH are minor for both chromosome-level and genome-wide FVE estimates. Moreover, when comparing decomposed FVE estimates with those obtained without decomposition, the decomposed estimates exhibit clear discrepancies both at the chromosome level and for the genome-wide estimates. This pattern indicates that the identified strong effects contribute meaningfully to the PVS, and that the proposed decomposed FVE estimation procedure provides a way for incorporating such effects into the overall estimation.

7.6 Sensitivity Analysis

The choice of screening set size represents a trade-off between effective SNP screening and reliable FVE estimation. While a larger estimation set enhances the precision of FVE estimates, a screening set that is too small may lack sufficient statistical power to identify SNPs with strong effects. To assess this balance, we conducted a sensitivity analysis for both BH and HOLP by varying the screening–estimation split across proportions $10\%, 15\%, 20\%$ for screening and the remaining $90\%, 85\%, 80\%$ for estimation, respectively. The results for BH are presented in Figure 12. Notably, the set of 52 SNPs identified across different screening proportions remains unchanged, consistent with the results reported in the previous section. Moreover, Figure 12 shows that the FVE estimation

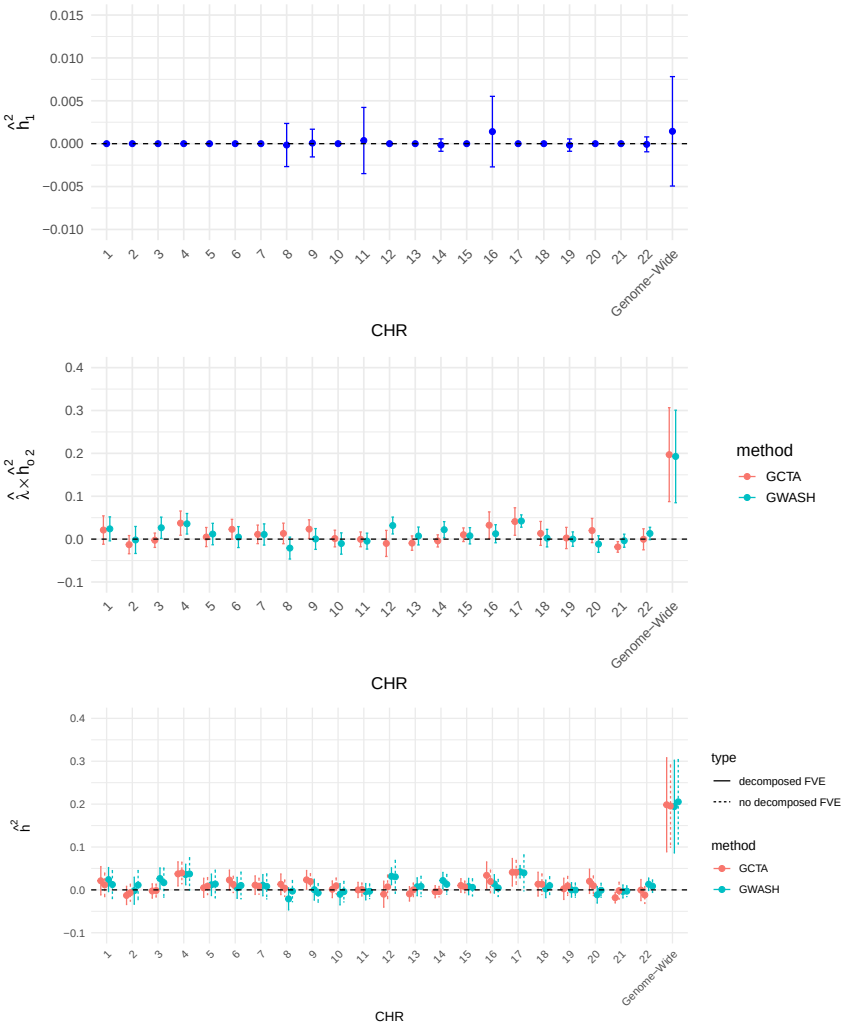

Figure 10: Chromosome-specific decomposed FVE estimates based on strong effects identified by the BH screening method using a screening set comprising 15% of the original sample. Top panel: FVE of the screened SNPs. Middle panel: Remaining high-dimensional FVE after removing the effect of the screened SNPs. Bottom panel: Total FVE estimates without and with decomposition. Note that the y-axis scales differ across the three panels.

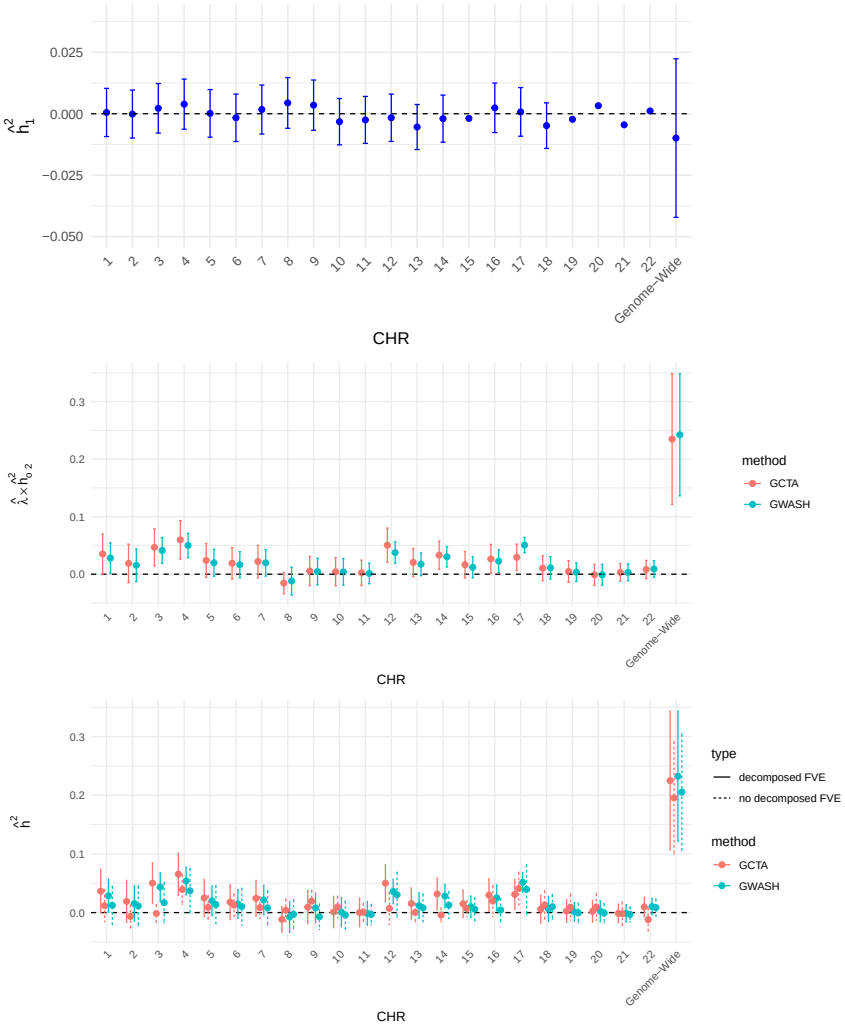

Figure 11: Chromosome-specific decomposed FVE estimates based on strong effects identified by the HOLP screening method using a screening set comprising 15% of the original sample. Top panel: FVE of the screened SNPs. Middle panel: Remaining high-dimensional FVE after removing the effect of the screened SNPs. Bottom panel: Total FVE estimates without and with decomposition. Note that the y-axis scales differ across the three panels.

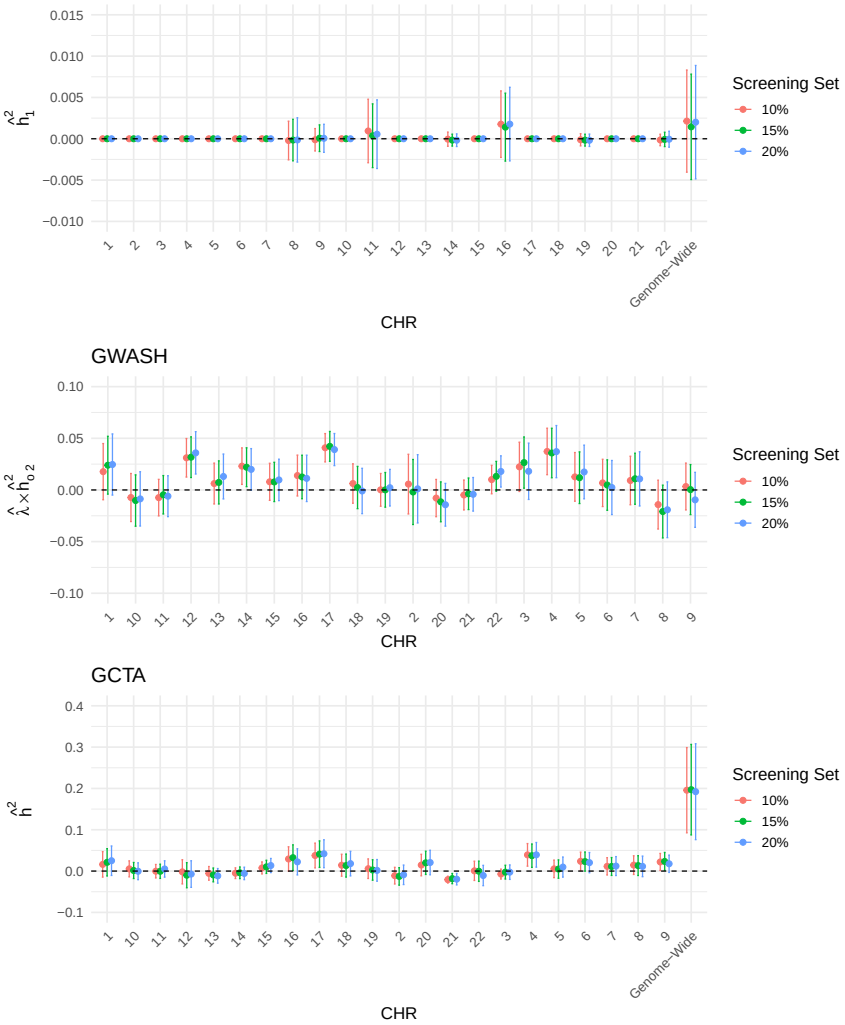

Figure 12: Sensitivity analysis of chromosome-specific decomposed FVE estimates based on strong effects identified by the BH screening method using a screening set comprising $\{10\%, 15\%, 20\%\}$ of the original sample. Note that the y-axis scales differ across the three panels.

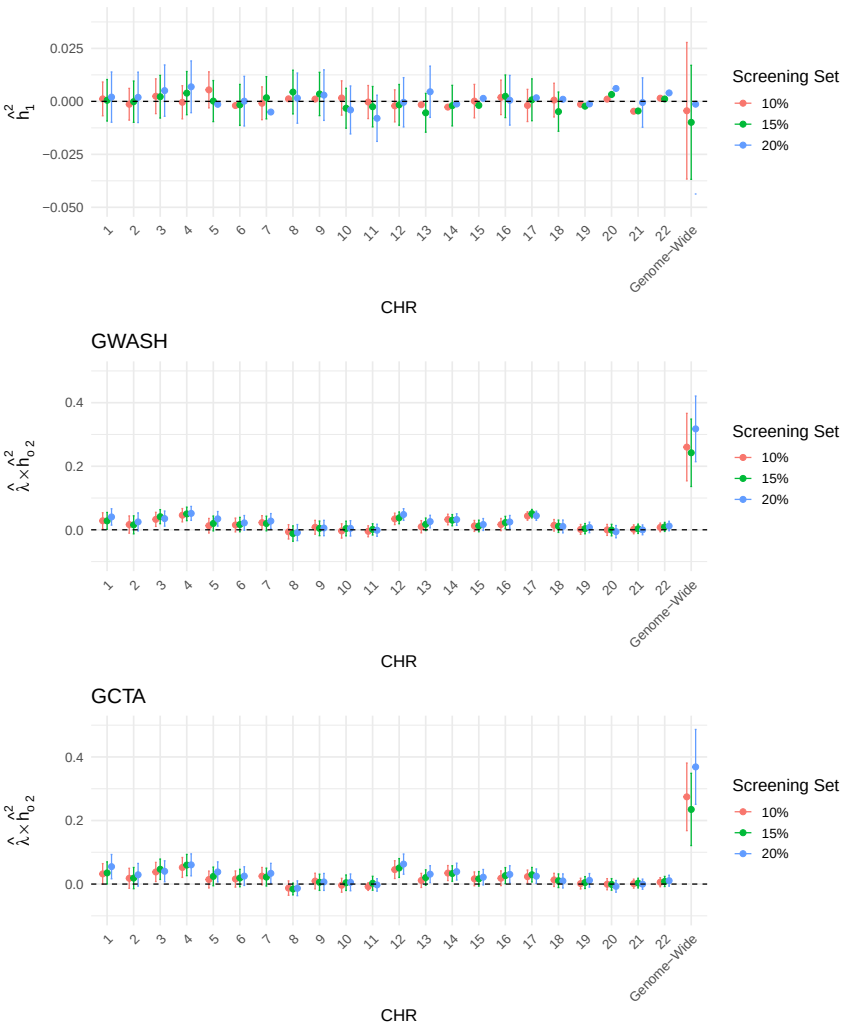

Figure 13: Sensitivity analysis of chromosome-specific decomposed FVE estimates based on strong effects identified by the HOLP screening method using a screening set comprising $\{10\%, 15\%, 20\%\}$ of the original sample. Note that the y-axis scales differ across the three panels.

results are stable across different choices of screening set sizes, indicating the robustness of the procedure. Accordingly, the BH method is recommended, as the real data analysis indicates that it demonstrates robustness with respect to the screening set size.

As for HOLP, we conducted a similar sensitivity analysis. Because the number of SNPs selected by HOLP depends directly on the screening sample size, the corresponding values of d_s were $\{101, 143, 184\}$ when the screening set was chosen as $\{10\%, 15\%, 20\%\}$ of the original sample, respectively. The results of this analysis are summarized in Figure 13. We observe that the decomposed FVE estimates, whether obtained using GWASH or GCTA, are sensitive to the choice of screening set. This instability likely arises because the number of identified SNPs depends on the size of the screening set.

8 Discussion

We have proposed a novel and general framework for decomposed FVE estimation in high-dimensional linear models where coefficients may have different orders of magnitude. To address limitations of existing methods, such as LDSC or GWASH, which rely on the BKE condition, we partition effects into strong and weak components: strong effects, treated as a low-dimensional problem, are estimated via adjusted R^2, while weak effects, after removing and adjusting for the strong ones, are assumed to satisfy the BKE condition and can be estimated using existing high-dimensional methods such as GWASH, LDSC, or GCTA. To help illustrate the proposed method and study its performance, we model the coefficients $\boldsymbol{\beta}$ using a mixture Gaussian model to control kurtosis and study BKE violations. We further derive a formal relationship between the kurtosis of $\boldsymbol{\beta}$ and that of the estimable quantity $\boldsymbol{\eta}$, which enables a hypothesis test for evaluating the validity of the BKE assumption. This relationship can also be empirically evaluated under the mixture Gaussian model. Finally, we incorporated several screening procedures into the proposed framework to identify strong effects, thereby facilitating the FVE decomposition. This proposed framework is also successfully applied to the ABCD dataset for chromosome-specific FVE estimation of the polyvoxel score.

Our proposed FVE decomposition framework exhibits superior accuracy in terms of both relative RMSE and bias. In particular, when the correlation among predictors is high and the kurtosis of $\boldsymbol{\beta}$ is large, the decomposed estimator substantially reduces both bias and RMSE. The proposed test for the BKE condition demonstrates well-controlled type I error rates and high statistical power under moderate predictor correlation across different settings. However, when the correlation becomes relatively high, the test fails to maintain proper type I error control. In addition, the incorporated screening procedures effectively identify strong effects, enabling the decomposition to achieve accurate estimation of the FVE.

There are several limitations and potential directions for future research. First, our estimation framework requires access to the original data and cannot be applied to datasets where only summary statistics are available. Extending the decomposed FVE estimation framework to operate with summary statistics represents a practically relevant avenue for further development. This extension, however, is technically challenging, as summary statistics do not provide sufficient information to perform regression and to residualize the data with respect to the set of SNPs with strong effects.

Second, in our simulation studies, we observed that the proposed test exhibited an inflated type I error rate when the AR(1) correlation parameter ρ exceeded 0.5, indicating a lack of robustness under highly correlated data structures. A plausible explanation is that the Jarque–Bera test relies on the i.i.d. assumption, whereas larger values of ρ induce dependence among the quantities η_j, rendering the computation of $\tilde{\eta}_j$ inappropriate in this context. A potential remedy is to invert the full covariance matrix $\boldsymbol{V}$, rather than only its diagonal elements, which would otherwise ignore the dependence among the η_j's when ρ is large. However, this is nontrivial, as $\boldsymbol{V}$ is not invertible when the feature dimension exceeds the sample size, and a more robust strategy is needed to obtain properly standardized η_j values. Consequently, developing a more stable and general procedure for testing the BKE condition under highly correlated predictor structures represents an important

direction for future research.

Third, in both the simulation studies and the real data analysis, we observed that the procedure may be sensitive to the screening sample size, depending on the screening method employed. Developing a version of the estimator that is more robust to the screening sample size is, therefore, an important direction for future work.

Acknowledgments

The authors thank Chun Chieh Fan and Anubhav Singh-Sachan from UCSD for their help in processing the ABCD data.

Funding

The work in this article was partially funded by National Institutes of Health (NIH) grant R01MH128923.

Conflict of Interest

The authors declare no conflict of interest.

References

David Azriel, Samuel Davenport, and Armin Schwartzman. Consistency of heritability estimation from summary statistics in high-dimensional linear models. *arXiv preprint arXiv:2502.11144*, 2025.

Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300, 1995.

Carlo Bonferroni. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R istituto superiore di scienze economiche e commericiali di firenze*, 8:3–62, 1936.

Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Nick Patterson, Mark J Daly, Alkes L Price, and Benjamin M Neale. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics*, 47(3):291–295, 2015.

Betty Jo Casey, Tariq Cannonier, May I Conley, Alexandra O Cohen, Deanna M Barch, Mary M Heitzeg, Mary E Soules, Theresa Teslovich, Danielle V Dellarco, Hugh Garavan, et al. The adolescent brain cognitive development (ABCD) study: imaging acquisition across 21 sites. *Developmental cognitive neuroscience*, 32:43–54, 2018.

Jan Solomon Cramer. Mean and variance of R^2 in small and moderate samples. *Journal of econometrics*, 35(2-3):253–266, 1987.

Lee H Dicker. Variance estimation in high-dimensional linear models. *Biometrika*, 101(2):269–284, 2014.

Luke M Evans, Rasool Tahmasbi, Scott I Vrieze, Gonçalo R Abecasis, Sayantan Das, Steven Gazal, Douglas W Bjelland, Teresa R De Candia, Haplotype Reference Consortium, Michael E Goddard, et al. Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. *Nature genetics*, 50(5):737–745, 2018.

Chun Chieh Fan, Robert Loughnan, Sylia Wilson, John K Hewitt, and ABCD Genetic Working Group. Genotype data and derived genetic instruments of adolescent brain cognitive development study® for better understanding of human brain development. *Behavior genetics*, 53(3):159–168, 2023.

Jianqing Fan and Jinchi Lv. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 70(5):849–911, 2008.

Hilary K Finucane, Brendan Bulik-Sullivan, Alexander Gusev, Gosia Trynka, Yakir Reshef, Po-Ru Loh, Verneri Anttila, Han Xu, Chongzhi Zang, Kyle Farh, et al. Partitioning heritability by

functional annotation using genome-wide association summary statistics. *Nature genetics*, 47(11): 1228–1235, 2015.

Ronald A Fisher. XV.—The correlation between relatives on the supposition of mendelian inheritance. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, 52(2):399–433, 1919.

Alexander Gusev, S Hong Lee, Gosia Trynka, Hilary Finucane, Bjarni J Vilhjálmsson, Han Xu, Chongzhi Zang, Stephan Ripke, Brendan Bulik-Sullivan, Eli Stahl, et al. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *The American Journal of Human Genetics*, 95(5):535–552, 2014.

Dominic Holland, Oleksandr Frei, Rahul Desikan, Chun-Chieh Fan, Alexey A Shadrin, Olav B Smeland, Vijay S Sundar, Paul Thompson, Ole A Andreassen, and Anders M Dale. Beyond snp heritability: Polygenicity and discoverability of phenotypes estimated with a univariate gaussian mixture model. *PLoS Genetics*, 16(5):e1008612, 2020.

Kangcheng Hou, Kathryn S Burch, Arunabha Majumdar, Huwenbo Shi, Nicholas Mancuso, Yue Wu, Sriram Sankararaman, and Bogdan Pasaniuc. Accurate estimation of SNP-heritability from biobank-scale data irrespective of genetic architecture. *Nature genetics*, 51(8):1244–1251, 2019.

Lucas Janson, Rina Foygel Barber, and Emmanuel Candes. Eigenprism: inference for high dimensional signal-to-noise ratios. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(4):1037–1065, 2017.

Vincent Laville, Jae H Kang, Clara C Cousins, Adriana I Iglesias, Réka Nagy, Jessica N Cooke Bailey, Robert P Igo Jr, Yeunjoo E Song, Daniel I Chasman, William G Christen, et al. Genetic correlations between diabetes and glaucoma: an analysis of continuous and dichotomous phenotypes. *American journal of ophthalmology*, 206:245–255, 2019.

Sang Hong Lee, Naomi R Wray, Michael E Goddard, and Peter M Visscher. Estimating missing heritability for disease from genome-wide association studies. *The American Journal of Human Genetics*, 88(3):294–305, 2011.

Gaorong Li, Heng Peng, Jun Zhang, and Lixing Zhu. Robust rank correlation based screening. *The Annals of Statistics*, 40(3):1846–1877, 2012.

Robert Loughnan, Jonathan Ahern, Mary Boyle, Terry L Jernigan, Donald J Hagler Jr, John R Iversen, Oleksandr Frei, Diana M Smith, Ole Andreassen, Noah Zaitlen, et al. Hemochromatosis neural archetype reveals iron disruption in motor circuits. *Science Advances*, 10(47):eadp4431, 2024.

Robert J Loughnan, Alexey A Shadrin, Oleksandr Frei, Dennis van der Meer, Weiqi Zhao, Clare E Palmer, Wesley K Thompson, Carolina Makowski, Terry L Jernigan, Ole A Andreassen, et al. Gen-

eralization of cortical mostest genome-wide associations within and across samples. *Neuroimage*, 263:119632, 2022.

Monica Luciana, James M Bjork, Bonnie J Nagel, Deanna M Barch, Raul Gonzalez, Sara Jo Nixon, and Marie T Banich. Adolescent neurocognitive development and impacts of substance use: Overview of the adolescent brain cognitive development (ABCD) baseline neurocognition battery. *Developmental cognitive neuroscience*, 32:67–79, 2018.

Yang Luo, Xinyi Li, Xin Wang, Steven Gazal, Josep Maria Mercader, Me Research Team, SIGMA Type 2 Diabetes Consortium, Benjamin M Neale, Jose C Florez, Adam Auton, et al. Estimating heritability and its enrichment in tissue-specific gene sets in admixed populations. *Human molecular genetics*, 30(16):1521–1534, 2021.

Michael Lynch, Bruce Walsh, et al. *Genetics and analysis of quantitative traits*, volume 1. Sinauer Sunderland, MA, 1998.

Tinca JC Polderman, Beben Benyamin, Christiaan A De Leeuw, Patrick F Sullivan, Arjen Van Bochoven, Peter M Visscher, and Danielle Posthuma. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature genetics*, 47(7):702–709, 2015.

Armin Schwartzman, Andrew J Schork, Rong Zablocki, and Wesley K Thompson. A simple, consistent estimator of snp heritability from genome-wide association studies. *The annals of applied statistics*, 13(4):2509, 2019.

Huwenbo Shi, Gleb Kichaev, and Bogdan Pasaniuc. Contrasting the genetic architecture of 30 complex traits from summary association data. *The American Journal of Human Genetics*, 99(1): 139–153, 2016.

Anubhav Nikunj Singh Sachan. *Estimating genetic effects under different data complexities*. PhD thesis, UC San Diego, 2025.

Doug Speed and David J Balding. Sumher better estimates the snp heritability of complex traits from summary statistics. *Nature genetics*, 51(2):277–284, 2019.

Doug Speed, Gibran Hemani, Michael R Johnson, and David J Balding. Improved heritability estimation from genome-wide snps. *The American Journal of Human Genetics*, 91(6):1011–1021, 2012.

Patrick F Sullivan and Daniel H Geschwind. Defining the genetic, genomic, cellular, and diagnostic architectures of psychiatric disorders. *Cell*, 177(1):162–183, 2019.

Daniel Taliun, Daniel N Harris, Michael D Kessler, Jedidiah Carlson, Zachary A Szpiech, Raul Torres, Sarah A Gagliano Taliun, André Corvelo, Stephanie M Gogarten, Hyun Min Kang, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*, 590(7845): 290–299, 2021.

Peter M Visscher, Naomi R Wray, Qian Zhang, Pamela Sklar, Mark I McCarthy, Matthew A Brown, and Jian Yang. 10 years of gwas discovery: biology, function, and translation. *The American Journal of Human Genetics*, 101(1):5–22, 2017.

Xiangyu Wang and Chenlei Leng. High dimensional ordinary least squares projection for screening variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 78(3):589–611, 2016.

Jichun Xie, T Tony Cai, John Maris, and Hongzhe Li. False discovery rate control for high-dimensional dependent data with an application to large-scale genetic association studies. *The Annals of Applied Statistics*, 5(2A):1207–1225, 2011.

Jian Yang, Beben Benyamin, Brian P McEvoy, Scott Gordon, Anjali K Henders, Dale R Nyholt, Pamela A Madden, Andrew C Heath, Nicholas G Martin, Grant W Montgomery, et al. Common SNPs explain a large proportion of the heritability for human height. *Nature genetics*, 42(7): 565–569, 2010.

Jian Yang, S Hong Lee, Michael E Goddard, and Peter M Visscher. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88(1):76–82, 2011.

A Further Details on Comparison Methods

A.1 Genome-wide Complex Trait Analysis (GCTA)

Genome-wide Complex Trait Analysis (GCTA; Yang et al. [2011]) is a widely used framework for estimating the proportion of phenotypic variance explained by genome-wide SNPs, commonly referred to as SNP heritability. Conceptually, GCTA is based on the same underlying linear model as Model (1), which assumes that the observed phenotype can be represented as the sum of additive genetic effects from a large number of SNPs and residual environmental noise.

In the framework of GCTA, all SNP effects are treated as random variables drawn from a normal distribution with a common variance component. This assumption implies that each SNP contributes an equal, infinitesimal amount to the overall genetic variance. The key task is therefore to estimate the variance components corresponding to the genetic and residual sources of variability. GCTA accomplishes this by fitting a linear mixed-effects model using restricted maximum likelihood (REML). Based on these estimates, SNP heritability is then computed as the ratio of the genetic variance to the total phenotypic variance, as defined in Equation (3).

To capture the dependence structure induced by shared genotypes, GCTA constructs a genetic relationship matrix (GRM) from standardized genotype data. The GRM quantifies pairwise genetic similarity between individuals across all SNPs and serves as the covariance matrix for the random genetic effects in the mixed model. Intuitively, the GRM allows GCTA to borrow strength across the entire genome to estimate the aggregate contribution of SNPs, even when individual effects are too small to detect through single-marker association tests.

A.2 Screening Methods

GWAS are commonly used statistical analyses aimed at identifying genetic variants, typically SNPs, associated with particular traits or diseases. GWAS involves testing each genetic variant individually for association with the phenotype of interest across the genome, often resulting in a massive multiple-testing scenario due to the large number of SNPs examined. Screening methods share conceptual similarities with GWAS in the aim to efficiently identifying significant predictors from large sets of candidates. In our context, screening methods specifically help in isolating predictors with strong effects, which may not be known in advance, facilitating the subsequent application of the decomposed estimation procedure.

Numerous screening methods have been proposed in the literature. Here, we select five representative methods and evaluate their effectiveness. A comparison between these methods can be found in Table 2.

- Bonferroni correction [Bonferroni, 1936] or the Benjamini–Hochberg procedure [Benjamini and Hochberg, 1995] are widely used approaches in multiple hypothesis testing and can be applied here to identify potential strong effects. The result in Equation (24) states $\tilde{\boldsymbol{\eta}}$ approximately follows a standard normal distribution, $\tilde{\boldsymbol{\eta}} \sim N(0, \boldsymbol{I}_m)$. The significance of each parameter $\tilde{\eta}_j$

is assessed using a two-sided p-value computed as

$$p_j = 2\left(1 - \Phi\left(|\tilde{\eta}_j|\right)\right),$$

where $\Phi(\cdot)$ denotes the cumulative distribution function (CDF) of the standard normal distribution. Since computing p_j for all $j = 1, \ldots, m$ gives rise to a multiple testing problem, we adjust the p-values $\{p_1, \ldots, p_m\}$ using either the Bonferroni correction or the Benjamini–Hochberg procedure. An effect is deemed strong if its adjusted p-value falls below a pre-specified significance level α.

- Sure Independence Screening (SIS, Fan and Lv [2008]) ranks predictors by their marginal Pearson correlations $\omega_j = \boldsymbol{X}_j^\top \boldsymbol{y}$, where $\boldsymbol{X}_j$ is j-th column of $\boldsymbol{X}$, retaining the top d (e.g. $d = \lfloor n/\log n \rfloor$ or $n-1$), which provides efficient dimensionality reduction under the assumption that important features exhibit strong marginal effects.
- High-dimensional Ordinary Least-squares Projection (HOLP, Wang and Leng [2016]) computes $\hat{\boldsymbol{\beta}}_{\text{HOLP}} = \boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{y}$. They show that $\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{X}^\top)^{-1}$ is the Moore-Penrose inverse of $\boldsymbol{X}$ when $m > n$; hence this estimator is an extension of the Ordinary Least-squares to the high-dimensional setting where $m > n$. HOLP then ranks variables by $|\hat{\beta}_{\text{HOLP},j}|$, thereby relaxing the "strong marginal correlation" requirement of SIS and improving screening performance when predictors are highly correlated.
- Robust Rank Correlation Screening (RRCS, Li et al. [2012]) replaces Pearson correlation with a Kendall-τ–type statistic, as defined in Equation (2.4) of Li et al. [2012], selecting predictors whose statistics exceed a threshold γ_n to enhance robustness against heavy-tailed distributions and nonlinear relationships.

Table 2: Summary of Selected Screening Methods

Method	**Screening Quantity**	**Screening Rule**
Bonferroni Correction/ Benjamini–Hochberg procedure	Two-sided p-values based on $\tilde{\eta}_j$ (24).	Select effects with adjusted p-values below a significance threshold α
Sure Independence Screening	Marginal Pearson correlation $\omega_j = \boldsymbol{X}_j^\top \boldsymbol{y}$ for each predictor.	Rank predictors by $\lvert\omega_j\rvert$ and select the top d predictors (e.g., $d = n-1$ or $\lfloor n/\log n \rfloor$).
High-dimensional Ordinary Least-squares Projection	Projected OLS coefficient $\hat{\beta}_{\text{HOLP},j} = [\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{y}]_j$	Rank predictors by $\lvert\hat{\beta}_{\text{HOLP},j}\rvert$ and select the top d predictors (e.g., $d = n-1$ or $\lfloor n/\log n \rfloor$).
Robust Rank Correlation Screening	Kendall-τ-type statistic for each predictor.	Select predictors whose statistic exceeds a given threshold γ_n.

To enable a more comprehensive comparison, we include two additional benchmark methods when evaluating the approaches discussed above. The first is the baseline GWASH or GCTA method without any decomposition and applied directly without screening. The second is the oracle method,

which uses the true indices of strong effects to perform decomposition. The performance comparison among all methods is presented in Section 6.5.

B Proofs

B.1 Proof of Proposition 3.1

Based on Equation (14), we have

$$\kappa_\beta - 3 = 3\left[\frac{\theta_0^2}{p} + \frac{(1-\theta_0)^2}{1-p} - 1\right].$$

(1) Consider scenarios in which the BKE condition is violated:

- Assume $mp \to c$ and then $m(1-p) \to \infty$ as $m \to \infty$,

$$\frac{1}{m}[\kappa_\beta - 3] = \frac{3\theta_0^2}{mp} + \frac{3(1-\theta_0)^2}{m(1-p)} - \frac{3}{m} \to \frac{3\theta_0^2}{c}.$$

- Assume $m(1-p) \to c$ and then $mp \to \infty$ as $m \to \infty$,

$$\frac{1}{m}[\kappa_\beta - 3] = \frac{3\theta_0^2}{mp} + \frac{3(1-\theta_0)^2}{m(1-p)} - \frac{3}{m} \to \frac{3(1-\theta_0)^2}{c}.$$

(2) Consider scenarios in which the BKE condition is satisfied:

- $p = \theta_0$:

$$\frac{1}{m}[\kappa_\beta - 3] = \frac{3\theta_0^2}{mp} + \frac{3(1-\theta_0)^2}{m(1-p)} - \frac{3}{m} = \frac{3\theta_0}{m} + \frac{3(1-\theta_0)}{m} - \frac{3}{m} \to 0.$$

- $p = 0$ or $p = 1$: In this case, the mixture Gaussian model (13) collapses to a single normal distribution, yielding $\kappa_\beta = 3$, which trivially satisfies the BKE condition.

□

B.2 Proof of Proposition 3.2

Let $\kappa_\beta = \text{Kurt}[\beta_j]$ and $\kappa_{\eta_j} = \text{Kurt}[\eta_j]$. From the definition of η_j in (5),

$$\eta_j = \frac{\sum_{k=1}^m \mathbf{\Sigma}_{jk}\beta_k}{\sigma_y\sqrt{\mathbf{\Sigma}_{jj}}} \qquad \Rightarrow \qquad \kappa_{\eta_j} = \text{Kurt}\left[\sum_{k=1}^m \mathbf{\Sigma}_{jk}\beta_k\right].$$

Assuming that the random variables β_k are independent with $\text{E}[\beta_k] = 0$ and $\text{Var}(\beta_k) = \sigma_\beta^2$, then $\text{E}[\beta_k^4] = \kappa_\beta[\text{Var}(\beta_k)]^2 = \kappa_\beta\sigma_\beta^4$. Define

$$w_k = \mathbf{\Sigma}_{jk} \quad \text{and} \quad W = \sum_{k=1}^m w_k\beta_k.$$

Then

$$\mathrm{Var}(W) = \sigma_\beta^2 \sum_{k=1}^{m} w_k^2.$$

When expanding W^4, we obtain:

$$W^4 = \sum_{k=1}^{m} w_k^4 \beta_k^4 + 6 \sum_{k<\ell} w_k^2 w_\ell^2 \beta_k^2 \beta_\ell^2 + (\text{other terms with odd moments}).$$

In this expansion of $\left(\sum_{k=1}^{m} w_k \beta_k\right)^4$, focus on the terms where exactly two distinct indices, say k and ℓ with $k \neq \ell$, appear twice. These terms have the form $w_k^2 w_\ell^2 \beta_k^2 \beta_\ell^2$. To see the combinatorial factor, note that there are 4 positions in the product. We need to choose 2 of these positions to be occupied by the index k (with the remaining two occupied by ℓ). The number of ways to choose 2 positions out of 4 is given by the binomial coefficient $\binom{4}{2} = 6$. This is the origin of the factor 6 in the expansion.

Taking expectations gives the fourth moment of W,

$$\begin{aligned}
\mathrm{E}[W^4] &= \sum_{k=1}^{m} w_k^4 \, \mathrm{E}[\beta_k^4] + 6 \sum_{k<\ell} w_k^2 w_\ell^2 \; \mathrm{Var}(\beta_k) \, \mathrm{Var}(\beta_\ell) + \mathrm{E}[\text{other terms with odd moments}] \\
&= \sigma_\beta^4 \left(\kappa_\beta \sum_{k=1}^{m} w_k^4 + 6 \sum_{k<\ell} w_k^2 w_\ell^2 \right).
\end{aligned}$$

where the terms involving odd powers vanish since $\mathrm{E}[\beta_k] = 0$.

The kurtosis of W is

$$\mathrm{Kurt}(W) = \frac{\mathrm{E}[W^4]}{\mathrm{Var}(W)^2} = \frac{\sigma_\beta^4 \left(\kappa_\beta \sum_{k=1}^{m} w_k^4 + 6 \sum_{k<\ell} w_k^2 w_\ell^2 \right)}{\sigma_\beta^4 \left(\sum_{k=1}^{m} w_k^2 \right)^2} = \frac{\kappa_\beta \sum_{k=1}^{m} w_k^4 + 6 \sum_{k<\ell} w_k^2 w_\ell^2}{\left(\sum_{k=1}^{m} w_k^2 \right)^2}.$$

Notice that

$$\left(\sum_{k=1}^{m} w_k^2 \right)^2 = \sum_{k=1}^{m} w_k^4 + 2 \sum_{k<\ell} w_k^2 w_\ell^2.$$

Therefore, the kurtosis can be rewritten as

$$\begin{aligned}
\mathrm{Kurt}(W) &= \frac{\kappa_\beta \sum_{k=1}^{m} w_k^4 + 6 \sum_{k<\ell} w_k^2 w_\ell^2}{\sum_{k=1}^{m} w_k^4 + 2 \sum_{k<\ell} w_k^2 w_\ell^2} \\
&= \frac{3 \sum_{k=1}^{m} w_k^4 + 6 \sum_{k<\ell} w_k^2 w_\ell^2 + (\kappa_\beta - 3) \sum_{k=1}^{m} w_k^4}{\sum_{k=1}^{m} w_k^4 + 2 \sum_{k<\ell} w_k^2 w_\ell^2} \\
&= 3 + (\kappa_\beta - 3) \frac{\sum_{k=1}^{m} w_k^4}{\left(\sum_{k=1}^{m} w_k^2 \right)^2}.
\end{aligned}$$

Returning to our original notation with $w_k = \mathbf{\Sigma}_{jk}$, we have

$$\kappa_{\eta_j} = \text{Kurt}\left[\sum_{k=1}^{m} \mathbf{\Sigma}_{jk}\beta_k\right] = 3 + (\kappa_\beta - 3)\frac{\sum_{k=1}^{m} \mathbf{\Sigma}_{jk}^4}{\left(\sum_{k=1}^{m} \mathbf{\Sigma}_{jk}^2\right)^2}.$$

□

B.3 Proof of Proposition 4.1

Step 1: Recall $\vec{\boldsymbol{x}}_{\bullet 2} = \vec{\boldsymbol{x}}_2 - \vec{\boldsymbol{x}}_1 \mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}$, $y_{\bullet 2} = y - \vec{\boldsymbol{x}}_1 \mathbf{\Sigma}_{11}^{-1} E(\vec{\boldsymbol{x}}_1^\top y)$, and

$$\boldsymbol{\beta} \;=\; \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \;=\; \mathbf{\Sigma}^{-1}\begin{pmatrix} \mathrm{E}(\vec{\boldsymbol{x}}_1^\top\, y) \\ \mathrm{E}(\vec{\boldsymbol{x}}_2^\top\, y) \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}.$$

Let

$$\mathbf{\Sigma}_{\bullet 22} = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\,\mathbf{\Sigma}_{11}^{-1}\,\mathbf{\Sigma}_{12},$$

then by the inverse-block formula,

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} \mathbf{\Sigma}_{11}^{-1} + \mathbf{\Sigma}_{11}^{-1}\,\mathbf{\Sigma}_{12}\,\mathbf{\Sigma}_{\bullet 22}^{-1}\,\mathbf{\Sigma}_{21}\,\mathbf{\Sigma}_{11}^{-1} & -\mathbf{\Sigma}_{11}^{-1}\,\mathbf{\Sigma}_{12}\,\mathbf{\Sigma}_{\bullet 22}^{-1} \\ -\mathbf{\Sigma}_{\bullet 22}^{-1}\,\mathbf{\Sigma}_{21}\,\mathbf{\Sigma}_{11}^{-1} & \mathbf{\Sigma}_{\bullet 22}^{-1} \end{pmatrix}.$$

The bottom block $\boldsymbol{\beta}_2$ is given by

$$\boldsymbol{\beta}_2 \;=\; \mathbf{\Sigma}_{\bullet 22}^{-1}\Big[\mathrm{E}(\vec{\boldsymbol{x}}_2^\top\, y) \;-\; \mathbf{\Sigma}_{21}\,\mathbf{\Sigma}_{11}^{-1}\,\mathrm{E}(\vec{\boldsymbol{x}}_1^\top\, y)\Big] = \mathbf{\Sigma}_{\bullet 22}^{-1}\mathrm{E}(\vec{\boldsymbol{x}}_{\bullet 2}^\top y_{\bullet 2}).$$

Step 2: In order to show Equation (18), it is equivalent to show

$$\boldsymbol{\beta}^\top\,\mathbf{\Sigma}\,\boldsymbol{\beta} \;=\; \boldsymbol{\beta}_1^{*T}\,\mathbf{\Sigma}_{11}\,\boldsymbol{\beta}_1^* \;+\; \boldsymbol{\beta}_2^\top\,\mathbf{\Sigma}_{\bullet 22}\,\boldsymbol{\beta}_2.$$

<u>Step 2.1</u>: By definition

$$\boldsymbol{\beta} \;=\; \mathbf{\Sigma}^{-1}\,E\big(\vec{\boldsymbol{x}}^\top\, y\big), \qquad \mathbf{\Sigma} \;=\; \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}.$$

Then we have $\mathbf{\Sigma}\boldsymbol{\beta} = \mathrm{E}(\vec{\boldsymbol{x}}^\top y)$, which becomes two vector equations in the block forms:

$$\begin{aligned} \mathbf{\Sigma}_{11}\,\boldsymbol{\beta}_1 + \mathbf{\Sigma}_{12}\,\boldsymbol{\beta}_2 &= \mathrm{E}(\vec{\boldsymbol{x}}_1^\top\, y), \\ \mathbf{\Sigma}_{21}\,\boldsymbol{\beta}_1 + \mathbf{\Sigma}_{22}\,\boldsymbol{\beta}_2 &= \mathrm{E}(\vec{\boldsymbol{x}}_2^\top\, y). \end{aligned}$$

Recall that $\boldsymbol{\beta}_1^* = \mathbf{\Sigma}_{11}^{-1}\,\mathrm{E}(\vec{\boldsymbol{x}}_1^\top\, y)$. From the first equation above, we can solve for $\boldsymbol{\beta}_1$:

$$\boldsymbol{\beta}_1 \;=\; \mathbf{\Sigma}_{11}^{-1}\,\mathrm{E}(\vec{\boldsymbol{x}}_1^\top\, y) \;-\; \mathbf{\Sigma}_{11}^{-1}\,\mathbf{\Sigma}_{12}\,\boldsymbol{\beta}_2 \;=\; \boldsymbol{\beta}_1^* \;-\; \mathbf{\Sigma}_{11}^{-1}\,\mathbf{\Sigma}_{12}\,\boldsymbol{\beta}_2.$$

Hence,

$$\vec{\boldsymbol{x}}_1\,\boldsymbol{\beta}_1 \;=\; \vec{\boldsymbol{x}}_1\,\boldsymbol{\beta}_1^* \;-\; \vec{\boldsymbol{x}}_1\,\big(\boldsymbol{\Sigma}_{11}^{-1}\,\boldsymbol{\Sigma}_{12}\,\boldsymbol{\beta}_2\big).$$

Recall that $\vec{\boldsymbol{x}}_{\bullet 2} = \vec{\boldsymbol{x}}_2 \;-\; \vec{\boldsymbol{x}}_1\,\boldsymbol{\Sigma}_{11}^{-1}\,\boldsymbol{\Sigma}_{12}$. Hence

$$\vec{\boldsymbol{x}}_2\,\boldsymbol{\beta}_2 \;=\; \big(\vec{\boldsymbol{x}}_1\,\boldsymbol{\Sigma}_{11}^{-1}\,\boldsymbol{\Sigma}_{12} + \vec{\boldsymbol{x}}_{\bullet 2}\big)\,\boldsymbol{\beta}_2 \;=\; \vec{\boldsymbol{x}}_1\,\big(\boldsymbol{\Sigma}_{11}^{-1}\,\boldsymbol{\Sigma}_{12}\,\boldsymbol{\beta}_2\big) \;+\; \vec{\boldsymbol{x}}_{\bullet 2}\,\boldsymbol{\beta}_2.$$

Substituting $\vec{\boldsymbol{x}}_1\,\boldsymbol{\beta}_1$ and $\vec{\boldsymbol{x}}_2\,\boldsymbol{\beta}_2$ into

$$\vec{\boldsymbol{x}}\,\boldsymbol{\beta} = \vec{\boldsymbol{x}}_1\,\boldsymbol{\beta}_1 + \vec{\boldsymbol{x}}_2\,\boldsymbol{\beta}_2,$$

we obtain:

$$\begin{aligned}\vec{\boldsymbol{x}}\,\boldsymbol{\beta} =\; & \vec{\boldsymbol{x}}_1\,\boldsymbol{\beta}_1^* \;-\; \vec{\boldsymbol{x}}_1\,\big(\boldsymbol{\Sigma}_{11}^{-1}\,\boldsymbol{\Sigma}_{12}\,\boldsymbol{\beta}_2\big) \;+\; \vec{\boldsymbol{x}}_1\,\big(\boldsymbol{\Sigma}_{11}^{-1}\,\boldsymbol{\Sigma}_{12}\,\boldsymbol{\beta}_2\big) \;+\; \vec{\boldsymbol{x}}_{\bullet 2}\,\boldsymbol{\beta}_2 \\ =\; & \vec{\boldsymbol{x}}_1\,\boldsymbol{\beta}_1^* \;+\; \vec{\boldsymbol{x}}_{\bullet 2}\,\boldsymbol{\beta}_2.\end{aligned}$$

<u>Step 2.2</u>: By definition of $\vec{\boldsymbol{x}}_{\bullet 2}$, we have

$$\vec{\boldsymbol{x}}_{\bullet 2} = \vec{\boldsymbol{x}}_2 - \vec{\boldsymbol{x}}_1\,\boldsymbol{\Sigma}_{11}^{-1}\,\boldsymbol{\Sigma}_{12} = \mathrm{E}\big(\vec{\boldsymbol{x}}_1^\top\,\vec{\boldsymbol{x}}_2\big) - \mathrm{E}\Big(\vec{\boldsymbol{x}}_1^\top\,\vec{\boldsymbol{x}}_1\,\boldsymbol{\Sigma}_{11}^{-1}\,\boldsymbol{\Sigma}_{12}\Big).$$

Then we have

$$\mathrm{E}\big(\vec{\boldsymbol{x}}_1^\top\,\vec{\boldsymbol{x}}_{\bullet 2}\big) \;=\; \mathrm{E}\big(\vec{\boldsymbol{x}}_1^\top\,\vec{\boldsymbol{x}}_2\big) \;-\; \mathrm{E}\Big(\vec{\boldsymbol{x}}_1^\top\,\vec{\boldsymbol{x}}_1\,\boldsymbol{\Sigma}_{11}^{-1}\,\boldsymbol{\Sigma}_{12}\Big) = \boldsymbol{\Sigma}_{12} \;-\; \boldsymbol{\Sigma}_{11}\,\boldsymbol{\Sigma}_{11}^{-1}\,\boldsymbol{\Sigma}_{12} \;=\; \mathbf{0}.$$

Since $\mathrm{E}(\vec{\boldsymbol{x}}_1^\top\,\vec{\boldsymbol{x}}_{\bullet 2}) = 0$,

$$\mathrm{E}\Big(\big[\vec{\boldsymbol{x}}_1\,\boldsymbol{\beta}_1^*\big]^\top\big[\vec{\boldsymbol{x}}_{\bullet 2}\,\boldsymbol{\beta}_2\big]\Big) \;=\; \mathrm{E}\Big(\boldsymbol{\beta}_1^{*T}\,\vec{\boldsymbol{x}}_1^\top\,\vec{\boldsymbol{x}}_{\bullet 2}\,\boldsymbol{\beta}_2\Big) \;=\; \boldsymbol{\beta}_1^{*T}\,\Big[\mathrm{E}\big(\vec{\boldsymbol{x}}_1^\top\,\vec{\boldsymbol{x}}_{\bullet 2}\big)\Big]\,\boldsymbol{\beta}_2 = 0.$$

Since $\mathrm{E}(\vec{\boldsymbol{x}}_1) = 0$ and

$$\mathrm{E}(\vec{\boldsymbol{x}}_{\bullet 2,j}) = \mathrm{E}(\vec{\boldsymbol{x}}_{2,j} - \vec{\boldsymbol{x}}_1\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12,\cdot j}) = \mathrm{E}(\vec{\boldsymbol{x}}_{2,j}) - \mathrm{E}(\vec{\boldsymbol{x}}_1)\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12,\cdot j} = 0,$$

we have

$$\mathrm{E}(\vec{\boldsymbol{x}}_1\,\boldsymbol{\beta}_1^*) = 0, \quad \mathrm{E}(\vec{\boldsymbol{x}}_{\bullet 2}\,\boldsymbol{\beta}_2) = 0.$$

Thus,

$$\mathrm{Cov}\big(\vec{\boldsymbol{x}}_1\,\boldsymbol{\beta}_1^*,\; \vec{\boldsymbol{x}}_{\bullet 2}\,\boldsymbol{\beta}_2\big) \;=\; \mathrm{E}\Big(\big[\vec{\boldsymbol{x}}_1\,\boldsymbol{\beta}_1^*\big]^\top\big[\vec{\boldsymbol{x}}_{\bullet 2}\,\boldsymbol{\beta}_2\big]\Big) = 0.$$

<u>Step 2.3</u>: In

$$\vec{\boldsymbol{x}}\,\boldsymbol{\beta} = \vec{\boldsymbol{x}}_1\,\boldsymbol{\beta}_1^* + \vec{\boldsymbol{x}}_{\bullet 2}\,\boldsymbol{\beta}_2,$$

the two terms are uncorrelated. Thus

$$\begin{aligned}\mathrm{E}\Big[\big(\vec{\boldsymbol{x}}\,\boldsymbol{\beta}\big)^2\Big] &= \mathrm{E}\Big[\big(\vec{\boldsymbol{x}}_1\,\boldsymbol{\beta}_1^* + \vec{\boldsymbol{x}}_{\bullet 2}\,\boldsymbol{\beta}_2\big)^2\Big] = \mathrm{E}\Big[\big(\vec{\boldsymbol{x}}_1\,\boldsymbol{\beta}_1^*\big)^2\Big] + \mathrm{E}\Big[\big(\vec{\boldsymbol{x}}_{\bullet 2}\,\boldsymbol{\beta}_2\big)^2\Big] \\ &= \boldsymbol{\beta}_1^{*T}\mathrm{E}\big(\vec{\boldsymbol{x}}_1^\top\vec{\boldsymbol{x}}_1\big)\boldsymbol{\beta}_1^* + \boldsymbol{\beta}_2^\top\mathrm{E}\big(\vec{\boldsymbol{x}}_{\bullet 2}^\top\vec{\boldsymbol{x}}_{\bullet 2}\big)\boldsymbol{\beta}_2 \\ &= \boldsymbol{\beta}_1^{*T}\boldsymbol{\Sigma}_{11}\boldsymbol{\beta}_1^* + \boldsymbol{\beta}_2^\top\boldsymbol{\Sigma}_{\bullet 22}\boldsymbol{\beta}_2\end{aligned}$$

Therefore,

$$\mathrm{E}\Big[\big(\vec{\boldsymbol{x}}\,\boldsymbol{\beta}\big)^2\Big] \;=\; \boldsymbol{\beta}_1^{*T}\,\boldsymbol{\Sigma}_{11}\,\boldsymbol{\beta}_1^* \;+\; \boldsymbol{\beta}_2^\top\,\boldsymbol{\Sigma}_{\bullet 22}\,\boldsymbol{\beta}_2.$$

Recall

$$\mathrm{E}\Big[\big(\vec{\boldsymbol{x}}\,\boldsymbol{\beta}\big)^2\Big] \;=\; \boldsymbol{\beta}^\top\,\boldsymbol{\Sigma}\,\boldsymbol{\beta}.$$

Hence

$$\boldsymbol{\beta}^\top\,\boldsymbol{\Sigma}\,\boldsymbol{\beta} \;=\; \boldsymbol{\beta}_1^{*T}\,\boldsymbol{\Sigma}_{11}\,\boldsymbol{\beta}_1^* \;+\; \boldsymbol{\beta}_2^\top\,\boldsymbol{\Sigma}_{\bullet 22}\,\boldsymbol{\beta}_2.$$

Finally, we have the decomposition of h^2 as

$$\begin{aligned}
h_\beta^2 &= \frac{\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta}}{\sigma_{y,\beta}^2} = \frac{\boldsymbol{\beta}_1^{*T}\boldsymbol{\Sigma}_{11}\boldsymbol{\beta}_1^*}{\sigma_{y,\beta}^2} + \frac{\boldsymbol{\beta}_{\mathbf{2}}^\top\boldsymbol{\Sigma}_{\bullet\mathbf{22}}\boldsymbol{\beta}_{\mathbf{2}}}{\sigma_{y,\beta}^2} \\
&= \frac{\boldsymbol{\beta}_1^{*T}\boldsymbol{\Sigma}_{11}\boldsymbol{\beta}_1^*}{\sigma_{y,\beta}^2} + \frac{\boldsymbol{\beta}_{\mathbf{2}}^\top\boldsymbol{\Sigma}_{\bullet\mathbf{22}}\boldsymbol{\beta}_{\mathbf{2}}}{\sigma_{y_{\bullet 2},\beta}^2}\frac{\sigma_{y_{\bullet 2},\beta}^2}{\sigma_{y,\beta}^2} \\
&= h_{1,\beta}^2 + \lambda_\beta h_{\bullet 2,\beta}^2,
\end{aligned}$$

where $\lambda_\beta = \sigma_{y_{\bullet 2},\beta}^2/\sigma_{y,\beta}^2$. □

B.4 Proof of Proposition 5.1

Part 1: Following Model (1), we have $\mathrm{E}[\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\beta}] = \boldsymbol{X}\boldsymbol{\beta}$ and $\mathrm{Var}(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\beta}) = \sigma_\varepsilon^2\boldsymbol{I}_n$. With the sample covariance matrix of $\boldsymbol{X}$ is $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X}$ and the correlation $\hat{\boldsymbol{\eta}} = \frac{1}{n}\boldsymbol{X}^\top\boldsymbol{y}$:

$$\begin{aligned}
\mathrm{E}(\hat{\boldsymbol{\eta}}|\boldsymbol{X},\boldsymbol{\beta}) &= \frac{1}{n}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\beta} = \hat{\boldsymbol{\Sigma}}\boldsymbol{\beta} \\
\mathrm{Var}[\hat{\boldsymbol{\eta}}|\boldsymbol{X},\boldsymbol{\beta}] &= \frac{\sigma_\varepsilon^2}{n^2}\boldsymbol{X}^\top\boldsymbol{X} = \frac{\sigma_\varepsilon^2}{n}\hat{\boldsymbol{\Sigma}}.
\end{aligned}$$

With the assumption that $\mathrm{E}(\boldsymbol{\beta}) = 0$ and $\mathrm{Var}[\boldsymbol{\beta}] = h^2\boldsymbol{I}/m$, by using the law of total variance, we can have the expression of the covariance matrix $\boldsymbol{V}$ given $\boldsymbol{X}$:

$$\begin{aligned}
\mathrm{Var}[\hat{\boldsymbol{\eta}}|\boldsymbol{X}] = \boldsymbol{V} &= \mathrm{E}(\mathrm{Var}[\hat{\boldsymbol{\eta}}|\boldsymbol{X},\boldsymbol{\beta}]) + \mathrm{Var}[\mathrm{E}(\hat{\boldsymbol{\eta}}|\boldsymbol{X},\boldsymbol{\beta})] \\
&= \mathrm{E}\left(\frac{\sigma_\varepsilon^2}{n}\hat{\boldsymbol{\Sigma}}\right) + \mathrm{Var}[\hat{\boldsymbol{\Sigma}}\boldsymbol{\beta}] \\
&= \frac{\sigma_\varepsilon^2}{n}\hat{\boldsymbol{\Sigma}} + \hat{\boldsymbol{\Sigma}}^\top\mathrm{Var}[\boldsymbol{\beta}]\hat{\boldsymbol{\Sigma}} \\
&= \frac{1-h^2}{n}\hat{\boldsymbol{\Sigma}} + \frac{h^2}{m}\hat{\boldsymbol{\Sigma}}^\top\hat{\boldsymbol{\Sigma}}.
\end{aligned}$$

Part 2: Substituting the model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, we have

$$\hat{\boldsymbol{\eta}} = \frac{1}{n}\boldsymbol{X}^\top\boldsymbol{y} = \frac{1}{n}\boldsymbol{X}^\top(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \hat{\boldsymbol{\Sigma}}\boldsymbol{\beta} + \frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\varepsilon}.$$

Since we assume $\boldsymbol{\beta} \sim N(\mathbf{0}, (h^2/m)\boldsymbol{I})$, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma_\varepsilon^2 \boldsymbol{I})$ and $\sigma_\varepsilon^2 = 1 - h^2$, with $\boldsymbol{\beta}$ being independent of $\boldsymbol{\varepsilon}$ in Model (1), we have

$$\hat{\boldsymbol{\Sigma}}\boldsymbol{\beta}|\boldsymbol{X} \sim N(\mathbf{0}, \frac{h^2}{m}\hat{\boldsymbol{\Sigma}}^2), \quad \frac{1}{n}\boldsymbol{X}^\top \boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(\mathbf{0}, \frac{1-h^2}{n^2}\boldsymbol{X}^\top \boldsymbol{X}) = N(\mathbf{0}, \frac{1-h^2}{n}\hat{\boldsymbol{\Sigma}}).$$

As both terms are Gaussian and independent given $\boldsymbol{X}$, it follows that their summation $\hat{\boldsymbol{\eta}}$ is also Gaussian conditional on $\boldsymbol{X}$:

$$\hat{\boldsymbol{\eta}} \mid \boldsymbol{X} \sim N\left(\mathbf{0}, \frac{1-h^2}{n}\hat{\boldsymbol{\Sigma}} + \frac{h^2}{m}\hat{\boldsymbol{\Sigma}}^2\right) = N(0, \boldsymbol{V}).$$

□

C Decomposed FVE Estimation Performance in terms of Relative Bias

As shown in Figure 14 and Figure 15, in both low- and high-dimensional settings, the decomposed estimates with GWASH and GCTA consistently exhibit smaller bias than the non-decomposed estimates.

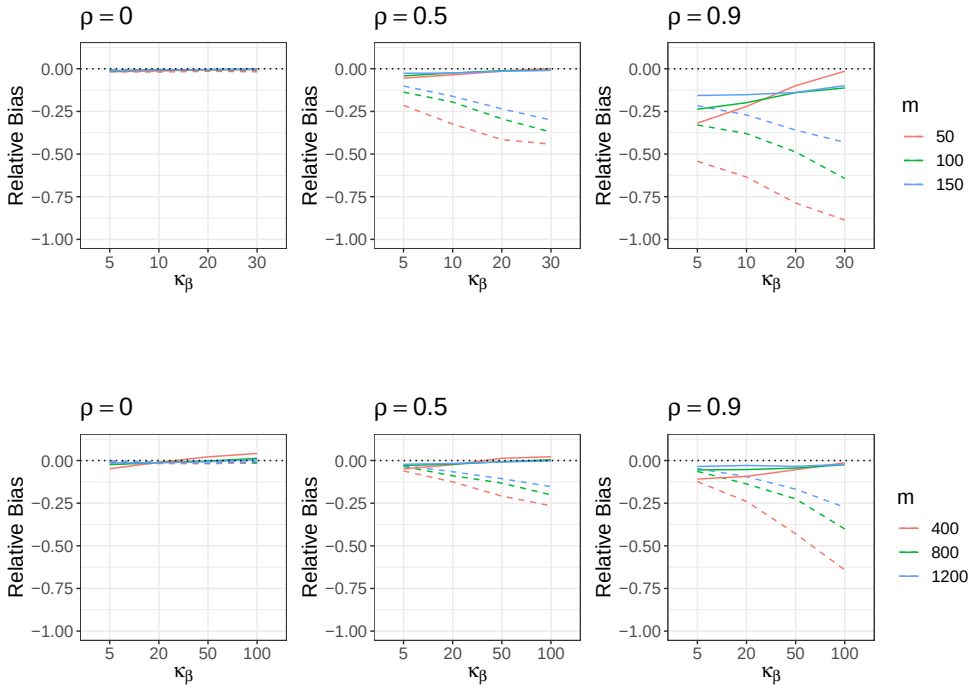

Figure 14: Estimation performance of GWASH without decomposition (dashed line) and decomposed GWASH (solid line) in the low-dimensional setting (top row) and high-dimensional setting (bottom row). The standard error from 1000 simulation instances is about 0.002 (low-dimension) and 0.003 (high-dimension).

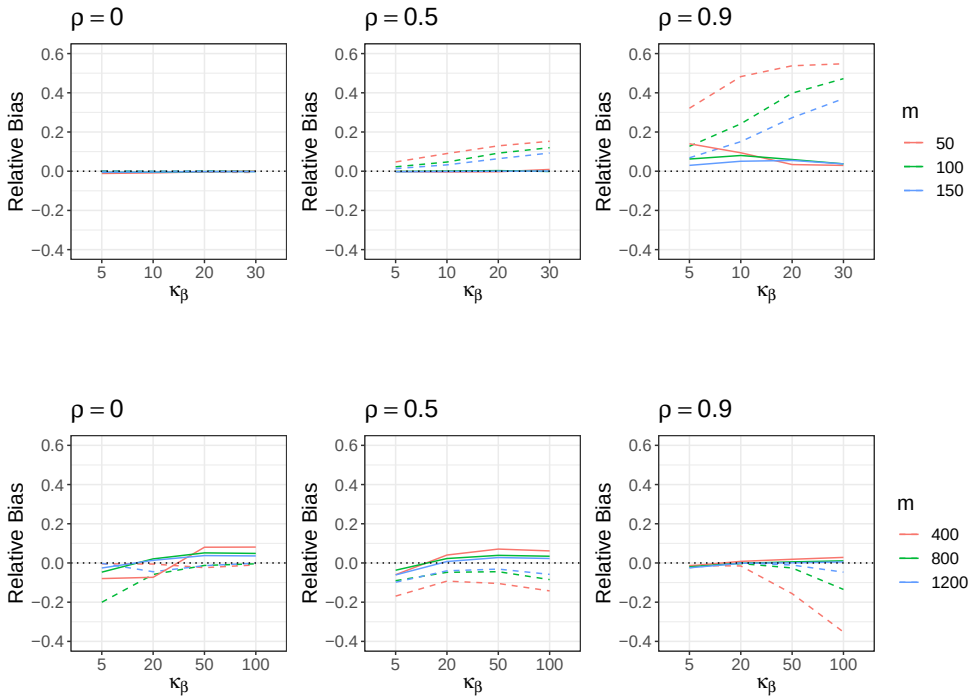

Figure 15: Estimation performance of the GCTA without decomposition (dashed line) and decomposed GCTA (solid line) under the high-dimension setting. The standard error from 1000 simulation instances is about 0.0015 (low-dimension) and 0.0018 (high-dimension).