Journal of Machine Learning Research 21 (2020) 1-67 Submitted 11/18; Revised 6/20; Published 6/20

A Regularization-Based Adaptive Test for High-Dimensional Generalized Linear Models

Chong Wu* CWU3@FSU.EDU
Department of Statistics, Florida State University, FL, USA

Gongjun Xu GONGJUN@UMICH.EDU
Department of Statistics, University of Michigan, MI, USA

Xiaotong Shen XSHEN@UMN.EDU
School of Statistics, University of Minnesota, MN, USA

Wei Pan* PANXX014@UMN.EDU
Division of Biostatistics, University of Minnesota, MN, USA

Editor: Jon McAuliffe

Abstract

In spite of its urgent importance in the era of big data, testing high-dimensional parameters in generalized linear models (GLMs) in the presence of high-dimensional nuisance parameters has been largely under-studied, especially with regard to constructing powerful tests for general (and unknown) alternatives. Most existing tests are powerful only against certain alternatives and may yield incorrect Type I error rates under high-dimensional nuisance parameter situations. In this paper, we propose the adaptive interaction sum of powered score (aiSPU) test in the framework of penalized regression with a non-convex penalty, called truncated Lasso penalty (TLP), which can maintain correct Type I error rates while yielding high statistical power across a wide range of alternatives. To calculate its p-values analytically, we derive its asymptotic null distribution. Via simulations, its superior finite-sample performance is demonstrated over several representative existing methods. In addition, we apply it and other representative tests to an Alzheimer's Disease Neuroimaging Initiative (ADNI) data set, detecting possible gene-gender interactions for Alzheimer's disease. We also put R package "*aispu*" implementing the proposed test on GitHub.

Keywords: Adaptive Test, Truncated Lasso Penalty, Gene-Environmental Interaction

1. Introduction

Statistical inference in high-dimensional models has been attracting increasing attentions, owing to the surge of high-dimensional data in many fields, such as in genetics and genomics. Accordingly, there is an increasing body of literature on significance testing in high-dimensional linear or generalized linear models (GLMs), mostly on low-dimensional regression coefficients. For example, Wasserman and Roeder (2009); Meinshausen et al. (2009) proposed random sample-splitting approaches to testing on a regression coefficient of interest in a high-dimensional model. Based on the idea of polyhedral selection, Lee et al. (2016) proposed an exact post-selection estimator conditional on the selection event.

∗. C.W. and W. P. are the corresponding authors.

Meanwhile, many researchers exploit the idea of projection or bias-correction to handle the impact of regularization and high-dimensional nuisance parameters (e.g., Javanmard and Montanari, 2014; Van de Geer et al., 2014; Zhang and Zhang, 2014; Lee et al., 2016; Shi et al., 2019; Ma et al., 2020). In spite of exciting progresses in the last few years, little work has been done to construct more general and powerful tests on high-dimensional regression coefficients in GLMs in the presence of high-dimensional nuisance parameters.

It is noted that, for high-dimensional problems, classical or popular tests may not perform well, even if their asymptotic properties (such as their null distributions) are well established. Fan (1996) gave a simple example: given a p-dimensional vector follows a normal distribution, $\mathbf{y} \sim N(\theta, I)$, to test H_0: $\theta = 0$ versus H_A: $\theta \neq 0$, the likelihood ratio test, Wald test, and score test statistics all share the same form $T = ||\mathbf{y}||_{\mathbf{2}}^{\mathbf{2}}$, which is a sum of squares-type statistic; even under an alternative H_A with $||\theta||_2^2 \to \infty$ as $p \to \infty$, as long as $||\theta||_2^2 = o(\sqrt{p})$, the power of the three tests vanishes (i.e. tending to the Type I error rate); in contrast, some adaptive tests can be much more powerful. This example convincingly demonstrates the importance of considering the power of a test in high-dimensional settings and this article aims at filling this gap.

This work was motivated by a critical problem in genetics to identify interaction effects between a genetic marker set and a complex disease like Alzheimer's. Although univariate single nucleotide polymorphism (SNP) based analyses for identifying gene-environment ($G \times E$) interactions are popular in the community, relatively few of the findings have been replicated (Manuck and McCaffery, 2014). To improve statistical power and enhance results interpretation, many genetic association studies have now considered an alternate/supplementary approach to jointly test the interaction effect of all SNPs in a biological meaningful marker set, e.g., SNPs in a gene or a pathway (Lin et al., 2013, 2016; Su et al., 2017). Jointly testing the interaction effect of a marker set can be formulated as testing on a high-dimensional parameter (i.e., interactions between possibly high-dimensional genetic variants and environmental factors) in the presence of high-dimensional nuisance parameters (to adjust for the main effects of the genetic variants and other covariates) under a high-dimensional GLM. Since such interactions in human genetics have proven difficult to detect, while specific interaction patterns are largely unknown, it is critical to develop and apply more general and adaptive tests that are powerful across a wide range of unknown alternatives.

To account for impact of high-dimensional nuisance parameters in $G \times E$ interactions testing problems, some variance-component score tests with the sum of squares-type statistics, coupled with the ridge regression to estimate the nuisance parameters under the null, have been proposed (Lin et al., 2013, 2016; Su et al., 2017). For example, Lin et al. (2013) proposed a test called gene-environment set association test (GESAT) by assuming that the $G \times E$ interaction effects follow an unspecified distribution with mean 0 and variance υ^2, then testing $H_0 : \upsilon^2 = 0$ for the overall $G \times E$ interaction. By noting that the ridge estimator is $\sqrt{n}$-consistent under suitable conditions (Knight and Fu, 2000), they derived the theoretical null distribution for GESAT. While enticing, the $\sqrt{n}$-consistency of the ridge estimator or asymptotic normality of the score vector may not be applicable under high-dimensional situations with finite samples, leading to incorrect Type I error rates. As to be shown in simulations, as the number of the covariates increases, methods based on the ridge penalty yield incorrect Type I error rates and substantial power loss.

Meanwhile, bias-correction based methods have been proposed (Dezeure et al., 2017; Zhang and Cheng, 2017). For example, inspired by the desparsifying Lasso estimator (Van de Geer et al., 2014) and the data-splitting strategy (Wasserman and Roeder, 2009), Zhang and Cheng (2017) proposed a three-step bootstrap-assisted procedure based on a supremum-type statistic to test on high-dimensional regression coefficients in high dimensional regression models. This method can control the Type I error rate well and yield high statistical power under highly sparse alternatives. However, due to the accumulation of estimation errors of the desparsifying Lasso estimator, the estimation errors might be out of control if a burden-type or sum of squares-type statistic is used (Zhang and Cheng, 2017). Moreover, although the data-splitting strategy adopted therein helps control the Type I error rate, it reduces the power as well.

To address those challenges, we propose an adaptive test, referred to as adaptive interaction sum of powered score (aiSPU) test, for testing high-dimensional regression coefficients under GLMs with high-dimensional nuisance parameters. The aiSPU test is new and appealing in two aspects. First, in aiSPU we apply the truncated Lasso penalty (TLP) (Shen et al., 2012), a non-convex penalty, to estimate the high-dimensional nuisance parameter under the null hypothesis. The TLP estimator consistently reconstructs the oracle estimator under mild assumptions, helping maintain correct Type I error rates under a high-dimensional situation. In contrast, the consistency of a convex penalty-based estimator, such as the ridge or Lasso estimator, holds under much stronger conditions. For example, the Lasso estimator is consistent under a strong irrepresentable (Wainwright, 2009), while the ridge estimator is consistent under the assumption that the sample covariance matrix of all the covariates converges to a non-singular matrix (Knight and Fu, 2000). Second, because the true alternative hypothesis is generally complex and unknown, we apply the idea of an adaptive testing (Pan et al., 2014). We first construct a group of interaction sum of powered score (iSPU) tests such that hopefully at least one of them would be powerful for a given alternative. The proposed adaptive test then data-adaptively selects the one with the most significant result with a proper adjustment for multiple testing to control the Type I error rate, and thus achieves high power.

To apply the proposed test, we establish its asymptotic null distribution. In particular, we derive the joint asymptotic distribution for a set of the iSPU tests. The marginal distribution of each iSPU test statistic converges to either a normal distribution or an extreme value distribution under some conditions. Based on the theoretical results, we develop an asymptotic way to calculate the p-values for the iSPU and aiSPU simultaneously. We demonstrate the superior performance of the proposed test with some theoretical power analyses under local alternatives. Further, as to be shown in simulations and real data analyses, the proposed aiSPU test would yield correct Type I error rates and higher statistical power than several existing methods under a wide range of high-dimensional alternative hypotheses, ranging from highly dense to highly sparse alternatives.

The rest of the paper is organized as follows. In Section 2, we review two representative tests before proposing our new aiSPU test. Results for simulations and analysis of the Alzheimer's Disease Neuroimaging Initiative (ADNI) data are presented in Sections 3 and 4, respectively. We conclude with a short discussion in Section 5. All technical details, proofs, and extensive simulation results are relegated to Appendices. We

have released open source R package *aispu* implementing the proposed test on GitHub (https://github.com/ChongWu-Biostat/aispu), and will upload it to CRAN soon.

2. Methods

2.1. Notation and model

Even though our study was motivated by detecting gene-environmental interactions, our proposed method is general and applicable to many other problems, thus we introduce our method in a general framework. Suppose we have n independent and identically distributed (IID) observations $\{(Y_i, Z_i, X_i) : i = 1, 2, \ldots, n\}$, for which we denote an n-vector outcome (response) $Y = (Y_1, ..., Y_n)'$, an $n \times q$ matrix $\mathbb{Z} = (Z_1', ..., Z_n')'$ for q nuisance covariates (including the intercept term) with $Z_i = (Z_{i1}, \ldots, Z_{iq})$, and an $n \times p$ matrix $\mathbb{X} = (X_1', .., X_n')'$ for p variables of interest with $X_i = (X_{i1}, \ldots, X_{ip})$. Without loss of generality, we assume that $E(X_i) = 0$ as otherwise each X_i can be re-centered by its sample mean. We consider a generalized linear model with the canonical link function,

$$E(Y|\mathbb{X}, \mathbb{Z}) = g^{-1}(\mathbb{X}\beta + \mathbb{Z}\vartheta), \tag{1}$$

where p-vector β and q-vector ϑ are unknown parameters, and g is the canonical link function. We are interested in testing

$$H_0 : \beta = 0 \qquad \text{versus} \qquad H_1 : \beta \neq 0, \tag{2}$$

while treating ϑ as the high-dimensional nuisance parameter. We target the situation with "large q and p."

Remark 1 *Numerous real-world problems can be formulated as testing high-dimensional parameters under GLMs in the presence of high-dimensional nuisance parameters. For example, when testing the interaction between a genetic marker set and a set of environmental variables, we can let $\mathbb{Z}$ be the environmental factors, genotypes in the marker-set, and some important covariates, and let $\mathbb{X}$ be the SNP-environment interaction variables. Here we consider a large number of SNPs in a marker-set, leading to high-dimensional q and p. Another example is testing gene-gene interactions (Cordell, 2009), a problem can be formulated with $\mathbb{Z}$ being all the SNPs from two genes and $\mathbb{X}$ being their interactions.*

2.2. Related existing methods

In this subsection, we review two representative methods: a variance component type test called GESAT (Lin et al., 2013) and a bias-correction based test (Zhang and Cheng, 2017).

By assuming β_j's follow an arbitrary distribution with mean zero and variance υ^2, GESAT converts testing $H_0 : \beta = 0$ to testing $H_0' : \upsilon^2 = 0$, which can be conducted via the following sum of squares-type statistic: $Q = (Y - \mu(\hat{\vartheta}))'\mathbb{X}\mathbb{X}'(Y - \mu(\hat{\vartheta}))$, where $\mu(\hat{\vartheta}) = g^{-1}(\mathbb{Z}\hat{\vartheta})$ and $\hat{\vartheta}$ is estimated under the null model,

$$E(Y|\mathbb{Z}) = g^{-1}(\mathbb{Z}\vartheta). \tag{3}$$

To account for high-dimensionality of ϑ, GESAT applies the ridge regression to estimate ϑ under the null model (3). Using the property that the ridge estimator $\hat{\vartheta}$ is $\sqrt{n}$-consistent

under suitable conditions (Knight and Fu, 2000), they showed that test statistic Q asymptotically follows a mixture of χ^2 distributions under the H_0, thus p-values can be calculated accordingly (Lin et al., 2013).

Meanwhile, a three-step bootstrap-assisted procedure (Zhang and Cheng, 2017) has been proposed to test on H_0. First, it randomly splits the sample into two sub-samples. Second, it screens out the irrelevant variables of $\mathbb{X}$ based on the first sub-sample. After screening, denote the reduced model $\mathcal{S} = \{j : j \notin \{\text{irrelevant variables}\}\}$. Third, it computes the desparsifying Lasso estimator $\{\check{\beta}_j\}_{j\in\mathcal{S}}$ and the corresponding variance estimator $\check{w}_{jj}$ based on the second sub-sample. The non-studentized (NST) and studentized (ST) supremum type statistic are $\max_{j\in\mathcal{S}} \sqrt{n}|\check{\beta}_j|$ and $\max_{j\in\mathcal{S}} \sqrt{n}|\check{\beta}_j|/\sqrt{\check{w}_{jj}}$, respectively. Zhang and Cheng (2017) then applied a bootstrap-assisted procedure to calculate their p-values.

Though appealing, both tests have limitations. First, a test based on the ridge penalty (such as GESAT) might yield incorrect Type I error rates when the dimensionality of nuisance parameters ϑ (i.e., q) is high. Note that the null distribution of GESAT is derived based on the $\sqrt{n}$-consistent ridge estimator, which requires that the sample covariance matrix of Z_i converges to a non-singular matrix (Knight and Fu, 2000); this assumption will not hold when $q > n$. This may explain incorrect Type I error rates of GESAT as to be shown in simulations. Second, the existing tests might be powerless under some alternatives. It is well known that a sum of squares-type statistic (for example, GESAT) and a supremum-type statistic (for example, NST and ST) are more powerful for dense and highly sparse nonzero signals, respectively (Pan et al., 2014). However, for moderately dense nonzero signals, neither may be powerful: there might not exist one or few components of β to represent a strong departure from H_0, whereas a sum of squares statistic might accumulate too much noises or estimation errors through summing over the non-informative components. Furthermore, both NST and ST only use a sub-sample to construct test statistics, further reducing power. As to be shown in simulations, the above methods would lose substantial power under some alternatives.

2.3. New test statistics

There are two main challenges for constructing a powerful test in a high-dimensional setting. First, estimating the high-dimensional ϑ under H_0 is not trivial. Second, because the underlying association patterns are unknown, it is crucial to construct an adaptive test such that it can maintain high power across a wide range of alternatives.

To accurately estimate the high-dimensional ϑ under the null model (3), we apply penalized regression by imposing the truncated Lasso penalty (TLP) (Shen et al., 2012) on the nuisance parameter ϑ. For gene-environmental interaction problems, we can impose no penalty on a subset of some pre-specified low-dimensional covariates to keep some important covariates, such as age and gender. TLP is defined as $\text{TLP}(x, \tau) = \min(|x|, \tau)$ for a scalar x and a tuning parameter τ. It can be regarded as the Lasso penalty for a small $|x| \le \tau$, but imposes no penalty for a large $|x| > \tau$. We use 10-fold cross-validation to select the tuning parameters for TLP and denote $\hat{\vartheta}$ as the TLP estimate of ϑ under H_0.

To maintain high power across various alternatives, we construct an adaptive test. Up to some constant, the score vector $U = (U_1, \ldots, U_p)'$ for β in (1) is $U_j = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\mu}_{0i})X_{ij}$

for $1 \le j \le p$, where $\hat{\mu}_{0i} = g^{-1}(Z_i\hat{\vartheta})$. Denote

$$U_{ij} = (Y_i - \hat{\mu}_{0i})X_{ij}$$

for $1 \le i \le n$ and $1 \le j \le p$. We first propose a class of test statistics called interaction sum of powered score (iSPU) with power index $\gamma > 0$ as

$$L(\gamma) = \sum_{j=1}^{p} \Big(\frac{1}{n}\sum_{i=1}^{n} U_{ij}\Big)^{\gamma}.$$

Since $L(\gamma)^{1/\gamma} \to \max_{1\le i\le p}\left|\frac{1}{n}\sum_{i=1}^{n} U_{ij}\right|$ as even integer $\gamma \to \infty$, we define $L(\infty)$ as

$$L(\infty) = \max_{1\le j\le p} \frac{n\left(\frac{1}{n}\sum_{i=1}^{n} U_{ij}\right)^2}{\check{\sigma}_{jj}},$$

where $\check{\mathbf{\Sigma}} = (\check{\sigma}_{kj})_{p\times p}$ is the covariance matrix with $\check{\sigma}_{kj} = \mathrm{cov}[U_{1k}, U_{1j}]$ for $1 \le k, j \le p$.

As to be shown in simulations, the power of $L(\gamma)$ depends on the unknown β under specific alternatives. Since in general there is no uniformly most-powerful test, to maintain high power across various alternatives, we propose the adaptive interaction sum of powered score (aiSPU) to combine the multiple iSPU tests with different γ:

$$T_{\mathrm{aiSPU}} = \min_{\gamma\in\Gamma} P_{L(\gamma)},$$

where $P_{L(\gamma)}$ is the p-value for $L(\gamma)$ and Γ contains the candidate values of γ, e.g., $\Gamma = \{1, 2, \ldots, 6, \infty\}$. We take the minimum p-value to approximately select the most powerful candidate test; T_{aiSPU} is the test statistic, but no longer a genuine p-value.

To emphasize the penalty we use, in some places, we denote iSPU(γ) and aiSPU explicitly with the penalty, say TLP, as iSPU(TLP,γ) and aiSPU(TLP), respectively.

Remark 2 *Accurate estimation of ϑ under the null is crucial in the situation with a high-dimensional nuisance parameter. Because the $\sqrt{n}$-consistency of the ridge estimator may not hold under a (relatively) high-dimensional situation, a test coupled with ridge regression may yield incorrect Type I error rates. The estimation errors of the desparsifying Lasso estimator might be out of control if a burden-type or sum of squares-type statistic is used (Zhang and Cheng, 2017), while the three-step bootstrap-assisted procedure based on a supremum-type statistic will not be powerful under dense alternatives. In contrast, because TLP enjoys the selection consistency and optimal parameter estimation under some mild conditions (Shen et al., 2012), aiSPU controls Type I error rates and achieves high power under a wide range of high-dimensional situations.*

Remark 3 *The proposed aiSPU test can be viewed as an extension of the aSPU test (Wu et al., 2019) to high-dimensional nuisance parameter situations. The aSPU test was proposed for the situations with large p but small q, while the aiSPU test targets situations with large p and large q. Thus, the Type I error rate can be controlled by aiSPU, but not by aSPU in high-dimensional nuisance parameter situations (large q).*

Remark 4 *Our proposed test may share some limitations of the standard score test with possible loss of power under H_A, which can be fixed by taking an approach as shown in Wang (2016).*

2.4. Asymptotic null distribution

In this subsection, we derive the asymptotic null distribution for iSPU. Before stating the theorem, we define necessary notation as follows. Let $\mu_0 \equiv (\mu_{01}, \ldots, \mu_{0n})'$ be the conditional mean of Y under H_0, where $\mu_{0i} = E(Y_i|H_0) = E(Y_i|Z_i) = g^{-1}(Z_i\vartheta^0)$ and ϑ^0 is the population value of ϑ. Write $S_{ij} = (Y_i - \mu_{0i})X_{ij}$ for $1 \le i \le n$ and $1 \le j \le p$. We further define the corresponding covariance matrix $\mathbf{\Sigma} = (\sigma_{kj})_{p\times p}$ with $\sigma_{kj} = \mathrm{Cov}[S_{1k}, S_{1j}]$ for $1 \le k, j \le p$. For simplicity, we denote $L(\gamma, \mu_0) = \sum_{j=1}^{p} L^{(j)}(\gamma, \mu_0)$ with $L^{(j)}(\gamma, \mu_0) = \left(\frac{1}{n}\sum_{i=1}^{n} S_{ij}\right)^{\gamma}$ for $1 \le i \le n$ and $1 \le j \le p$. Then the mean and variance of $L(\gamma, \mu_0)$ can be denoted by $\psi(\gamma) = \sum_{j=1}^{p} \psi^{(j)}(\gamma)$ with $\psi^{(j)}(\gamma) = E\left[L^{(j)}(\gamma, \mu_0)|H_0\right]$, and by $\omega^2(\gamma) = \mathrm{var}[L(\gamma, \mu_0)|H_0]$, respectively.

Theorem 1 shows that (1) each iSPU(γ) converges to either a normal distribution or an extreme value distribution; (2) iSPU(∞) and iSPU with a finite γ are asymptotically independent under the H_0.

Theorem 1. *Under assumptions C1–C7 stated in Appendix A and under the null hypothesis H_0, for any fixed and finite Γ set we have:*
(i) For finite candidate values γ in Γ, that is, $\Gamma' = \Gamma \setminus \{\infty\}$, the vector of the iSPU test statistics $[\{L(\gamma) - \psi(\gamma)\}/\omega(\gamma)]'_{\gamma\in\Gamma'}$ converges weakly to a normal distribution with mean 0 and covariance matrix $\mathbf{R}(\Gamma') = (\rho_{st})$, i.e., $N(0, \mathbf{R}(\Gamma'))$ as $n, p \to \infty$, where $\psi(\gamma)$, $\omega(\gamma)$, and $\mathbf{R}(\Gamma')$ are defined in Appendix B.
(ii) For $\gamma = \infty$, let $a_p = 2\log p - \log\log p$, for any real number x, $Pr\{L(\infty) - a_p \le x\} \to \exp\{-\pi^{-1/2}\exp(-x/2)\}$.
(iii) $[\{L(\gamma) - \psi(\gamma)\}/\omega(\gamma)]'_{\gamma\in\Gamma'}$ is asymptotically independent of $L(\infty)$, that is, the joint distribution of $[\{L(\gamma) - \psi(\gamma)\}/\omega(\gamma)]'_{\gamma\in\Gamma'}$ and $L(\infty) - a_p$ converges weakly to the product of the limiting distributions given in (i) and (ii).

Remark 5 *We leave the technical details and assumptions into the Appendices A–C. Intuitively speaking, $L(\gamma, \mu_0)$ with a finite γ follows a normal distribution asymptotically when S_{ij_1} and S_{ij_2} are independent for $j_1 \ne j_2$. Under moment assumptions that put constraints on correlation structures, we prove that the asymptotically normal still holds for $L(\gamma)$ with a finite γ and a TLP-based estimate $\hat{\vartheta}$. For $L(\infty)$, we derive its distribution based on theorems in Cai et al. (2014). Of note, Wu et al. (2019) derived a similar Theorem under a much simpler context; our current proof is different and more challenging due to the technical complications under the adopted penalized regression framework to deal with the presence of high-dimensional nuisance parameters.*

Remark 6 *In Theorem 1, we assume technical assumptions, such as a sparsity assumption regarding the effect from $\mathbb{Z}$ (C3) and a feature selection assumption involving Hellinger distance (C7). Because C7 is hard to validate in practice, we propose a stronger than needed beta-min like condition C7*, which is a sufficient assumption for C7. As to be shown in simulations, when the effect from $\mathbb{Z}$ is non-sparse but with sparse strong signals, our proposed method still works. In other words, under situations where C3 and C7* have been violated but C7 might hold, our proposed method still works.*

These technical assumptions are used to establish the difference between μ_0 and $\hat{\mu}_0$ is a small order term, which can be ignored in the theoretical derivation. Once we have a good

estimate of the conditional mean of Y under H_0 (i.e., μ_0), our proposed method works. In principle, our proposed aiSPU test can be extended to consider higher-order interactions within $\mathbb{Z}$ if μ_0 can be well estimated. We leave this interesting topic for future study.

Next, we briefly discuss how to calculate p-values and leave the detailed procedures into the Appendix B. According to Theorem 1, we can calculate p-values asymptotically. The p-values for individual iSPU(γ) can be calculated via either a normal or an extreme value distribution. The p-value for aiSPU can be calculated by the following two steps. First, by $\text{cov}[L(t,\mu_0), L(s,\mu_0)] = o(pn^{-(t+s)/2})$ if $s+t$ is odd and by Theorem 1 part three, iSPU with even γ, odd γ, $\gamma = \infty$ are asymptotically independent to each other (see Appendix B for details). Because for a finite γ, $L(\gamma) - \psi(\gamma)/\omega(\gamma)$ follows a standard normal distribution, taking the minimum p-value as test statistics equals to taking the maximum of $|L(\gamma) - \psi(\gamma)|/\omega(\gamma)$ as the test statistics. Further define $t_O = \max_{\text{odd } \gamma\in\Gamma} \big|\big(L(\gamma) - \psi(\gamma)\big)\big|/\omega(\gamma)$ and $t_E = \max_{\text{even } \gamma\in\Gamma} \big(L(\gamma) - \psi(\gamma)\big)/\omega(\gamma)$ as the observed test statistics from the data and calculate the p-values for t_O, t_E, and $L(\infty)$ as $p_O = Pr[\max_{\text{odd } \gamma\in\Gamma} |\big(L(\gamma) - \psi(\gamma)\big)/\omega(\gamma)| > t_O]$, $p_E = Pr[\max_{\text{even } \gamma\in\Gamma} \big(L(\gamma) - \psi(\gamma)\big)/\omega(\gamma) > t_E]$, and p_∞ equals to the p-value of iSPU(∞). Specifically, we use *pmvnorm()* in R package *mvrnorm* to calculate the normal tail probabilities of p_O and p_E. Second, we take the minimum p-value from the above three categories, that is, $p_{\min} = \min\{p_O, p_E, p_\infty\}$. By the asymptotic independence among p_O, p_E, and p_∞, the asymptotic p-value for the aiSPU test is $p_{\text{aiSPU}} = 1 - (1 - p_{\min})^3$.

Of note, calculating $\psi(\gamma)$, $\omega(\gamma)$ and $\mathbf{R}(\Gamma')$ involves $\mathbf{\Sigma} = (\sigma_{kj})$, which is unknown and has to be estimated in practice. We apply either the banding method of Bickel and Levina (2008) or a parametric bootstrap-based method to estimate covariance matrix $\mathbf{\Sigma}$ (see Remark S2 in Appendix B for details).

Meanwhile, we can calculate p-values by the parametric bootstrap (see Appendix B for details). The parametric bootstrap may estimates more accurately the p-values than the asymptotics-based method, but it is highly computational extensive, especially at a high significance level. To facilitate data analyses in the wider community, we have developed an R package "*aispu*", implementing both methods.

Remark 7 *We recommend using the asymptotic-based method when p is large and using the parametric bootstrap-based method when p is small. Our proposed asymptotic-based method may have a better performance when p is large due to two reasons. First, residual bootstrap and pairs bootstrap are known to be problematic under a high-dimensional setting (El Karoui and Purdom, 2018). We expect that parametric bootstrap may have a similar problem when p is large. Second, by Theorem 1, estimation error can be ignored as both n and p go to infinity. On the other hand, when p is small, the parametric bootstrap-based method may achieve superior performance than the asymptotic-based method because the asymptotic theory in Theorem 1 may not hold.*

2.5. Asymptotic power analysis

We analyze the asymptotic power of the aiSPU test. Under an alternative $H_A : \beta \neq 0$, we first derive approximations to the mean and variance of $L(\gamma, \mu_0)$ with $\gamma < \infty$, denoted by $\psi_A(\gamma) = E[L(\gamma,\mu_0)|H_A]$ and by $\omega_A^2(\gamma) = \text{var}[L(\gamma,\mu_0)|H_A]$, respectively. Then we derive

the asymptotic power under a local alternative. In the end, we discuss the choice of the Γ set. To save space, we put technical details into the Appendix D.

First, we define some necessary notations. Let β^0 be the true value of β and $\mu_0^A \equiv (\mu_{01}^A, \ldots, \mu_{0n}^A)'$ with $\mu_{0i}^A = E(Y_i|X_i, Z_i; H_A) = g^{-1}(X_i\beta^0 + Z_i\vartheta^0)$ being the conditional mean of Y_i under H_A. We further define $\tilde{\psi}(\gamma) = E\left[L(\gamma, \mu_0^A)|H_A\right]$ and $\tilde{\omega}^2(\gamma) = \text{var}[L(\gamma, \mu_0^A)|H_A]$.

The high dimensionality of X_i makes the identification of the leading order term of the test statistic $L(\gamma)$ quite challenging. Here, we consider a local alternative such that for $j = 1, 2, \ldots, p$, $\Delta_j = E[(\mu_{01}^A - \mu_{01})X_{1j}] = O(n^{-1/2}(\log p)^\kappa)$ with $\kappa > 0$, which allows the identification of the leading order term. This condition restricts that Δ_j is a small term, which further implies that $\psi_A(\gamma) - \tilde{\psi}(\gamma)$ and $\omega_A^2(\gamma) - \tilde{\omega}^2(\gamma)$ are relatively small. Under the local alternative, we denote the set of locations of the signal variables by $\mathcal{S}_\eta = \{j : \Delta_j \neq 0; 1 \leq j \leq p\}$ and the cardinality of $\mathcal{S}_\eta$ by $p^{1-\eta}$, where $0 \leq \eta \leq 1$ is the parameter controlling the sparsity level.

We now analyze the power of the proposed aiSPU test. Let p_α be the critical threshold for the aiSPU test under H_0 with the significance level α. Because $T_{\text{aiSPU}} = \min_{\gamma\in\Gamma} P_{L(\gamma)}$, the statistical power under H_A satisfies $Pr(T_{\text{aiSPU}} = \min_{\gamma\in\Gamma} P_{\text{iSPU}(\gamma)} < p_\alpha) \geq Pr(P_{\text{iSPU}(\gamma)} < p_\alpha)$. Thus the asymptotic power of aiSPU is 1 if there exists a $\gamma \in \Gamma$ such that $Pr(P_{\text{iSPU}(\gamma)} < p_\alpha) \to 1$. In other words, to study the asymptotic power of the aiSPU, we only need to discuss the power of iSPU(γ) for $\gamma \in \Gamma$. For that purpose, Theorem 2 shows the asymptotic distribution of $L(\gamma, \mu_0)$ with any finite and fixed γ under $0 \leq \eta < 1/2$.

Theorem 2. *Under the assumptions C8–C9 in Appendix A and the alternative H_A with $0 \leq \eta < 1/2$ and $\Delta_j = O(n^{-1/2}(\log p)^\kappa)$ with $\kappa > 0$, for any fixed and finite Γ' set, $[\{L(\gamma, \mu_0) - \psi_A(\gamma)\}/\omega_A(\gamma)]'_{\gamma\in\Gamma'}$ converges weakly to a multivariate normal distribution with mean zero as $n, p \to \infty$.*

Remark 8 *Under the local alternative $0 \leq \eta < 1/2$, by noting that $(\log p)^{c\kappa}/p^\eta = o(1)$, we have $\psi_A(\gamma) - \tilde{\psi}(\gamma) = \sum_{j=1}^p \sum_{c=1}^\gamma \binom{\gamma}{c} \Delta_j^c O(n^{-(\gamma-c)/2}) = o(pn^{-\gamma/2})$. Similarly, we have $\omega_A^2(\gamma) - \tilde{\omega}^2(\gamma) = o(pn^{-\gamma})$. Then a proof similar to that of Theorem 1 for any fixed and finite Γ set (part one) yields Theorem 2.*

For simplicity, we assume μ_0 is known under H_A and derive Theorem 2 with $L(\gamma) = L(\gamma, \mu_0)$. While this simplification ignores the estimation errors of $\hat{\mu}_0$ and thus induces a gap between Theorem 2 and our proposed test, Theorem 2 still provides useful insights regarding which iSPU(γ) achieves the highest power under different alternatives. These insights are in line with our simulation results. To establish Theorem 2 with estimated $\hat{\mu}_0$ is quite challenging because we need to estimate and quantify the estimation error of $\hat{\mu}_0$ under a misspecified model, which is unknown and an interesting question. We leave it for future research.

Theorem 2 gives the asymptotic power of iSPU(γ) at the significance level p_α as

$$Pr(P_{\text{iSPU}(\gamma)} < p_\alpha) = \begin{cases} \Phi\Big\{\frac{\psi_A(\gamma) - \tilde{\psi}(\gamma) - z_{p_\alpha}\tilde{\omega}(\gamma)}{\omega_A(\gamma)}\Big\}, & \gamma \text{ is even}, \\ \Phi\Big\{\frac{\psi_A(\gamma) - \tilde{\psi}(\gamma) - z_{p_\alpha/2}\tilde{\omega}(\gamma)}{\omega_A(\gamma)}\Big\} + \Phi\Big\{-\frac{\psi_A(\gamma) - \tilde{\psi}(\gamma) + z_{p_\alpha/2}\tilde{\omega}(\gamma)}{\omega_A(\gamma)}\Big\}, & \gamma \text{ is odd}, \end{cases}$$

where Φ and z_{p_α} is the standard normal cumulative distribution function and its $(1 - p_\alpha)$th quantile, respectively. Because $\tilde{\omega}(\gamma)/\omega_A(\gamma)$ is bounded, the asymptotic power of iSPU(γ) is

mainly determined by $\{\psi_A(\gamma)-\tilde{\psi}(\gamma)\}/\omega_A(\gamma)$. Further note that $\omega_A(\gamma)$ is of order $p^{1/2}n^{-\gamma/2}$ and thus the power goes to 1 if $(\psi_A(\gamma)-\tilde{\psi}(\gamma))n^{\gamma/2}p^{-1/2} \to \infty$. In particular, the asymptotic power of iSPU(1) and iSPU(2) goes to 1 if $p^{-1/2}n^{1/2}\sum_i \Delta_i \to \infty$ and $p^{-1/2}n\sum_i \Delta_i^2 \to \infty$, respectively.

Note that iSPU(∞) is expected to lose power substantially when $\max_j |\Delta_j|$ is small, i.e., $\max_j |\Delta_j| = o(\log(p)^{1/2}n^{-1/2})$ (Cai et al., 2014), while iSPU(1) and iSPU(2) are expected to be powerful under dense but weak signals (e.g., $\max_j |\Delta_j| = o(n^{-1/2})$) alternatives. Thus, we discuss dense alternatives ($0 \le \eta < 1/2$) and sparse alternatives ($\eta \ge 1/2$) separately.

Under different dense alternatives, different iSPU(γ) tests achieve the highest power. To further study the power of different iSPU tests and gain insights about how to choose the Γ set, we consider a particular alternative where the Δ_j is fixed at the same level. To be specific, we consider the local alternative such that $\Delta_1 = \cdots = \Delta_p = \Delta = n^{-1/2}r^{1/2}$, where $r \to 0$ as $n, p \to \infty$. As shown in the Appendix D, under this alternative, iSPU(1) is more powerful than any other iSPU(γ) tests. Similarly, we show that iSPU(2) is asymptotically more powerful than other iSPU(γ) tests under the alternative where the absolute values of the Δ_j are the same but about half being positive while the other half being negative.

We then briefly discuss the sparse alternatives with $\eta > 1/2$. Under the sparse H_A with $\eta \ge 1/2$, any iSPU test with a finite γ loses power. For example, for any $\eta < 1/2$, the power of iSPU(1) converges to 1 when $p^{-1/2}n^{1/2}\sum_j \Delta_j \to \infty$; however, $\Delta_j = O\big(n^{-1/2}(\log p)^{\kappa}\big)$ and $\sum_j \Delta_j = p^{1-\eta}O\big(n^{-1/2}(\log p)^{\kappa}\big)$, leading to $p^{-1/2}n^{1/2}\sum_j \Delta_j \sim p^{1/2-\eta}(\log p)^{\kappa} \to 0$ when $\eta > 1/2$. Thus the asymptotic power of iSPU(1) is strictly less than 1 when $\eta \ge 1/2$. For other finite γ, we have similar results. On the other hand, a supremum-type test like iSPU(∞) is known to be powerful against sparse alternatives (Cai et al., 2014), therefore, the asymptotic power of aiSPU is 1 if that of iSPU(∞) converges to 1.

Overall, we recommend including small γ values such as $1, 2$ to maintain high power under dense alternatives. As to be shown in simulations, iSPU with a medium γ value is often the most powerful in a finite sample. To achieve a balance between the asymptotic and finite-sample performances, including medium γ values such as $3, \ldots, 6$ in Γ is recommended. This recommendation is also supported by our previous studies (Xu et al., 2016; Wu et al., 2019). Because iSPU(∞) is powerful under the sparse alternative, we recommend including ∞ in Γ. In summary, we recommend use $\Gamma = \{1, 2, \ldots, 6, \infty\}$ as our default setting.

3. Simulations

3.1. Simulation settings

To facilitate fair and unbiased comparisons, we adopted the simulation settings similar to those in Lin et al. (2013); Zhang and Cheng (2017).

Simulation settings for $G \times E$ interactions. We simulated genotypes as in Wang and Elston (2007). First, a latent vector $s = (s_1, \ldots, s_p)'$ was generated from a multivariate normal distribution $N(0, \mathbb{V})$, where $\mathbb{V} = (V_{kj})$ had a first-order autoregressive covariance structure with $V_{kj} = \rho^{|k-j|}$. Second, a haplotype was generated by dichotomizing the latent vector s with some pre-specified minor allele frequencies (MAFs), each of which was randomly sampled from a uniform distribution between 0.1 and 0.3 for common variants (unless otherwise stated for rare variants). Third, the above two steps were repeated to gen-

erate two independent haplotypes and for subject i, the genotype value $G_i = (G_{i1}, \ldots, G_{ip})'$ was the sum of the two haplotypes. We set $\rho = 0$ to generate independent SNPs unless otherwise stated.

As in Lin et al. (2013), we generated a binary outcome by the following logistic regression model

$$\text{logit}[P(Y_i = 1|Z_i, E_i, G_i)] = \vartheta_0 + \vartheta_1 Z_{1i} + \vartheta_2 Z_{2i} + \vartheta_3 E_i + \vartheta_4' G_i + \beta' G_i \times E_i,$$

where $\vartheta_0 = \log(0.4/0.6)$, $\vartheta_1 = 0.05$, $\vartheta_2 = 0.057$, $\vartheta_3 = 0.64$, and

$$\vartheta_4 = (\underbrace{0.4, \ldots, 0.4}_{q_1}, \underbrace{-0.4, \ldots, -0.4}_{q_2}, \underbrace{0, \ldots, 0}_{p-q_1-q_2})'.$$

Z_1 was generated from a normal distribution while Z_2 was generated from a Bernoulli distribution. Environmental variable E was generated from a Bernoulli distribution, taking on 1 and -1 with an equal probability. $G_i \times E_i$ is the gene-environmental interaction for subject i. As in a case-control study, we sampled $n/2$ cases and $n/2$ controls in each data set. We were interested in testing H_0 to see whether there is any gene-environment interaction. Under H_A, the gene-environmental interaction effect patterns are generally complex and unknown. For example, for xeroderma pigmentosum, there is no main genetic effect, but both environmental (ultraviolet light) effect and gene-environmental interaction effect exist (Hunter, 2005). To consider various scenarios, we randomly chose $\lfloor ps \rfloor$ elements in β to be non-zero and their values were generated from a uniform distribution $U(-c, c)$ unless otherwise stated.

Simulation settings for high-dimensional linear models. We generated $\mathbb{X}_{n\times p}$ and $\mathbb{Z}_{n\times q}$ from a multivariate normal distribution; that is, we had independent draws $X_i \sim N(0, \mathbf{\Xi}_1)$ and $Z_i \sim N(0, \mathbf{\Xi}_2)$ for $i = 1, \ldots, n$, where $\mathbf{\Xi}_1$ and $\mathbf{\Xi}_2$ were block diagonal symmetric matrices. The response Y was generated from a high-dimensional linear model:

$$Y = \mathbb{Z}\vartheta + \mathbb{X}\beta + \epsilon,$$

where $\vartheta = (\vartheta_1, \ldots, \vartheta_q)'$, $\beta = (\beta_1, \ldots, \beta_p)'$, and each element of ϵ followed a standard normal distribution. We set $\vartheta_1 = \vartheta_2 = 0.4$ and other $\vartheta_j = 0$. We considered testing H_0 and H_A in (2).

Under H_A, $\lfloor ps \rfloor$ elements in β were set to be non-zero, where $s \in [0, 1]$ controlled the level of signal sparsity. The indices of non-zero elements of β were uniformly distributed, and their values were generated from a uniform distribution $U(-c, c)$ unless specified otherwise. We set $n = 200$, $q = 1000$, and $p = 1000$.

For each simulation setting, we generated 1,000 data sets to evaluate the empirical size and power at the significance level $\alpha = 0.05$. The candidate set of γ for the aiSPU was taken to be $\Gamma = \{1, \ldots, 6, \infty\}$ unless otherwise stated.

To evaluate the effect of penalization, we further presented the results of aiSPU with two different ways of estimating the nuisance parameter ϑ under H_0. First, we considered the oracle estimator, which is defined as the MLE with the knowledge/oracle about which covariates are non-informative (i.e. their effect size is 0) under H_0, denoted as aiSPU(Oracle).

Second, under the situation with $n > p$, we considered using the MLE to estimate ϑ, denoted as aiSPU(Full). Note that aiSPU(Full) equals to the aSPU (Wu et al., 2019).

For comparison, under $G \times E$ interaction settings, we applied GESAT (Lin et al., 2013) for common variants, and applied both iSKAT (Lin et al., 2016) and MiSTi (Su et al., 2017) for rare variants. To confirm that the theoretical null distribution of GESAT may not hold under a relatively high-dimensional situation, we calculated the p-value of GESAT by a simulation-based method, denoted as GESAT-sim. As a benchmark, we further considered the univariate minimum p-value (UminP) test, which first tests for SNP-environment interaction for each SNP, then takes their minimum p-value as the test statistic, and finally performs a corresponding Bonferroni adjustment. Under high-dimensional linear model settings, we conducted the three-step procedure with NST and ST statistics (Zhang and Cheng, 2017).

3.2. Results for $G \times E$ interactions

In many set-based $G \times E$ testing applications, the number of genetic variants p is relatively large but still smaller than the sample size n. Thus, we conducted two types of simulations: $n > p$ or $n < p$.

Simulations with $n > p$. First, we conducted simulations with $n = 2000$, $q_1 = 2$, $q_2 = 0$, and varying p to evaluate Type I error rates of different tests under different scenarios, ranging from low-dimensional to relatively high-dimensional. Note that the dimension of the nuisance parameter ϑ was $q = p + 4$, while that of the parameter β being tested was p. Table 1 shows the empirical Type I error rates, indicating that GESAT (Lin et al., 2013) yielded an inflated Type I error rate when p was large. Of note, even though by searching a much larger upper bound for the tuning parameter (say, $\sqrt{n}$ instead of the default $\sqrt{n/\log(n)}$) somewhat alleviated the problem, GESAT still yielded an inflated Type I error rate. For example, the Type I error rate of GESAT with tuning parameter searching up to $\sqrt{n}$ was 0.253 for the situation with $n = 2000$ and $p = 300$. In contrast, GESAT-sim maintained the correct Type I error rate, confirming that the theoretical null distribution of GESAT was not applicable when q was relatively large. As expected, aiSPU(Full) maintained the correct Type I error rate when q was relatively small and yielded an inflated Type I error rate when q was large, indicating penalized estimation of ϑ was necessary when q was relatively large. As expected, both aiSPU(Oracle) and aiSPU(TLP) yielded well-controlled Type I error rates for all the situations considered.

Next, we studied the effect of the number of non-zero nuisance parameters. Here we evaluated the performance of iSPU and aiSPU with some popular penalties, such as the Lasso and ridge. Table 2 shows the results of $n = 2000$, $p = 300$, and varying $q_1 = q_2$. When $q_1 = q_2$ was relatively large ($q_1 = q_2 = 20$ or $q_1 = q_2 = 30$), both the ridge and Lasso yielded slightly conservative Type I error rates and thus power loss (Figure 1). In contrast, aiSPU(TLP) provided results that were similar to those of aiSPU(Oracle). Again, GESAT yielded inflated Type I error rates because its theoretical null distribution was not applicable with relatively larger p and $p > n$. The results of $n = 2000$, $p = 200$, and varying $q_1 = q_2$ show similar conclusions (Table S1 in Appendix E).

To evaluate empirical power, we considered two cases: (a) under relatively low dimensional situations; (b) under relatively high-dimensional situations. Figure 1 shows the

Table 1: Empirical Type I error rates of various tests for $G \times E$ interaction in simulations with $n = 2000$, $q_1 = 2$, $q_2 = 0$, and varying p. n, p, q_1, and q_2 stand for the sample size, number of terms in $G \times E$ interaction, number of the positive genetic main effects, and number of the negative genetic main effects, respectively. * Inflated Type I error rates.

p	25	50	70	100	200	300	400	500
GESAT	0.061	0.055	0.090*	0.103*	0.277*	0.636*	0.944*	1.000*
GESAT-sim	0.050	0.048	0.062	0.050	0.051	0.044	0.051	0.047
aiSPU(Full)	0.071	0.057	0.080*	0.085*	0.199*	0.551*	0.944*	1.000*
aiSPU(Oracle)	0.067	0.049	0.064	0.052	0.052	0.046	0.057	0.047
aiSPU(TLP)	0.061	0.054	0.057	0.053	0.053	0.042	0.060	0.047

Table 2: Empirical Type I error rates of various tests for $G \times E$ interaction in simulations with $n = 2000$, $p = 300$ and varying $q_1 = q_2$. n, p, q_1, and q_2 stand for the sample size, number of terms in $G \times E$ interaction, number of the positive genetic main effects, and number of the negative genetic main effects, respectively. * Inflated Type I error rates; ** Conservative Type I error rates.

$q_1 = q_2$	2	5	7	10	20	30
GESAT	0.637*	0.636*	0.628*	0.641*	0.657*	0.633*
GESAT-sim	0.043	0.030	0.026	0.010**	0.004**	0.002**
aiSPU(Ridge)	0.058	0.046	0.045	0.027	0.023**	0.017**
aiSPU(Lasso)	0.048	0.039	0.035	0.028	0.023**	0.016**
aiSPU(Full)	0.584*	0.594*	0.598*	0.634*	0.690*	0.712*
aiSPU(Oracle)	0.054	0.057	0.054	0.056	0.062	0.055
aiSPU(TLP)	0.058	0.052	0.057	0.053	0.058	0.057

power of different methods under relatively high-dimensional situations with $n = 2000$, $p = 300$, and $q_1 = q_2 = 20$. Because both the Lasso and ridge yielded slightly conservative Type I error rates, aiSPU(Ridge) and aiSPU(Lasso) were less powerful than aiSPU(TLP). Perhaps because TLP better approximated the optimal L_0 constraint (Shen et al., 2012), aiSPU(TLP) achieved higher power than aiSPU(MCP) and aiSPU(SCAD). As a benchmark, UminP performed relatively well when the signal was sparse. Figure S1 shows that iSPU with different γ was more powerful under different sparsity levels. However, due to its adaptivity, aiSPU was the overall winner (Figure S1 in Appendix E). The results for correlated SNPs ($\rho = 0.3$) or $q_1 = q_2 = 50$ showed similar patterns as in Figure 1 and thus were relegated to the Appendix E (Figures S2 and S3). Under relatively low-dimensional situations with $n = 2000$ and $p = 25$ or 50 (Figures S4 and S5), GESAT yielded well-controlled Type I error rates and achieved very similar power as GESAT-sim and iSPU(2). As expected, GESAT achieved higher power than iSPU(∞) under dense signal situations, but lower power than iSPU(∞) under sparse signal situations. In comparison, aiSPU achieved robustly high power under various scenarios. For the situation with $n = 2000$ and $p = 75$, while GESAT (regardless of how larger a searching region for the tuning parameter) had a slightly inflated Type I error rate, the results showed similar patterns as before (Figure S6 in Appendix E).

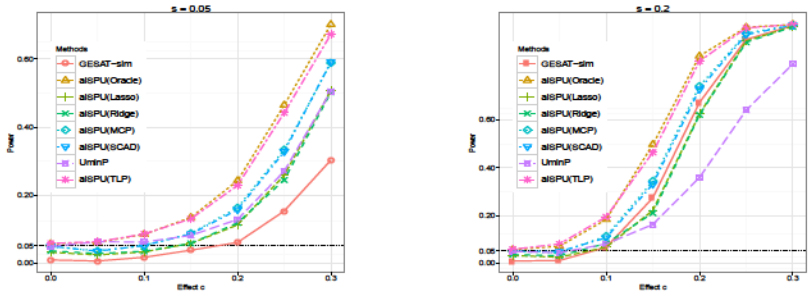

Figure 1: Power comparison for different methods in under $G \times E$ interaction simulations with $n = 2000$, $p = 300$ and $q_1 = q_2 = 20$. n, p, q_1, and q_2 stand for the sample size, number of terms in $G \times E$ interaction, number of the positive genetic main effects, and number of the negative genetic main effects, respectively. We varied the sparsity level s.

Next, similar to that in Su et al. (2017), we considered rare variants by generating SNPs with MAFs ranging from 0.005 to 0.05 while keeping the other simulation aspects unchanged. As expected, when p was relatively high, both iSKAT and MiSTi yielded inflated Type I error rates due to the theoretical null distribution is not applicable under relatively larger p and $p > n$ situations. In contrast, aiSPU(TLP) maintained the correct Type I error rate under relatively high-dimensional situation (Tables S2 and S3 in Appendix E). Figure S7 shows the power comparison under the different low dimensional situations. Again, even though different tests may be more powerful under certain situations, aiSPU achieved robust high power across all the situations considered.

Simulations with $n < p$. We conducted simulations with $n = 200$, $p = 1000$, $q_1 = 2$, $q_2 = 2$, and varying sparsity level s. Since GESAT yielded incorrect Type I error rates in high dimensional settings, the results of GESAT were not shown here.

First, we evaluated the performance of the asymptotic theory in Theorem 1 for finite samples. Table 3 shows the empirical Type I error rates and statistical power under $s = 0.005$. The iSPU and aiSPU yielded well-controlled Type I error rates. The results of the tests based on asymptotics were close to those based on the bootstrap, supporting Theorem 1. The results of other simulation settings ($s = 0.001$, $s = 0.01$,$s = 0.05$, $s = 0.2$, and informative variables in β were generated from a uniform distribution $U(0, c)$) showed similar patterns and were relegated into the Appendix E (Tables S4–S8). We further studied the situation when both main effects and interaction effects exist for the same set of SNPs and again showed similar patterns as expected (Table S9 in Appendix E).

Table 3: Empirical Type I errors and power (in percentage) of various tests under $G \times E$ interactions with $p = 1000$ and $n = 200$. Zero signal strength $c = 0$ represents Type I errors, while $c \neq 0$ represents powers. The sparsity level was $s = 0.005$, leading to 5 non-zero elements in β. The results outside and inside parentheses were calculated from parametric bootstrap- and asymptotics-based methods, respectively.

c	0	1	2	3	4	5
iSPU(1)	4.9 (4.8)	5.9 (5.6)	6.2 (6.1)	6.4 (6.6)	5.8 (6)	5.8 (5.7)
iSPU(2)	2.6 (5.2)	6.8 (11.8)	22.2 (28.5)	43.5 (47.5)	58.2 (61.3)	64.2 (67.8)
iSPU(3)	5.8 (5.6)	8.5 (8.1)	29.9 (28.7)	52.2 (51.4)	63.9 (62.1)	70.3 (69.2)
iSPU(4)	3 (3.9)	14.9 (17.1)	65.7 (67.1)	89.7 (90.4)	96 (96)	98.2 (98.4)
iSPU(5)	5.9 (5)	17 (15.6)	61.5 (60)	82.8 (81.3)	90.1 (88.9)	92.3 (92.5)
iSPU(6)	3.7 (3.2)	21.8 (19)	75.6 (74)	94.9 (93.7)	98.4 (97.9)	99.2 (99.2)
iSPU(∞)	8.5 (7.5)	26.8 (22.2)	85 (83.3)	97.6 (97.4)	99.6 (99.6)	100 (100)
aiSPU	5.8 (6.1)	20.7 (21.5)	79.4 (80.5)	95.4 (96.1)	98.8 (99.4)	99.8 (99.7)

Next, we compared statistical power. Figure 2 shows the empirical power for the tests under different sparsity levels s. When the signal was highly sparse, iSPU(∞) was more powerful than other tests ($s = 0.001$ and $s = 0.005$). As signal became relatively sparse ($s = 0.05$), iSPU(4) was the most powerful, closely followed by iSPU(6) and aiSPU, demonstrating the power gain by using some iSPU(γ) test with $2 < \gamma < \infty$ in a finite sample situation. When the signal became relatively dense with different association directions ($s = 0.2$), iSPU(2) was more powerful. For last sub-figure of Figure 2, we generated non-zero values of the parameter from a uniform distribution $U(0, c)$ instead, and iSPU(1) was the winner. All these simulation results confirmed the previous asymptotic power analysis. By combining information from different iSPU tests, aiSPU was an overall winner, either achieving the highest power or having power close to that of the winner in any setting. In comparison, UminP achieved relatively high power when the signal was sparse ($s = 0.001$, $s = 0.005$, and $s = 0.01$), but lost power substantially when the signal was dense ($s = 0.05$ and $s = 0.2$).

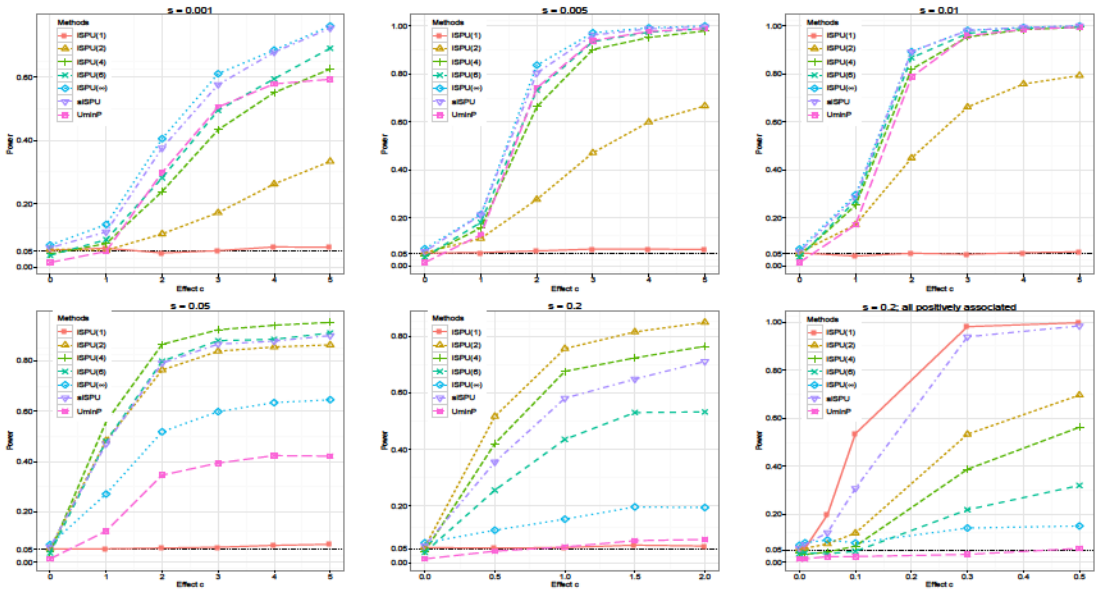

Figure 2: Power comparison for different methods under $G \times E$ interaction simulations with $n = 200$, $p = 1000$. We varied the sparsity level s. In last sub-figure, we generated informative variables in β from a uniform distribution $U(0, c)$.

Next, we briefly discussed the sensitivity of the aiSPU test to the choice of Γ set. Figure 3 shows the results of aiSPU with different Γ sets under different sparsity levels ($s = 0.01$, $s = 0.05$, and $s = 0.2$), indicating that the aiSPU test was robust to the choice of Γ. The results for other settings showed similar patterns and were relegated to the Appendix E (Figure S8).

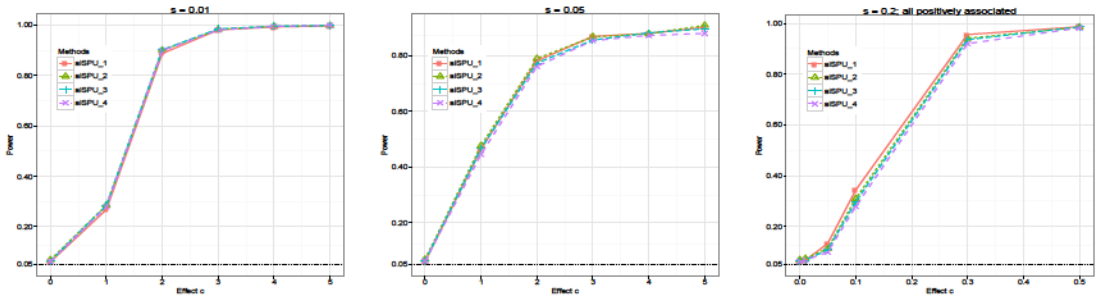

Figure 3: Empirical power of aiSPU with different Γ set under $G \times E$ interaction simulations with $n = 200$, $p = 1000$. aiSPU_1, aiSPU_2, aiSPU_3, aiSPU_4 represent aiSPU with $\Gamma_1 = \{1, 2, 3, 4; \infty\}$, $\Gamma_2 = \{1, 2, \ldots, 6, \infty\}$, $\Gamma_3 = \{1, \ldots, 8, \infty\}$, and $\Gamma_4 = \{1, 2, \ldots, 10, \infty\}$, respectively. We varied the sparsity level s. In last sub-figure, we generated none-zero elements of β from a uniform distribution $U(0, c)$.

Next, we briefly evaluated the robustness of the aiSPU test. Theorem 1 assumes that the effect of $\mathbb{Z}$ is sparse and strong. While this assumption is usually required by a penalized regression method, it might be violated in real applications. For example, under an omnigenic model (Liu et al., 2019), many variables in $\mathbb{Z}$ (i.e., SNPs) have weak effects, and only a few variables have strong effects. To evaluate the impact of the violation of the sparse effect assumption on $\mathbb{Z}$, we kept the simulation setting unchanged except that we randomly selected a pre-specified number of variables in $\mathbb{Z}$ and set non-zero small effect sizes for those selected variables. Figure 4 shows that aiSPU yielded well-controlled Type I error rates and achieved high power. Perhaps because the contribution of the small-effect variables in $\mathbb{Z}$ is relatively small to the estimation of $\hat{Y}$, the results of the tests based on asymptotics were close to those based on the bootstrap, indicating Theorem 1 is relatively robust to the violation of sparse effect assumption on $\mathbb{Z}$. We further varied the effect size for the randomly selected small effect variables in $\mathbb{Z}$ and obtained similar results (Figure S9 in Appendix E).

Next, we investigated whether aiSPU with other non-convex penalties such as SCAD (Fan and Li, 2001) and MCP (Zhang, 2010) would yield results similar to that with TLP. Perhaps because TLP enjoys the selection consistency and optimal parameter estimation under some mild assumptions (Shen et al., 2012), aiSPU(TLP) often achieved higher power than both aiSPU(MCP) and aiSPU(SCAD) (Figure S10). Interestingly, aiSPU(SCAD) yielded inflated Type I error rates under a linear model setting (Figure S11). In Summary, aiSPU(TLP) generally achieved higher power and controlled Type I error rates. Furthermore, we have provided some theoretical guarantee for aiSPU(TLP) and thus recommend using aiSPU with TLP as our default setting.

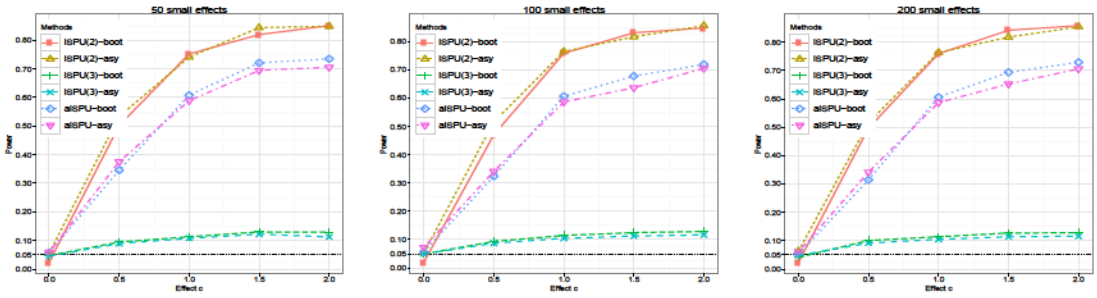

Figure 4: Empirical power of aiSPU under $G \times E$ interaction simulations with $n = 200$, $p = 1000$, and sparsity level $s = 0.2$. We randomly selected a pre-specified number of variables in $\mathbb{Z}$ and set the effect size followed a uniform distribution $U(-0.01, 0.01)$. -boot and -asy stand for the results based on bootstrap and asymptotics, respectively.

3.3. Results for linear models

First, the aiSPU test maintained correct Type I error rates, for which the asymptotics- and bootstrap-based methods gave similar results under different sparsity levels and association

directions (Tables S10–S13 in Appendix E). Similarly, both NST and ST yielded well-controlled Type I error rates (NST: 0.055 and ST: 0.061 at the significance level $\alpha = 0.05$).

Next, we assess statistical power. Figure 5 shows the empirical power for the tests under different sparsity levels s. Because the TLP estimator could consistently reconstruct the oracle estimator under mild assumptions (Shen et al., 2012), aiSPU(TLP) and aiSPU(Oracle) yielded similar results. Note that both NST and ST base their test statistics on a subsample, while aiSPU is on the whole sample; partly due to this difference in using the sample, aiSPU and iSPU(∞) were more powerful than both NST and ST even under a highly sparse alternative (i.e., with only one nonzero component in β; $s = 0.001$). Under other denser alternatives, aiSPU was way more powerful than both NST and ST. As in the simulations for $G \times E$ interaction, when the signal was relatively sparse ($s = 0.01$), iSPU(6) was the most powerful, highlighting the power gain by using some iSPU(γ) test with $2 < \gamma < \infty$. In contrast, SPU(2) was more powerful when the signal became dense ($s = 0.2$). Again, all these simulation results confirmed the previous asymptotic power analysis. By combining different iSPU tests, aiSPU maintained high power across a wide range of alternative scenarios.

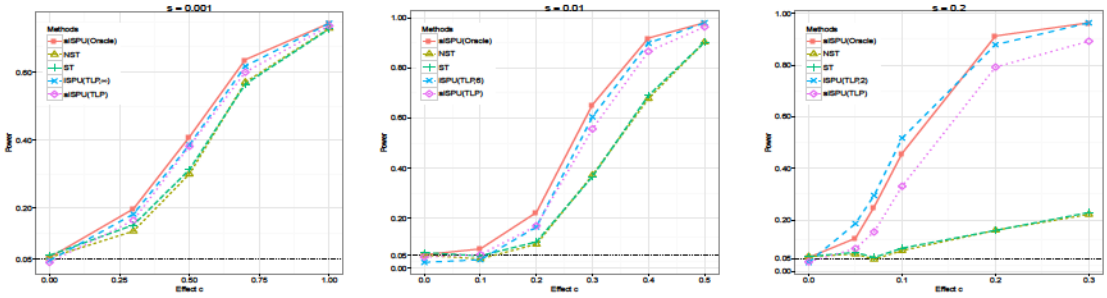

Figure 5: Power comparison for different tests under high-dimensional linear models simulations. We varied the sparsity level s.

In the end, we briefly compared the computational time among some competing methods, the parametric bootstrap-based aiSPU, and the asymptotics-based aiSPU (Figure S12 in Appendix E), showing that the asymptotic-based aiSPU was generally computationally more efficient. Of note, we implemented penalized regression with TLP in R, which is not computationally efficient in high-dimensional settings. We expect that the computational time for the asymptotics-based aiSPU can be further reduced once we implement aiSPU in C or other more efficient computer languages.

In summary, owing to its adaptivity, the power of aiSPU remained high, being either the winner or close to the winner in any setting. In particular, the aiSPU(TLP) test performed similarly to aiSPU(Oracle) and yielded well-controlled Type I error rates, presumably because the TLP estimator could consistently reconstruct the oracle estimator under mild conditions.

4. Real data analyses

Alzheimer's disease (AD) is the most common form of dementia, affecting millions of patients worldwide. The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a longitudinal, multisite observational study of elderly subjects with normal cognitive (healthy controls), mild cognitive impairment, or AD (Jack et al., 2008). The major goal of ADNI is to better understand the underlying mechanism of mild cognitive impairment (MCI) and AD (Jack et al., 2008). ADNI1 has recruited 819 elderly subjects to participate in the research. See `www.adni-info.org` for the latest information.

Several case-control studies suggest that AD is far more pronounced in females and gene-gender interaction may play roles in AD. Thus, we reanalyzed the ADNI1 data set to study whether the effect of genetic variants on AD risk is modified by gender.

Following set-ups in Altmann et al. (2014), we used the data of the Caucasian subjects in either the healthy control or MCI group, who had complete information on the environmental factor (gender) and covariates (age, years of education, and intracranial volume measured at baseline). For the outcome of interest, we set $Y_i = 1$ for any subject i in the MCI group, while setting $Y_i = 0$ for the other group. For the genotype data, we ran standard quality control steps to pre-process the data. In brief, we filtered out all SNPs with a genotyping rate < 0.95, those with a minor allele frequency < 0.05, and those failing to pass the Hardy-Weinberg equilibrium test (p-value $< 10^{-5}$). Further, we imputed the missing SNPs by a Michigan Imputation Server (Das et al., 2016) with the 1000 Genomes Phase 1 v3 European samples as the reference panel. We restricted our analysis to the HapMap3 SNP subset and pruned SNPs with a criterion of linkage disequilibrium $r^2 > 0.2$ using a sliding window of size 200 SNPs and a moving step of 20. According to the human genome reference hg19, we obtained the genomic coordinates of SNPs and genes, and assigned an SNP to a gene if it is located within 5,000 base pairs upstream or downstream of the gene's coding region. We extracted candidate pathways from the KEGG database (Kanehisa et al., 2009). As other pathway-based analyses (O'Dushlaine et al., 2015; Pan et al., 2015), we restricted our analyses to the pathways containing between 10 and 200 genes. In total, we analyzed 578 subjects and 96 KEGG pathways. To account for multiple testing, we applied the Bonferroni correction and used a slightly conservative cutoff $0.05/100 = 5 \times 10^{-4}$. Because other studies have reported an *APOE* gene and gender interaction on AD (Altmann et al., 2014), we tested the *APOE* and gender interaction as well. For testing main genetic effects, we applied aSPU (Pan et al., 2014) while adjusting for the same covariates as in testing $G \times E$ interactions.

Table 4 summarizes the results of our analysis. aiSPU identified one significant pathway "Fructose and mannose metabolism" (hsa00051, p-value $= 0.0003$) for $G \times E$ interaction, while GESAT failed to identify any significant pathways, showcasing possibly improved power of aiSPU over GESAT. Note that pathway "Fructose and mannose metabolism" contained 134 SNPs and thus, relative to the sample size, can be regarded as high-dimensional. The p-value of aiSPU was smaller than that of iSPU(1) and iSPU(2) but larger than iSPU(∞). Interestingly, aSPU failed to reject the null hypothesis of no main effects of the pathway (p-value $= 0.54$ by aSPU).

Next, we tested the *APOE* and gender interaction. Note that *APOE* contained 5 SNPs and can be viewed as a low dimensional situation. aSPU yielded a p-value of 0.007,

confirming the strong association of *APOE* on AD. Further, aiSPU yielded a p-value of 0.039 for the $G \times E$ interaction, suggesting a potential *APOE* and gender interaction. In contrast, GESAT yielded a p-value of 0.56, failing to detect any $G \times E$ interactions. Similarly, with a Bonferroni-adjusted p-value of 0.30, UminP also failed to detect $G \times E$ interactions. By analyzing a large, multisite, longitudinal data from National Alzheimer's Coordinating Center, Altmann et al. (2014) discovered *APOE*-gender interaction. They found that healthy female *APOE4* carriers had an almost 2-fold increased risk to develop MCI or AD when compared to female noncarriers (Altmann et al., 2014). By contrast, healthy male *APOE4* carriers had little increase in risk (Altmann et al., 2014). These findings support a possible interaction between *APOE* and gender on AD. In summary, our analyses have demonstrated that aiSPU is more powerful than GESAT in identifying gene-environment interactions when analyzing the ADNI1 data set.

Table 4: P-values from the association analysis of the ADNI1 data set to detect interactions between gender and genetic variants (in KEGG pathway hsa00051 or gene APOE).

	iSPU(γ)							aiSPU	GESAT
	$\gamma=1$	$\gamma=2$	$\gamma=3$	$\gamma=4$	$\gamma=5$	$\gamma=6$	$\gamma=\infty$		
hsa00051	0.017	0.017	0.014	0.010	0.006	0.003	0.0001	0.0003	0.016
APOE	0.022	0.032	0.042	0.059	0.068	0.079	0.112	0.039	0.56

5. Discussion

In this paper, we have proposed and studied an adaptive aiSPU test for high-dimensional parameters in GLMs in the presence of high-dimensional nuisance parameters. Our proposed aiSPU test takes advantage of both the TLP estimator (Shen et al., 2012) and data adaptive testing ideas (Pan et al., 2014), and thus enjoys several theoretical and practical benefits: first, the Type I error rate is well controlled; second, it maintains high statistical power under various scenarios, ranging from highly sparse to highly dense alternatives; third, it is computationally efficient as its p-values can be calculated via its asymptotic null distribution.

Several new methods (Ma et al., 2020; Shi et al., 2019; Sur and Candès, 2019; Fei and Li, 2019; Zhu et al., 2019) have recently been proposed for statistical inference with high-dimensional generalized linear models. However, they mainly focused on related but different questions with different approaches. Specifically, Ma et al. (2020) considered a global testing problem using a debiased Lasso based method with generalized low-dimensional projection. Sur and Candès (2019) quantified and corrected the bias of maximum likelihood estimators when the sample size and the dimensionality of parameters are in the same order. Fei and Li (2019) proposed a multi-sample splitting and averaging method to test a fixed subset of parameters. Shi et al. (2019) and Zhu et al. (2019) extended the score/Wald/likelihood ratio tests to (non-convex) penalized/constrained regression to test a subset of parameters of size much smaller than the sample size. In principle, due to its data-adaptive feature, aiSPU (with suitable modifications) may be a powerful tool to tackle

these related problems, though rigorous investigation is warranted. We leave it for future research.

We conclude with several potential extensions of our approach. First, as transcriptome-wide association studies (TWAS) (Gamazon et al., 2015; Gusev et al., 2016) that incorporate eQTL-derived weights into a weighted Sum test (Xu et al., 2017) to both improve statistical power and enhance biological interpretation, our proposed method can incorporate eQTL-derived weights into the test statistics of iSPU(γ) and aiSPU. Also, some other functional weights (He et al., 2017; Ma and Wei, 2019) can be equally applied. We expect that integrating functional genomic information will improve power and gain insights into the mechanisms of complex traits. Second, we mainly considered interactions between a genetic marker set and an environmental variable. We expect the same approach can be applied to other biological problems. For example, by replacing the environmental variable E with a treatment, we can test for interactions between a genetic marker set and the treatment, which is at the core of personalized medicine. More generally, our method can be potentially applied to other high-dimensional problems. For example, with some technical modifications, our method may be capable of simultaneous inference on submatrices of a high-dimensional precision matrix. The proposed method can also be extended to the asymptotically independent U-statistics framework as recently introduced in He et al. (2020). We leave these for future research.

Acknowledgments

We thank reviewers and the action editor for helpful comments. The authors thank Xianyang Zhang for sharing the R code implementing three-step procedures. This research was supported by the National Institutes of Health (NIH) grants R01GM113250, R01GM126002, R01HL105397, R01AG065636 and R01HL116720, by NSF grants DMS 1711226, DMS 1712717, DMS 1952539, SES 1659328 and SES 1846747, and by the Minnesota Supercomputing Institute. The investigators within the ADNI contributed to the design and implementation of ADNI and provided data, but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`.

Appendix A. Assumptions

We further decompose ϑ^0 into two parts: $\vartheta^0 \equiv (\vartheta_1^0, \ldots, \vartheta_q^0)' = (\vartheta_{A_0}^0, 0_{A_0^c})'$, where $A_0 \equiv \{j : \vartheta_j^0 \neq 0\}$ is the set of nonzero coefficients of ϑ^0 with size $|A_0| = q_0$, and $0_{A_0^c}$ is a vector of 0's. Define $\tilde{\mathbb{Z}}$ as the (nuisance) covariate matrix containing the variables in A_0, and the oracle estimate $\hat{\vartheta}^o$ as the maximum likelihood estimate (MLE) given that A_0 is known priori.

We need the following assumptions to establish the asymptotic null distribution.

C1. The eigenvalues of $\mathbf{\Sigma}$ are bounded, that is, $B^{-1} \leq \lambda_{\min}(\mathbf{\Sigma}), \lambda_{\max}(\mathbf{\Sigma}) \leq B$ for some finite constant B, where $\lambda_{\min}(\mathbf{\Sigma})$ and $\lambda_{\max}(\mathbf{\Sigma})$ denote the minimum and maximum eigenvalues of matrix $\mathbf{\Sigma}$, respectively. Moreover, the absolute value of any corresponding correlation element is strictly smaller than 1, i.e., $\max_{1\leq i\neq j\leq p} |\sigma_{ij}|/\sqrt{\sigma_{ii}\sigma_{jj}} < 1 - \xi$ for some constant $\xi > 0$.

C2. Under $H_0 : \beta = 0$, we have $E\left[S_{1j}^3\right] = 0$ for $1 \leq j \leq p$. There exist some constants ϱ and $K_0 > 0$ such that $E\big[\exp\big(\varrho S_{1j}^2/\sigma_{jj}\big)\big] \leq K_0$ for $1 \leq j \leq p$.

C3. $\tilde{\mathbb{Z}}$ is uniformly bounded. We further assume $E(X_{1j}|\tilde{\mathbb{Z}}) \neq 0$ only holds for $j \in P_0 \subset \{1, \ldots, p\}$ with the size of P_0, denoted by p_0, satisfying $p_0 = O(p^{\eta_1})$ for a small positive η_1.

C4. We assume $\frac{1}{p}\sum_{j_1,j_2} \big|E[S_{1j_1}S_{1j_2}]\big| = O(1)$ and $\frac{1}{p}\sum_{j_1\notin P_0, j_2\notin P_0} \big|E[X_{ij_1}X_{ij_2}|\tilde{\mathbb{Z}}]\big| = O(1)$.

C5. We assume $q \leq \exp\big(nC_{\min}(\vartheta^0)/d_0\big)$ and $pq_0^4/n^2 = o(1)$, where d_0 is some constant, $C_{\min}(\vartheta^0) \equiv \inf_{\{\vartheta^A=(\vartheta_A, 0_{A^c}: A\neq A_0, |A|\leq q_0\}} -\log(1 - h^2(\vartheta^A, \vartheta^0)/\max(|A_0 \setminus A|, 1)$, and $h(\cdot,\cdot)$ is the Hellinger distance. We further assume the model is sparse under the null, that is, $q_0 = O(n^{\eta_2})$ for a small positive η_2.

C6. There exist some positive constants K_1 and K_2 such that $K_1 < E[\epsilon_{0i}^2|\mathbb{Z} = z] < K_2$, where $\epsilon_{0i} = Y_i - \mu_{0i}$, $1 \leq i \leq n$. We further assume $\liminf_{n\to\infty} n^{-1}\lambda_{\min}(\tilde{\mathbb{Z}}'\mathbb{W}\tilde{\mathbb{Z}}) > 0$, where $\mathbb{W} = \mathrm{diag}\{E(\epsilon_{01}^2|\mathbb{Z}), \ldots, E(\epsilon_{0n}^2|\mathbb{Z})\}$.

C7. $-\log(1 - h^2(\vartheta, \vartheta^0)) \geq -d_1\log(1 - h^2(\vartheta_{\tau^+}, \vartheta^0)) - d_3 q\tau^{d_2}$ for some constants d_1, d_2, and d_3, where $d_1 - d_3 > 0$, $\vartheta_{\tau^+} = \big(\vartheta_1 I(|\vartheta_1| \geq \tau), \ldots, \vartheta_q I(|\vartheta_q| \geq \tau)\big)$ and $h(\vartheta, \vartheta^0)$ is the Hellinger distance between the two probability distributions specified by ϑ and ϑ^0. For some constant c_0 and any $\epsilon/2^4 < t < \epsilon \leq 1$, $H(t, \mathcal{B}_A) \leq c_0(\log q)^2|A|\log(2\epsilon/t)$, with $|A| \leq q_0$, where $H(\cdot, \mathcal{B}_A)$ is the bracketing Hellinger metric entropy of $\mathcal{B}_A$, $\mathcal{B}_A = \mathcal{F}_A \cap \{h(\vartheta, \vartheta^0) \leq 2\epsilon\}$ is a local parameter space, and $\mathcal{F}_A = \{g^{1/2}(\vartheta, y)$ is a collection of square root densities with $g(\vartheta, y)$ be a probability density for Y_1.

C8. Under $H_A : \beta \neq 0$, we have $E\big[(\tilde{S}_{ij})^3\big] = 0$ for $1 \leq j \leq p$.

C9. $\tilde{W} = \{\tilde{W}^{(j)} = (\tilde{S}_{ij}, i = 1, \ldots, n) : j \geq 1\}$ is α-mixing such that $\alpha_{\tilde{W}}(s) \leq M\delta^s$, where $\delta \in (0, 1)$ and M is some constant.

Remark S1 *Assumption C1 is commonly used in the high-dimensional setting (Cai et al., 2014; Wu et al., 2019), assuming that the underlying true covariance matrix* $\mathbf{\Sigma}$ *is non-singular. Assumption C2 assumes sub-Gaussian-type tails of* S_{1j}*, which is also common. Both assumptions C1 and C2 are only used to establish the weak convergence of* $L(\infty, \mu_0)$ *and not needed for* $L(\gamma, \mu_0)$ *with a finite* γ*.*

Assumption C3 assumes the underlying true model under H_0 *is sparse, which is often reasonable in real data applications and penalized regression framework. Note that we assume that each* X_j *is centered, which partially supports the assumption that* $E[X_{ij}|\tilde{\mathbb{Z}}] \neq 0$

only for $j \in P_0$ with the size of P_0 in a small order of p, i.e., $p_0 = O(p^{\eta_1})$. For example, in our motivating gene-environment interaction testing problems, $X_{ij} = G_{ij} \times E_i$ and $E[X_{ij}|\tilde{\mathbb{Z}}] \neq 0$ holds if and only if $E[G_{ij}|\tilde{\mathbb{Z}}] \neq 0$, where $\tilde{\mathbb{Z}}$ contains common covariates, environmental factors, and important SNPs selected by our penalized regression model. Of note, genome-wide association studies with around a hundred thousand subjects only identified from a few hundred to a few thousand significant SNPs for each of the traits, which were some tiny proportions of all the SNPs being tested (about 10 million) (Buniello et al., 2018). In other words, the majority of SNPs G_j are independent of common covariates. Furthermore, because linkage disequilibrium (LD) is often local, SNP G_j is only correlated with a small proportion of the SNPs being tested (see Figure S13 for an example). Then $E[X_{ij}|\tilde{\mathbb{Z}}] \neq 0$ only for $j \in P_0$ with the size of P_0, $p_0 = O(p^{\eta_1})$. One caveat is that even though C3 usually holds for genetic and genomic data, C3 may fail in other applications, perhaps leading to Theorem 1 invalid. We leave this interesting topic for future research.

Assumption C4 is a moment assumption and assumes a weak dependence structure. Intuitively speaking, many random vectors meet this moment assumption. For example, random vectors $\zeta = (\zeta_1, \zeta_2, \dots)'$, where ζ_i only correlates a finite number of ζ_j; then ζ satisfies moment condition. It also includes an α-mixing type weak dependence as a special case, which has been broadly used in time series and spatial statistics and adopted previously in high-dimensional testing problems (Xu et al., 2016; Wu et al., 2019). To account for the effects of nuisance parameters, we further assume conditionally moment assumption, which is a natural extension of the moment assumption.

Assumption C5 is a relatively strong assumption needed to prove Theorem 1. It imposes some restrictions on the growth rate of p such that $p = O(n^{2-\eta_3})$ for a small positive η_3. Zhang and Cheng (2017) assumed $(\log(pn))^7/n \leq N_1 n^{-N_2}$ for some positive constants N_1 and N_2 to establish the theory for a bias-correction based test statistic, which is weaker than C5. A stronger condition is needed here to establish the joint asymptotic distribution of $L(\gamma)$ with different γ's. Nevertheless, C5 allows $p/n \to \infty$. Assumption C5 imposes a weak restriction on q, allowing exponentially many nuisance parameters $q = \exp\left(nC_{\min}(\vartheta^0)/d_0\right)$. Shen et al. (2012) showed that this is a necessary condition for TLP to be selection consistent. C5 also assumes the sparsity on the ϑ^0, which is common adopted by penalized regression and by bias correction-based methods. For example, under the nearly optimal condition $q_0 = o(n/(\log p)^2)$, the debiased Lasso estimator follows a Gaussian distribution asymptotically (Javanmard and Montanari, 2018). Also, the sparsity assumption regarding q_0 may be relaxed. For example, the sparsity assumption is $q_0 = o(n/\log p)$ in a directed graphical model with TLP constraint (Li et al., 2019). More importantly, the sparsity assumption might not be essential for our proposed method. As shown in the simulation section (Figures 4 and S9), our proposed method still worked when the sparsity assumption was violated.

The first part of Assumption C6 is common in GLMs, for example, as adopted in Fan and Song (2010); Guo and Chen (2016). By Theorem 5.39 in Vershynin (2010), we have $\liminf_{n\to\infty} n^{-1}\lambda_{\min}(\tilde{\mathbb{Z}}'\mathbb{W}\tilde{\mathbb{Z}}) > 0$ with high probability. To simplify the technical details in the proof of the weak convergence result in Theorem 1, here we assume $\liminf_{n\to\infty} n^{-1}\lambda_{\min}(\tilde{\mathbb{Z}}'\mathbb{W}\tilde{\mathbb{Z}}) > 0$.

Assumption C7 is needed for feature selection consistency and optimal parameter estimation by TLP for GLMs (Shen et al., 2012). However, for linear and logistic regression,

Assumption C7 can be substituted by the following C7 for verification purpose (Shen et al., 2012).*

C7.* $\gamma_{\min}^2 \min_{A:|A|\le q_0, A\neq A_0} \lambda_{\min}\big(\mathbf{\Theta}_{A_0\setminus A} - \mathbf{\Theta}_{A_0\setminus A,A}\mathbf{\Theta}_A^{-1}\mathbf{\Theta}_{A,A_0\setminus A}\big) \ge d_0\frac{\log q}{n}$, *where* d_0 *is some constant independent of* (n, q, q_0), $\mathbf{\Theta}$ *is the covariance of* $\mathbb{Z}$ *with the* $jkth$ *element* $cov(Z_j, Z_k)$, $\mathbf{\Theta}_{A_0\setminus A}$ *is a submatrix of* $\mathbf{\Theta}$ *by keeping rows and columns corresponding to a subset* $A_0 \setminus A$ *of predictors, and* $\mathbf{\Theta}_{A_0\setminus A,A}$ *is a submatrix of* $\mathbf{\Theta}$ *by keeping rows corresponding to a subset* $A_0 \setminus A$ *of predictors and columns corresponding to a subset* A *of predictors,* $\gamma_{\min} = \gamma_{\min}(\vartheta^0) \equiv \min\{|\vartheta_k^0| : \vartheta_k^0 \neq 0\}$ *is the resolution level of the true regression coefficients, and* $\lambda_{\min}\big(\mathbf{\Theta}_{A_0\setminus A} - \mathbf{\Theta}_{A_0\setminus A,A}\mathbf{\Theta}_A^{-1}\mathbf{\Theta}_{A,A_0\setminus A}\big) \ge \min_{B\supset A_0:|B|\le 2q_0} \lambda_{\min}\big(\mathbf{\Theta}_B\big)$.

Of note, C7 involves Hellinger distance, which is hard to verify in practice. For verification purposes, we propose a stronger than needed assumption C7, which is sufficient for C7. C7* imposes a lower bound for coefficient strength (like a beta-min condition), which might be violated in practice. However, our proposed method might still work when assumption C7* (i.e., beta-min like assumption) is violated but C7 holds. In addition, we use this technical assumption to prove the TLP-based estimator achieves feature selection consistency, and thus the difference between* μ_0 *and* $\hat{\mu}_0$ *can be well controlled. In practice, as long as we have a good estimate* $\hat{\mu}_0$*, our proposed method works. For example, as shown in simulations, our proposed method aiSPU still worked when the coefficient for* $\mathbb{Z}$ *was non-sparse but with sparse and strong signals, which violated the assumption C7* but not necessarily C7. To further illustrate this, we provide an example. Suppose the coefficient* $\vartheta^0 = (1, \frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \ldots, \frac{1}{\sqrt{n}})'$, $\tau = \frac{1}{2\sqrt{2}}$, *and* $d_1 = 1$. *Then* $\log(1 - h^2(\vartheta, \vartheta^0))$ *equals to* $\log(1 - h^2(\vartheta_{\tau^+}, \vartheta^0))$*. This leads to the assumption C7 holds, coefficient for* $\mathbb{Z}$ *is non-sparse, and C7* may not hold. A similar example has been provided for TLP in the context of constrained maximum likelihood inference (Zhu et al., 2019). We leave the exciting topic on further relaxing Assumption C7 for future research.*

Appendix B. Calculating p-values

In this subsection, we describe how to calculate p-values by both parametric bootstrap and asymptotics-based methods in details.

Asymptotics-based method

First, we calculate p-values for iSPU separately. We apply the Theorem 1 to approximate $\psi(\gamma)$, $\omega(\gamma)$ and $\mathbf{R}(\Gamma') = (\rho_{st})$, respectively. Specifically, $\psi(1) = 0$, $\psi(\gamma) = \frac{\gamma!}{d!2^d} n^{-d} \sum_{j=1}^p \sigma_{jj}^d + o(pn^{-d})$ if $\gamma = 2d$ and $\psi(\gamma) = o(pn^{-(d+1)})$ if $\gamma = 2d+1$; $\omega^2(1) = \frac{1}{n}\sum_{1\le i,j\le p}\sigma_{ij} + o(pn^{-1})$ and

$$\omega^2(\gamma) = \psi(2\gamma) - \sum_{j=1}^p \{\psi^{(j)}(\gamma)\}^2 + \frac{1}{n^\gamma}\sum_{i\neq j}\sum_{\substack{2c_1+c_3=\gamma\\ 2c_2+c_3=\gamma\\ c_3>0}} \frac{(\gamma!)^2}{c_3!c_1!c_2!2^{c_1+c_2}}\sigma_{ii}^{c_1}\sigma_{jj}^{c_2}\sigma_{ij}^{c_3} + o(pn^{-\gamma})$$

if $\gamma \ge 2$;

$$\begin{aligned}&\operatorname{cov}[L(t,\mu_0),L(s,\mu_0)]\\&=\psi(t+s)-\sum_{j=1}^{p}\psi^{(j)}(t)\psi^{(j)}(s)+\frac{1}{n^c}\sum_{k\neq j}\sum_{\substack{2c_1+c_3=t\\2c_2+c_3=s\\c_3>0}}\frac{t!s!}{c_3!c_1!c_2!2^{c_1+c_2}}\sigma_{kk}^{c_1}\sigma_{jj}^{c_2}\sigma_{kj}^{c_3}+o(pn^{-(t+s)/2})\end{aligned}$$

if $s+t$ is even and $\operatorname{cov}[L(t,\mu_0),L(s,\mu_0)]=o(pn^{-(t+s)/2})$ if $s+t$ is odd; $\rho_{ss}=1$ for $s\in\Gamma'$ and $\rho_{st}=\operatorname{cov}\big[L(s,\mu_0),L(t,\mu_0)\big]/(\omega(s)\omega(t))$ for $s\neq t\in\Gamma'$. Then by Theorem 1, the p-values for individual iSPU(γ) can be calculated via either a normal or an extreme value distribution.

Remark S2 *In practice, $\mathbf{\Sigma}=(\sigma_{kj})$ is unknown and has to be estimated, which is a challenging problem under a high-dimensional setting. We discussed two situations separately: when $\mathbf{\Sigma}$ satisfies certain structures and when the structure is unknown.*

When the covariance matrix $\mathbf{\Sigma}$ satisfies certain structures, we can apply some existing methods, such as banding and thresholding techniques (Bickel and Levina, 2008; Cai and Liu, 2011). See Fan et al. (2016) for an excellent review. For example, we can apply the banding method of Bickel and Levina (2008) to estimate covariance matrix $\mathbf{\Sigma}$ if the following α-mixing assumption holds: $W=\{W^{(j)}=(S_{ij},i=1,\ldots,n):j\geq 1\}$ is α-mixing such that $\alpha_W(s)\leq M_1\delta_1^s$, where $\delta_1\in(0,1)$ and M_1 is some constant. Specifically, we calculate the sample covariance matrix $\mathbb{S}=(\mathfrak{s}_{kj})$ and then calculate the bandable covariance matrix as $\hat{\mathbf{\Sigma}}_{k_n}=\big(\mathfrak{s}_{kj}I(|k-j|\leq k_n)\big)$. An optimal bandwidth k_n has been selected by five-fold cross-validation. For a properly chosen $k_n=o(n^{1/2})$, the difference induced by estimating $\mathbf{\Sigma}$ is ignorable (Xu et al., 2016; Wu et al., 2019). We further define $\hat{\psi}(\gamma)$ and $\omega^2(\gamma)$ as the corresponding estimated $\psi(\gamma)$ and $\omega^2(\gamma)$ by replacing $\mathbf{\Sigma}$ with $\hat{\mathbf{\Sigma}}_{k_n}$. Under the mixing assumption, for any j,k, and $\epsilon>0$, there exists some constant C such that $\sigma_{kj}\leq C\delta^{|k-j|\epsilon/(2+\epsilon)}$, where $\delta\in(0,1)$ (Guyon, 1995). Then for $k_n\to\infty$ as $n\to\infty$, the summation of terms involving σ_{kj} with $|k-j|>k_n$ in $\omega^2(\gamma)$ is ignorable, i.e., $n^{-\gamma}\sum_{k\neq j;|k-j|>k_n}C\sigma_{kk}^{c_1}\sigma_{jj}^{c_2}\sigma_{kj}^{c_3}=o(pn^{-\gamma})$. Furthermore, there are $O(k_np)$ terms in $\omega^2(\gamma)$ involving σ_{kj} with $|k-j|\leq k_n$. By noting that $\mathfrak{s}_{kj}=\sigma_{kj}+O_p(n^{-1/2})$, we have $\omega^2(\gamma)-\omega^2(\gamma)=o_p(pn^{-\gamma})$ if $k_n=o(n^{1/2})$. Because $\omega^2(\gamma)\sim O(pn^{-\gamma})$, we have $\omega^2(\gamma)=(1+o(1))\omega^2(\gamma)$. Similarly, we can derive $\hat{\psi}(\gamma)=(1+o(1))\psi(\gamma)$. Under our motivating gene-environmental interaction problems, the genetic variants are weakly dependent, and the dependent structure is often local. In other words, the α-mixing assumption holds and the banding method of Bickel and Levina (2008) works.

On the other hand, when the structure of the covariance matrix $\mathbf{\Sigma}$ is unknown, applying the banding method of Bickel and Levina (2008) might be problematic. As an alternative, we propose a parametric bootstrap-based method to estimate $\psi(\gamma)$, $\omega^2(\gamma)$ and $\mathbf{R}(\Gamma')=(\rho_{st})$, which circumvents the challenging problem of estimating the covariance matrix $\mathbf{\Sigma}$. Specifically, we first fit a penalized regression model under H_0 to obtain $\hat{\mu}_{0i}$ and then simulate a new set of responses $Y_i^{(b)}$ based on $\hat{\mu}_{0i}$ for $b=1,2,\ldots,B$. Second, we refit the model with the same tuning parameters and calculate the corresponding score vector $U^{(b)}$ and null statistics $L(\gamma)^{(b)}=\sum_{j=1}^{p}\big(\frac{1}{n}\sum_{i=1}^{n}U_{ij}^{(b)}\big)^{\gamma}$. In practice, we only need to repeat the above procedures for a relatively small B (say, 100) times. Then we estimate $\psi(\gamma)$, $\omega^2(\gamma)$ and $\mathbf{R}(\Gamma')=(\rho_{st})$ by $\hat{\psi}(\gamma)=\sum_{b=1}^{B}L(\gamma)^{(b)}/B$, $\hat{\omega}^2(\gamma)=\sum_{b=1}^{B}(L(\gamma)^{(b)}-\hat{\psi}(\gamma))^2/(B-1)$, and $\hat{R}(\Gamma')=cor(L(\Gamma')^{(b)})$, where cor is the sample correlation estimated by cor() function in R.

Second, we calculate the p-value for aiSPU. Because $\text{cov}[L(t,\mu_0), L(s,\mu_0)] = o(pn^{-(t+s)/2})$ if $s+t$ is odd (Theorem 1 part one), iSPU with even γ and odd γ are asymptotically uncorrelated. By Theorem 1 part one, iSPU with finite γs converge jointly and weakly to a multivariate normal distribution as $n, p \to \infty$, leading to iSPU with even γ and odd γ are asymptotically independent. Then by Theorem 1 part three, iSPU with even γ, odd γ, $\gamma = \infty$ are asymptotically independent to each other. Because for a finite γ, $L(\gamma) - \psi(\gamma)/\omega(\gamma)$ follows a standard normal distribution, taking the minimum p-values as test statistics equals to taking the maximum of $|L(\gamma) - \psi(\gamma)|/\omega(\gamma)$ as test statistics. Then we can analytically calculate the p-value for aiSPU by the following two steps. First, define $t_O = \max_{\text{odd } \gamma \in \Gamma} \big|\big(L(\gamma) - \psi(\gamma)\big)\big|/\omega(\gamma)$ and $t_E = \max_{\text{even } \gamma \in \Gamma} \big(L(\gamma) - \psi(\gamma)\big)/\omega(\gamma)$ as the observed test statistics from the data and calculate the p-values for t_O and t_E as $p_O = Pr[\max_{\text{odd } \gamma \in \Gamma} |\big(L(\gamma) - \psi(\gamma)\big)/\omega(\gamma)| > t_O]$ and $p_E = Pr[\max_{\text{even } \gamma \in \Gamma} \big(L(\gamma) - \psi(\gamma)\big)/\omega(\gamma) > t_E]$. We use *pmvnorm()* in R package *mvrnorm* to calculate the normal tail probabilities of p_O and p_E. Further, let p_∞ denote the p-value of iSPU(∞), which can be calculated by a extreme value distribution (Theorem 1 part 2). Second, take the minimum p-value from the above three categories, that is, $p_{\min} = \min\{p_O, p_E, p_\infty\}$. By the asymptotic independence, the asymptotic p-value for the aiSPU test is $p_{\text{aiSPU}} = 1 - (1 - p_{\min})^3$.

Parametric bootstrap-based method

We can calculate p-values by parametric bootstrap as follows: first, we fit a penalized regression model under H_0 to obtain $\hat{\mu}_{0i}$ and then simulate a new set of responses $Y_i^{(b)}$ based on $\hat{\mu}_{0i}$ for $b = 1, 2, \ldots, B$; second, we refit the model with the same tuning parameters and calculate the corresponding score vector $U^{(b)}$ and null statistics $L(\gamma)^{(b)} = \sum_{j=1}^{p} \big(\frac{1}{n}\sum_{i=1}^{n} U_{ij}^{(b)}\big)^\gamma$; third, the p-value of iSPU(γ) is $P_{L(\gamma)} = [1 + \sum_{b=1}^{B} I(|L(\gamma)^{(b)}| \geq |L(\gamma)|)]/(B+1)$, and the p-value for aiSPU, $P_{\text{aiSPU}} = [1 + \sum_{b=1}^{B} I(T_{\text{aiSPU}}^{(b)} \leq T_{\text{aiSPU}})]/(B+1)$, with $T_{\text{aiSPU}}^{(b)} = \min_{\gamma \in \Gamma} P_{L(\gamma)}^{(b)}$ and $P_{L(\gamma)}^{(b_1)} = [\sum_{b \neq b_1} I(|L(\gamma)^{(b)}| \geq |L(\gamma)^{(b_1)}|)]/B$.

In practice, selecting a good B is important for saving computational sources. Here, we start with a smaller B, say $B = 1000$ to scan all the tests and then repeatedly increase B for the tests that that pass the following criterion: p-value $< 5/B$ in the previous step (Pan et al., 2014). Of note, the accuracy is bounded by the number of bootstraps B and calculating a very small p-value requires a very large B. This is different from asymptotics-based method, in which we only use a relatively small number of bootstraps (say, $B = 100$) to estimate mean, variance, and covariance matrix of iSPU and calculate the p-values by Theorem 1.

Appendix C. Proof of Theorem 1

We prove each part in Theorem 1 separately.

(i) Similar to the proof of Theorem 1 in Wu et al. (2019), we can show that if the conditional mean of Y, μ_0, is known, Theorem 1 holds. Specifically, by assumption C4, the order of double summation (across j_1 and j_2) of terms involving $\sigma_{j_1 j_2}$ is $O(p)$. Then by similar techniques used in Wu et al. (2019), we can calculate $\psi(\gamma)$, $\omega(\gamma)$, and $\mathbf{R}(\Gamma')$ as shown in Appendix B. We can further use Bernstein's block idea (Ibragimov and Linnik, 1971) to prove iSPU with finite γs follows a multivariate normal distribution asymptotically. Of note,

in our previous work (Wu et al., 2019), we derive a similar Theorem under the α-mixing assumption, which is a special case of Theorem 1.

Then for any fixed and finite γ, we prove Theorem 1 holds by showing that the difference between $L(\gamma, \mu_0)$ and $L(\gamma, \hat{\mu}_0)$ with the TLP estimates is negligible.

Note that the Hellinger distance for linear regression is

$$h^2(\vartheta, \vartheta^0) = 1 - E\Big[\exp\big(-\frac{1}{8}||\mathbb{Z}\vartheta - \mathbb{Z}\vartheta^0||^2\big)\Big]$$

and for logistic regression is

$$h^2(\vartheta, \vartheta^0) = \frac{1}{2}E\Big[\nu^{1/2}\big((\vartheta^0)'\mathbb{Z}\big) - \nu^{1/2}(\vartheta'\mathbb{Z}) + \big(1-\nu\big((\vartheta^0)'\mathbb{Z}\big)\big)^{1/2} - \big(1-\nu\big(\vartheta'\mathbb{Z}\big)\big)^{1/2}\Big],$$

where $\nu(s) = (1+\exp(s))^{-1}$. We decompose $A \equiv \{j : 1 \le j \le q\}$ into two parts: $A = A^{\tau^+} \cup A^{\tau^-}$, where $A^{\tau^+} \equiv \{j : |\vartheta_j| \ge \tau\}$ and $A^{\tau^-} \equiv \{j : |\vartheta_j| < \tau\}$. Further note that $\big|\frac{\partial h^2(\vartheta,\vartheta^0)}{\partial \vartheta_j}\big| \le 1/2E[|Z_j|]$ for $1 \le j \le q$ and $\vartheta \in \mathcal{R}^q$. Then

$$\big|h^2(\vartheta,\vartheta^0) - h^2(\vartheta_{\tau^+},\vartheta^0)\big| \le \tau\Big|\sum_{j\in A^{\tau^-}} \frac{\partial h^2(\vartheta,\vartheta^0)}{\partial\vartheta_j}\Big| \le 2\tau \sum_{j\in A^{\tau^-}} E[|Z_j|] \le 2\tau q \max_j \mathbf{\Sigma}_{jj},$$

where $\vartheta_{\tau^+} = (\vartheta_1 I(|\vartheta_1| \ge \tau), \ldots, \vartheta_q I(|\vartheta_q| \ge \tau))$. Then by assumption C7*, $-\log(1-x) > x$ for any $0 < x < 1$, and the derivative of $1-\exp(-1/8x^2)$ and $(1+\exp(x))^{-1/2}$ are bounded away from zero, the assumption C7 can be validated.

By assumption C5 and C7, through tuning, Theorem 2 in Shen et al. (2012) yields $P(\hat{\vartheta} \ne \hat{\vartheta}^o) \le \exp(-cn + 2\log(q+1) + 3)$, where c is some positive constant. Then we can apply Theorem 2 in Shen et al. (2012) and get the feature selection consistency for $\hat{\vartheta}$, that is, $E[h^2(\hat{\vartheta},\vartheta^0)] = E[h^2(\hat{\vartheta}^o,\vartheta^0)] = O(q_0/n) \to 0$ as $n \to \infty$. Then by the consistency property of MLE $||\hat{\vartheta}^o - \vartheta^0|| = O_p(q_0 n^{-1/2})$, we have $||\hat{\vartheta} - \vartheta^0|| = O_p(q_0 n^{-1/2})$.

Using Taylor expansion and the approach in Le Cessie and Van Houwelingen (1991), we have

$$\hat{\mathbb{D}} = \big(\mathbb{I}_n - \mathbb{W}\tilde{\mathbb{Z}}\{\mathbb{I}(\vartheta)\}^{-1}\tilde{\mathbb{Z}}'\big)\mathbb{D} + o_p(n^{-1/2}),$$

where $\hat{\mathbb{D}} = (Y - \hat{\mu}_0) = \{Y_1 - \hat{\mu}_{01}, \ldots, Y_n - \hat{\mu}_{0n}\}'$, $\mathbb{D} = (Y - \mu_0) = \{Y_1 - \mu_{01}, \ldots, Y_n - \mu_{0n}\}'$, $\mathbb{I}_n$ is the $n \times n$ identity matrix, $\mathbb{W}$ is a diagonal matrix, which is defined as $\mathbb{W} = \mathrm{diag}\{E(\epsilon_{01}^2|\mathbb{Z}), \ldots, E(\epsilon_{0n}^2|\mathbb{Z})\}$, $\tilde{\mathbb{Z}}$ contains the variables corresponding to $A_0 = \{j : \vartheta_j^0 \ne 0\}$, and $\mathbb{I}(\vartheta)$ is a $q_0 \times q_0$ matrix given by $\mathbb{I}(\vartheta) = \tilde{\mathbb{Z}}'\mathbb{W}\tilde{\mathbb{Z}}$. Since the smaller order term $o_p(n^{-1/2})$ can be ignored, we focus on the leading term in the subsequent analysis. For notation simplicity, further define $\mathbb{B} = \mathbb{W}\tilde{\mathbb{Z}}\{\mathbb{I}(\vartheta)\}^{-1}\tilde{\mathbb{Z}}' = (b_{ij})_{n\times n}$. By Cauchy-Schwarz inequality,

$$b_{ij} = W_{ii}\tilde{Z}_i\{\mathbb{I}(\vartheta)\}^{-1}\tilde{Z}_j \le W_{ii}(\tilde{Z}_i\{\mathbb{I}(\vartheta)\}^{-1}\tilde{Z}_i)^{1/2}(\tilde{Z}_j\{\mathbb{I}(\vartheta)\}^{-1}\tilde{Z}_j)^{1/2}.$$

By assumption C6, $\mathbb{W}$ is uniformly bounded and $\liminf n^{-1}\lambda_{\min}(\mathbb{I}(\vartheta)) > 0$. Then by assumption C3, $\tilde{\mathbb{Z}}$ is uniformly bounded, we have $\tilde{Z}_i\{\mathbb{I}(\vartheta)\}^{-1}\tilde{Z}_i \le O(q_0) \times \lambda_{\min}(\mathbb{I}(\vartheta))^{-1} = O(q_0 n^{-1})$. This leads to $b_{ij} = O(q_0 n^{-1})$ uniformly over i, j. By linear algebra, we have $\mu_{0i} - \hat{\mu}_{0i} = \sum_{l=1}^n b_{il}\epsilon_{0l}$ for $1 \le i \le n$, where $b_{il} = O(q_0 n^{-1})$. To prove the difference between $L(\gamma, \mu_0)$ and $L(\gamma, \hat{\mu}_0)$ can be ignored, we discuss two cases: $\gamma = 1$ and $1 < \gamma < \infty$ separately. To simplify the notation, we denote all the constants by C which may vary from place to place.

For $\gamma = 1$: we decompose the statistic $L(1, \hat{\mu}_0)$ as

$$L(1,\hat{\mu}_0) = \frac{1}{n}\sum_{j=1}^{p}\sum_{i=1}^{n}(Y_i - \hat{\mu}_{0i})X_{ij} = \sum_{j=1}^{p}\sum_{i=1}^{n}\frac{1}{n}S_{ij} + \sum_{j=1}^{p}\sum_{i=1}^{n}\frac{(\mu_{0i} - \hat{\mu}_{0i})X_{ij}}{n} = T_{10} + T_{11}.$$

Under the null hypothesis and proposed assumptions, Theorem 1 in Wu et al. (2019) leads to

$$T_{10}/\omega(1) \xrightarrow{d} N(0,1) \qquad \text{as } n \to \infty \text{ and } p \to \infty.$$

For T_{11}, since $\mu_{0i} - \hat{\mu}_{0i} = \sum_{l=1}^{n} b_{il}\epsilon_{0l}$, we have

$$\begin{aligned}
E[(T_{11})^2] &= \frac{1}{n^2}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\Big[(\mu_{0i_1} - \hat{\mu}_{0i_1})X_{i_1j_1}(\mu_{0i_2} - \hat{\mu}_{0i_2})X_{i_2j_2}\Big] \\
&= \frac{1}{n^2}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\Big[X_{i_1j_1}X_{i_2j_2}\sum_{l=1}^{n}\epsilon_{0l}b_{i_1l}\sum_{l=1}^{n}\epsilon_{0l}b_{i_2l}\Big] \\
&= \frac{1}{n^2}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\Big[X_{i_1j_1}X_{i_2j_2}\big(\epsilon_{0i_1}b_{i_1i_1} + \epsilon_{0i_2}b_{i_1i_2} + \sum_{l\neq i_1,i_2}\epsilon_{0l}b_{i_1l}\big) \\
&\qquad\qquad \times \big(\epsilon_{0i_1}b_{i_2i_1} + \epsilon_{0i_2}b_{i_2i_2} + \sum_{l\neq i_1,i_2}\epsilon_{0l}b_{i_2l}\big)\Big].
\end{aligned}$$

Since i_1 and i_2 are symmetrical, we have

$$\begin{aligned}
&E[(T_{11})^2] \\
=&\frac{C}{n^2}\sum_{j_1,j_2,i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}\big] + \frac{C}{n^2}\sum_{j_1,j_2,i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_1}\epsilon_{0i_2}b_{i_2i_2}\big] \\
&+ \frac{C}{n^2}\sum_{j_1,j_2,i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_1}\sum_{l\neq i_1,i_2}\epsilon_{0l}b_{i_2l}\big] + \frac{C}{n^2}\sum_{j_1,j_2,i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}\sum_{l\neq i_1,i_2}\epsilon_{0l}^2 b_{i_1l}b_{i_2l}\big] \\
=&E[T_{111}] + E[T_{112}] + E[T_{113}] + E[T_{114}].
\end{aligned}$$

We discuss the order of each term and show that $|T_{11}| = o_p(\sqrt{p}n^{-1/2})$ and thus can be ignored. By assumption C3, $|E[X_{ij}|\tilde{\mathbb{Z}}]| \neq 0$ only holds for $j \in P_0$, then

$$\begin{aligned}
E[T_{111}|\tilde{\mathbb{Z}}] &= \frac{C}{n^2}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\tilde{\mathbb{Z}}\big] \\
&= \frac{C}{n^2}\sum_{j_1\notin P_0}\sum_{j_2\notin P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\tilde{\mathbb{Z}}\big] \\
&\quad + \frac{C}{n^2}\sum_{j_1\in P_0}\sum_{j_2\notin P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\tilde{\mathbb{Z}}\big] \\
&\quad + \frac{C}{n^2}\sum_{j_1\notin P_0}\sum_{j_2\in P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\tilde{\mathbb{Z}}\big] \\
&\quad + \frac{C}{n^2}\sum_{j_1\in P_0}\sum_{j_2\in P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\tilde{\mathbb{Z}}\big].
\end{aligned}$$

For the first term, by Assumptions C3 and C4, we have

$$\begin{aligned}
&\frac{C}{n^2}\sum_{j_1\notin P_0}\sum_{j_2\notin P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\tilde{\mathbb{Z}}\big] \\
=&O(n^{-2})\sum_{j_1\notin P_0}\sum_{j_2\notin P_0}\sum_{i_1=1}^{n} E\big[X_{i_1j_1}X_{i_1j_2}\epsilon_{0i_1}^2|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-2}) \\
=&O(n^{-2})\times O(np^2)\times O(q_0^2n^{-2}) \\
=&O(pn^{-1})\times O(pq_0^2n^{-2}) = o(pn^{-1}).
\end{aligned}$$

Of note, because b_{ij} is a function of $\tilde{\mathbb{Z}}$, it can be taken out of the expectation when conditional on $\tilde{\mathbb{Z}}$. For the second term, we have

$$\begin{aligned}
&\frac{C}{n^2}\sum_{j_1\in P_0}\sum_{j_2\notin P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\tilde{\mathbb{Z}}\big] \\
=&O(n^{-2})\sum_{j_1\in P_0}\sum_{j_2\notin P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-2}) \\
=&O(n^{-2})\times O(pp_0n^2)\times O(q_0^2n^{-2}) = O(pp_0q_0^2n^{-2}) = o(pn^{-1}).
\end{aligned}$$

By noting that $p_0 = O(p^{\eta_1})$ for a small positive η_1 and assumption C5 $pq_0^4/n^2 = o(1)$, we can derive the last equation. Similar to the derivation of the second term, for the third term, we have

$$\frac{C}{n^2}\sum_{j_1\notin P_0}\sum_{j_2\in P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\tilde{\mathbb{Z}}\big] = o(pn^{-1}).$$

For the last term, we have

$$\begin{aligned}
&\frac{C}{n^2}\sum_{j_1\in P_0}\sum_{j_2\in P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2 b_{i_1i_1}b_{i_2i_1}|\tilde{\mathbb{Z}}\big]\\
=&O(n^{-2})\sum_{j_1\in P_0}\sum_{j_2\in P_0}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-2})\\
=&O(n^{-2})\times O(p_0^2n^2)\times O(q_0^2n^{-2})=O(n^{-2}q_0^2p_0^2)=o(pn^{-1}).
\end{aligned}$$

By combining the above derivations, we have $E[T_{111}|\tilde{\mathbb{Z}}]=o(pn^{-1})$. Importantly,

$$\sum_{j_1,j_2}\sum_{i_1,i_2}E[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}^2|\tilde{\mathbb{Z}}]=o(pq_0^2n^2).$$

Next, we discuss the order of $E[T_{112}|\tilde{\mathbb{Z}}]$. By noting that $E[X_{ij}\epsilon_{0i}|\tilde{\mathbb{Z}}]=0$, we have

$$\begin{aligned}
E[T_{112}|\tilde{\mathbb{Z}}]&=\frac{C}{n^2}\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_1}\epsilon_{0i_2}b_{i_2i_2}|\tilde{\mathbb{Z}}\big]\\
&=O(n^{-2})\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{i_1=1}^{n}E\big[X_{i_1j_1}\epsilon_{0i_1}^2X_{i_1j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-2})\\
&=O(n^{-2})\times O(p^2n)\times O(q_0^2n^{-2})=O(pn^{-1}pq_0^2n^{-2})=o(pn^{-1}).
\end{aligned}$$

Next, we discuss the order of $E[T_{113}|\tilde{\mathbb{Z}}]$:

$$\begin{aligned}
E[T_{113}|\tilde{\mathbb{Z}}]=&\frac{C}{n^2}\sum_{j_1,j_2,i_1,i_2}E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_1}\sum_{l\neq i_1,i_2}\epsilon_{0l}b_{i_2l}|\tilde{\mathbb{Z}}\Big]\\
=&O(n^{-2})\sum_{j_1,j_2,i_1,i_2}E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_1}|\tilde{\mathbb{Z}}\Big]E\Big[\sum_{l\neq i_1,i_2}\epsilon_{0l}b_{i_2l}|\tilde{\mathbb{Z}}\Big]\\
=&O(n^{-2})\sum_{j_1,j_2,i_1,i_2}E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_1}|\tilde{\mathbb{Z}}\Big]\times 0=0.
\end{aligned}$$

Next, we discuss the order of $E[T_{114}|\tilde{\mathbb{Z}}]$. Similarly, we decompose $E[T_{114}|\tilde{\mathbb{Z}}]$ into three parts and bound each part separately.

$$\begin{aligned}
E[T_{114}|\tilde{\mathbb{Z}}] &= \frac{C}{n^2}\sum_{j_1,j_2,i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}\sum_{l\neq i_1,i_2}\epsilon_{0l}^2 b_{i_1l}b_{i_2l}|\tilde{\mathbb{Z}}\big] \\
&= O(n^{-2})\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}\big]E\big[\sum_{l\neq i_1,i_2}\epsilon_{0l}^2 b_{i_1l}b_{i_2l}|\tilde{\mathbb{Z}}\big] \\
&= O(n^{-2})\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-1}) \\
&= O(n^{-2})\sum_{j_1\notin P_0,j_2\notin P_0}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-1}) \\
&\quad + O(n^{-2})\sum_{j_1\notin P_0,j_2\in P_0}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-1}) \\
&\quad + O(n^{-2})\sum_{j_1\in P_0,j_2\in P_0}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-1}).
\end{aligned}$$

By assumption C3, $E(X_{1j}|\tilde{\mathbb{Z}}) = 0$ for $j \notin P_0$. Then by assumption C4, for the first part, we have

$$\begin{aligned}
&O(n^{-2})\sum_{j_1\notin P_0,j_2\notin P_0}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-1}) \\
=&O(q_0^2n^{-3})\sum_{i_1=1}^{n}\sum_{j_1\notin P_0,j_2\notin P_0} E\big[X_{i_1j_1}X_{i_1j_2}|\tilde{\mathbb{Z}}\big] \\
=&O(q_0^2n^{-3})\times O(pn) \\
=&O(pn^{-1}n^{-1}q_0^2) = o(pn^{-1}).
\end{aligned}$$

Similarly, we have the following for the second part:

$$\begin{aligned}
&O(n^{-2})\sum_{j_1\notin P_0,j_2\in P_0}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-1}) \\
=&O(q_0^2n^{-3})\sum_{j_1\notin P_0,j_2\in P_0}\sum_{i_1=1}^{n} E\big[X_{i_1j_1}X_{i_1j_2}|\tilde{\mathbb{Z}}\big] \\
=&O(q_0^2n^{-3})\times O(p_0pn) \\
=&O(pn^{-1}p_0q_0^2n^{-1}) = o(pn^{-1}).
\end{aligned}$$

Note that by assumptions C3 and C5, $p_0 = O(p^{\eta_1})$ for a small positive η_1 and $q_0 = O(n^{\eta_2})$ for a small positive η_2. Then $O(p_0q_0^2n^{-1}) = O(p^{\eta_1}n^{1-2\eta_2}) = o(1)$ and the above last equation holds.

For the last part, we have

$$\begin{aligned}
&O(n^{-2})\sum_{j_1\in P_0,j_2\in P_0}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-1}) \\
=&O(n^{-2})\times O(p_0^2n^2)\times O(q_0^2n^{-1}) = O(q_0^2p_0^2n^{-1}) = o(pn^{-1}).
\end{aligned}$$

By Combining the above equations, we have $E[T_{114}|\tilde{\mathbb{Z}}] = o(pn^{-1})$. Importantly, we have

$$\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}\big] = o(pn^2q_0^{-2}).$$

In summary, we have $E[(T_{11})^2] = E\big[E[T_{111}|\tilde{\mathbb{Z}}] + E[T_{112}|\tilde{\mathbb{Z}}] + E[T_{113}|\tilde{\mathbb{Z}}] + E[T_{113}|\tilde{\mathbb{Z}}]\big] = o(pn^{-1})$, leading to $|T_{11}| = o_p(n^{-1/2}\sqrt{p})$. Thus T_{11} can be ignored and this completes the proof for $\gamma = 1$.

For $1 < \gamma < \infty$: we decompose the statistic $L(\gamma, \hat{\mu}_0)$ as

$$\begin{aligned}
L(\gamma,\hat{\mu}_0) &= \sum_{j=1}^{p}\Big(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\mu}_{0i})X_{ij}\Big)^{\gamma} \\
&= \sum_{j=1}^{p}\Big(\frac{1}{n}\sum_{i=1}^{n}\big((Y_i-\mu_{0i}) + (\mu_{0i}-\hat{\mu}_{0i})\big)X_{ij}\Big)^{\gamma} \\
&= \sum_{j=1}^{p}\Big(\frac{1}{n}\sum_{i=1}^{n}S_{ij}\Big)^{\gamma} + \sum_{1\le v\le\gamma}\binom{\gamma}{v}\sum_{j=1}^{p}\Big(\frac{1}{n}\sum_{i=1}^{n}S_{ij}\Big)^{\gamma-v}\Big(\frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}-\hat{\mu}_{0i})X_{ij}\Big)^{v} \\
&= T_{\gamma 0} + \sum_{v=1}^{\gamma}T_{\gamma v}, \qquad \text{say.}
\end{aligned}$$

Under the null hypothesis and proposed assumptions, we have $\{T_{\gamma 0} - \psi(\gamma)\}/\omega(\gamma) \xrightarrow{d} N(0,1)$ as $n, p \to \infty$. Then we discuss two cases: $v = 1$ and $v > 1$ separately for the orders of $T_{\gamma v}$, $1 \le v \le \gamma$.

When $v = 1$, we have

$$\begin{aligned}
&E[(T_{\gamma 1})^2] \\
&= E\Big[\frac{C}{n^2}\sum_{j_1}^{p}\sum_{j_2}^{p}\Big(\sum_{i=1}^{n}\frac{1}{n}S_{ij_1}\Big)^{\gamma-1}\Big(\sum_{i=1}^{n}\frac{1}{n}S_{ij_2}\Big)^{\gamma-1}\sum_{i=1}^{n}((\mu_{0i}-\hat{\mu}_{0i})X_{ij_1})\sum_{i=1}^{n}((\mu_{0i}-\hat{\mu}_{0i})X_{ij_2})\Big] \\
&= \frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\times\big(\sum_{l\in\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n} + \sum_{l\notin\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-1} \\
&\qquad\qquad \times\big(\sum_{l\in\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n} + \sum_{l\notin\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-1}\Big].
\end{aligned}$$

By Binomial theorem, we have

$$\begin{aligned}
\big(\sum_{l}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-1} \le{}& \big(\sum_{l\notin\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-1} + C\sum_{l\in\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big(\sum_{l\notin\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-2} + \dots \\
&+ C\big(\sum_{l\in\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-2}\sum_{l\notin\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big) + C\big(\sum_{l\in\{i_1,\dots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-1}.
\end{aligned}$$

Then

$$E[(T_{\gamma 1})^2] = \sum_{k_1=1}^{\gamma}\sum_{k_2=1}^{\gamma}\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{k_1-1} \times \big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{k_2-1}\big(\sum_{l\notin\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-k_1}\big(\sum_{l\notin\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-k_2}\Big]$$
$$= \sum_{k_1=1}^{\gamma}\sum_{k_2=1}^{\gamma} T_{\gamma 1 k_1 k_2}, \qquad \text{say.}$$

To prove the order of $|T_{\gamma 1}|$ is ignorable, we discuss two situations: $k_1+k_2 \le 6$ and $k_1+k_2 > 6$. First, we focus on the situation with $k_1+k_2 \le 6$ and discuss the order of each term separately. For $T_{\gamma 111}$, we have

$$\begin{aligned}
&T_{\gamma 111}\\
=&\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\notin\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-1}\big(\sum_{l\notin\{i_1,i_2,i_3,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-1}\Big]\\
=&\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big]E\Big[\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-1}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-1}\Big]\\
=&O(n^{-2})\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big]\times O(n^{-(\gamma-1)})\\
=&O(n^{-2})\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big]\times O(n^{-(\gamma-1)}).
\end{aligned}$$

Note that

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}|\tilde{\mathbb{Z}}\big]\times O(n^{-(\gamma-1)})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}\epsilon_{0i_4}|\tilde{\mathbb{Z}}\big]\times O(q_0^2 n^{-(\gamma+1)})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}^2|\tilde{\mathbb{Z}}\big]\times O(q_0^2 n^{-(\gamma+1)})\\
&+\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}\epsilon_{0i_1}X_{i_2j_2}\epsilon_{0i_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2 n^{-(\gamma+1)})\\
&+\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}\epsilon_{0i_1}^2X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2 n^{-(\gamma+1)}).
\end{aligned}$$

We discuss each term separately. By a similar discussion of $E[T_{114}|\tilde{\mathbb{Z}}]$, for the first term in $T_{\gamma 111}$, we have

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}^2|\tilde{\mathbb{Z}}\big] \times O(q_0^2 n^{-(\gamma+1)}) \\
=&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}\big] \times O(q_0^2 n^{-\gamma}) \\
=&o(pn^2 q_0^{-2}) \times O(q_0^2 n^{-\gamma}) = o(pn^{-\gamma+2}).
\end{aligned}$$

For the second term in $T_{\gamma 111}$, by noting that $E[X_{ij}\epsilon_{0i}|\mathbb{Z}] = 0$, we have

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}\epsilon_{0i_1}X_{i_2j_2}\epsilon_{0i_2}|\tilde{\mathbb{Z}}\big] \times O(q_0^2 n^{-(\gamma+1)}) \\
=&\sum_{j_1,j_2}\sum_{i_1} E\big[X_{i_1j_1}X_{i_1j_2}\epsilon_{0i_1}^2|\tilde{\mathbb{Z}}\big] \times O(q_0^2 n^{-(\gamma+1)}) \\
=&O(p^2 n) \times O(q_0^2 n^{-(\gamma+1)}) = O(q_0^2 p^2 n^{-\gamma}) = o(pn^{-\gamma+2}).
\end{aligned}$$

For the third term in $T_{\gamma 111}$, similar to the discussion of T_{111}, we have

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}\epsilon_{0i_1}^2 X_{i_2j_2}|\tilde{\mathbb{Z}}\big] \times O(q_0^2 n^{-(\gamma+1)}) \\
=&o(pn^2 q_0^{-2}) \times O(q_0^2 n^{-(\gamma+1)}) = o(pn^{-\gamma+1}) = o(pn^{-\gamma+2}).
\end{aligned}$$

By combing the above three equations, we have $T_{\gamma 111} = o(pn^{-\gamma})$.

Similarly,

$$\begin{aligned}
&T_{\gamma 121} \\
=&\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n} \\
&\qquad\times\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-2}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-1}\Big] \\
=&\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\Big] \\
&\qquad\times E\Big[\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-2}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-1}\Big] \\
=&\frac{C}{n^3}\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i1j_1}^2X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_3}b_{i_2i_3}\epsilon_{0i_3}^2 + X_{i_1j_1}X_{i_2j_2}b_{i_1i_3}b_{i_2i_3}\epsilon_{0i_3}^3X_{i_3j_1}\big] \times O(n^{-(\gamma-1)}) \\
=&O(n^{-\gamma-2})\sum_{j_1,j_2}\sum_{i_1,i_2,i_3}\Big(E\big[X_{i1j_1}^2X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_3}b_{i_2i_3}\epsilon_{0i_3}^2\big] + E\big[X_{i_1j_1}X_{i_2j_2}b_{i_1i_3}b_{i_2i_3}\epsilon_{0i_3}^3X_{i_3j_1}\big]\Big).
\end{aligned}$$

Again, we discuss each term separately. Note that since $E[\epsilon|\mathbb{X},\tilde{\mathbb{Z}}]=0$, we have $E[X_{ij}^2\epsilon_{0i}|\tilde{\mathbb{Z}}]=0$. Thus

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}^2X_{i_2j_2}\epsilon_{0i_1}b_{i_1i_3}b_{i_2i_3}\epsilon_{0i_3}^2|\tilde{\mathbb{Z}}\big]\times O(n^{-\gamma-2})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}^2X_{i_2j_2}\epsilon_{0i_1}^3|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma-4})\\
=&O(p^2n^2)\times O(q_0^2n^{-\gamma-4})=O(pn^{-\gamma}pq_0^2n^{-2})=o(pn^{-\gamma}).
\end{aligned}$$

Similarly, for the second term in $T_{\gamma 121}$, we have

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}X_{i_2j_2}b_{i_1i_3}b_{i_2i_3}\epsilon_{0i_3}^3X_{i_3j_1}|\tilde{\mathbb{Z}}\big]\times O(n^{-\gamma-2})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}\big]E\big[\epsilon_{0i_3}^3X_{i_3j_1}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma-4})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma-3})\\
=&o(pn^2q_0^{-2})\times O(q_0^2n^{-\gamma-3})=o(pn^{-\gamma}).
\end{aligned}$$

Combining the above two equations, we have $T_{\gamma 121}=o(pn^{-\gamma})$.

For $T_{\gamma 122}$, we have

$$\begin{aligned}
&T_{\gamma 122}\\
=&\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\\
&\qquad\times\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-2}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-2}\Big]\\
=&\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\Big]\\
&\qquad\times E\Big[\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-2}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-2}\Big]\\
=&\frac{C}{n^4}\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_1}\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_2}\big]\times O(n^{-\gamma+2}).
\end{aligned}$$

Note that

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_1}\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_2}|\tilde{\mathbb{Z}}\big]\times O(n^{-\gamma+2})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}^2X_{i_2j_2}^2\epsilon_{0i_1}^2\epsilon_{0i_2}^2+X_{i_1j_1}^2X_{i_1j_2}X_{i_2j_2}\epsilon_{0i_1}^4|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma})\\
&+\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}^2\epsilon_{0i_1}X_{i_2j_2}^2\epsilon_{0i_2}\epsilon_{0i_3}^2|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma})\\
&+\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2X_{i_2j_2}\epsilon_{0i_3}^2|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma})\\
&+\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}^4X_{i_3j_1}X_{i_3j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma})\\
&+\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\big[X_{i_1j_1}X_{i_2j_2}X_{i_3j_1}\epsilon_{0i_3}^2X_{i_4j_2}\epsilon_{0i_4}^2|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma}).
\end{aligned}$$

Then we discuss each term separately. For the first term, we have

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2} E\Big[X_{i_1j_1}^2X_{i_2j_2}^2\epsilon_{0i_1}^2\epsilon_{0i_2}^2+X_{i_1j_1}^2X_{i_1j_2}X_{i_2j_2}\epsilon_{0i_1}^4|\tilde{\mathbb{Z}}\Big]\times O(q_0^2n^{-\gamma})\\
=&O(p^2n^2)\times O(q_0^2n^{-\gamma})=o(pn^{-\gamma+4}).
\end{aligned}$$

For the second term, by noting that $E[X_{ij}^2\epsilon_{0i}|\tilde{\mathbb{Z}}]=0$, we have

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}^2\epsilon_{0i_1}X_{i_2j_2}^2\epsilon_{0i_2}\epsilon_{0i_3}^2|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}^2\epsilon_{0i_1}X_{i_2j_2}^2\epsilon_{0i_2}|\tilde{\mathbb{Z}}\big]\times E\big[\epsilon_{0i_3}^2|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_3} E\big[X_{i_1j_1}^2\epsilon_{0i_1}^2X_{i_1j_2}^2|\tilde{\mathbb{Z}}\big]\times E\big[\epsilon_{0i_3}^2|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma})\\
=&O(p^2n)\times O(n)\times O(q_0^2n^{-\gamma})=o(pn^{-\gamma+4}).
\end{aligned}$$

For the third term, we have

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2X_{i_2j_2}\epsilon_{0i_3}^2|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times\sum_{i_3} E\big[\epsilon_{0i_3}^2|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma+1})\\
=&\sum_{j_1,j_2}\sum_{i_1\neq i_2} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma+1})+\sum_{j_1,j_2}\sum_{i_1} E\big[X_{i_1j_1}^2X_{i_1j_2}^2\epsilon_{0i_1}^2|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma+1}).
\end{aligned}$$

We discuss each term separately as follows. First,

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1\neq i_2} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma+1})\\
=&\sum_{i_1\neq i_2}\sum_{j_1\notin P_0,j_2\notin P_0} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2|\tilde{\mathbb{Z}}\big]\times E\big[X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma+1})\\
&+\sum_{i_1\neq i_2}\sum_{j_1\notin P_0,j_2\in P_0} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2|\tilde{\mathbb{Z}}\big]\times E\big[X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma+1})\\
&+\sum_{i_1\neq i_2}\sum_{j_1\in P_0,j_2\notin P_0} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2|\tilde{\mathbb{Z}}\big]\times E\big[X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma+1})\\
&+\sum_{i_1\neq i_2}\sum_{j_1\in P_0,j_2\in P_0} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2|\tilde{\mathbb{Z}}\big]\times E\big[X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma+1})
\end{aligned}$$

By noting that $E[X_{i_2j_2}|\tilde{\mathbb{Z}}]=0$, we have

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1\neq i_2} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma+1})\\
=&\sum_{i_1\neq i_2}\sum_{j_1\notin P_0,j_2\in P_0} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2|\tilde{\mathbb{Z}}\big]\times E\big[X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma+1})\\
&+\sum_{i_1\neq i_2}\sum_{j_1\in P_0,j_2\in P_0} E\big[X_{i_1j_1}^2X_{i_1j_2}\epsilon_{0i_1}^2|\tilde{\mathbb{Z}}\big]\times E\big[X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma+1})\\
=&\big[O(n^2pp_0)+O(n^2p_0^2)\big]\times O(q_0^2n^{-\gamma+1})\\
=&O(pn^{-\gamma+4})\times\big[O(p_0q_0^2n^{-1})+O(p_0^2p^{-1}q_0^2n^{-1})\big]\\
=&o(pn^{-\gamma+4}).
\end{aligned}$$

According to assumptions C3 and C4, $p_0=O(p^{\eta_1})$ for a small positive η_1 and $q_0=O(n^{\eta_2})$ for a small positive η_2. Then $p^{1/2-\eta_3/2}q_0^2n^{-1}=o(n^{-1})$ and we can derive the last equation accordingly.

Second, by the discussion of T_{111}, we have

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1} E\big[X_{i_1j_1}^2X_{i_1j_2}^2\epsilon_{0i_1}^2|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma+1})\\
=&O(p^2n)\times O(q_0^2n^{-\gamma+1})=O(pn^{-\gamma+4})\times O(pq_0^2n^{-2})=o(pn^{-\gamma+4}).
\end{aligned}$$

For the fourth term, similar to the derivation of $E[T_{114}|\tilde{\mathbb{Z}}]$, we have

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}^4X_{i_3j_1}X_{i_3j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3} E\big[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}\big]E\big[\epsilon_{0i_3}^4X_{i_3j_1}X_{i_3j_2}|\tilde{\mathbb{Z}}\big]\times O(q_0^2n^{-\gamma})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2} E\big[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}\big]\times O(n)\times O(q_0^2n^{-\gamma})\\
=&o(pn^2q_0^{-2})\times O(q_0^2n^{-\gamma+1})=o(pn^{-\gamma+4}).
\end{aligned}$$

For the last term, by the derivation of $E[T_{114}|\tilde{\mathbb{Z}}]$, we have

$$\begin{aligned}
&\sum_{j_1,j_2}\sum_{i_1,i_2,i_3,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}X_{i_3j_1}\epsilon_{0i_3}^2X_{i_4j_2}\epsilon_{0i_4}^2|\tilde{\mathbb{Z}}\Big]\times O(q_0^2n^{-\gamma})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2} E[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}]\times O(n^2)\times O(q_0^2n^{-\gamma})\\
=&\sum_{j_1,j_2}\sum_{i_1,i_2} E[X_{i_1j_1}X_{i_2j_2}|\tilde{\mathbb{Z}}] = o(pn^2q_0^{-2})\times O(q_0^2n^{-\gamma+2}) = o(pn^{-\gamma+4}).
\end{aligned}$$

Combining the above equations, we have $T_{\gamma 122} = o(pn^{-\gamma})$.

Then we discuss the order of $T_{\gamma 131}$, we have

$$\begin{aligned}
T_{\gamma 131} =& \frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^2\\
&\times\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-3}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-1}\Big]\\
=& O(n^{-2})\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^2\Big]\\
&\times E\big[\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-3}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-1}\Big]\\
=& O(n^{-2})\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^2\Big]\times O(n^{-\gamma+2}).
\end{aligned}$$

$(\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_1}/n)^2$ and $(\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_1}/n)\times(\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_2}/n)$ have the same effect to the order. Similar to the discussion of $T_{\gamma 122}$, we have $T_{\gamma 131} = o(pn^{-\gamma})$. Next, we discuss the order of $T_{\gamma 132}$:

$$\begin{aligned}
&T_{\gamma 132}\\
=&\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^2\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\\
&\times\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-3}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-2}\Big]\\
=&O(n^{-\gamma-3})\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_1}\big)^2\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_2}\Big].
\end{aligned}$$

For a fixed j_1 and j_2,

$$\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_1}\big)^2\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_2}]$$

contains 5 ϵ_{0i} and $E[\epsilon_{0i}|\mathbb{X},\tilde{\mathbb{Z}}] = 0$. Then it at most contains n^2 terms with non-zero expectation. Since $b_{ij} = O(q_0n^{-1})$, we have

$$T_{\gamma 132} = O(n^{-\gamma-3})\times O(p^2n^2q_0^2n^{-2}) = O(p^2q_0^2n^{-\gamma-3}) = O(pn^{-\gamma})\times O(pq_0^2n^{-3}) = o(pn^{-\gamma}).$$

Similarly, we can prove $T_{\gamma 141} = o(pn^{-\gamma})$. For $T_{\gamma 133}$, we have

$$\begin{aligned}
&T_{\gamma 133}\\
=&\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^2\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^2\\
&\qquad\times\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-3}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-3}\Big]\\
=&O(n^{-\gamma-3})\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_1}\big)^2\big(\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_2}\big)^2\Big].
\end{aligned}$$

Since $\sum_{i_1,\ldots,i_4} E[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}(\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_1})^2(\sum_{l\in\{i_1,\ldots,i_4\}}\epsilon_{0l}X_{lj_2})^2]$ contains 6 ϵ_{0i} and $E[\epsilon_{0i}|\mathbb{X},\tilde{\mathbb{Z}}]=0$, for a fixed j_1 and j_2, it at most contains n^3 terms with non-zero expectation. Thus

$$T_{\gamma 133} = O(n^{-\gamma-3})\times O(p^2n^3q_0^2n^{-2}) = O(q_0^2p^2n^{-\gamma-2}) = O(pn^{-\gamma})\times O(pq_0^2n^{-2}) = o(pn^{-\gamma}).$$

Similarly, we can prove $T_{\gamma 1k_1k_2} = o(pn^{-\gamma})$ for $k_1+k_2=6$.

For $k_1+k_2\geq 7$, we have

$$\begin{aligned}
&\frac{C}{n^2}\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\times\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{k_1-1}\\
&\times\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{k_2-1}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-k_1}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-k_2}\Big]\\
=&O(n^{-2})\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{k_1-1}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{k_2-1}\Big]\\
&\times E\Big[\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{\gamma-k_1}\big(\sum_{l\notin\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{\gamma-k_2}\Big]\\
=&O(n^{-2})\sum_{j_1,j_2}\sum_{i_1,\ldots,i_4} E\Big[X_{i_1j_1}X_{i_2j_2}\epsilon_{0i_3}b_{i_1i_3}\epsilon_{0i_4}b_{i_2i_4}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_1}}{n}\big)^{k_1-1}\big(\sum_{l\in\{i_1,\ldots,i_4\}}\frac{\epsilon_{0l}X_{lj_2}}{n}\big)^{k_2-1}\Big]\\
&\times O(n^{-\gamma+\lfloor(k_1+k_2)/2\rfloor})\\
=&O(n^{-2})O(p^2\times n^4\times q_0^2n^{-2}\times n^{-(k_1+k_2-2)})\times O(n^{-\gamma+\lfloor(k_1+k_2)/2\rfloor})\\
=&O(pn^{-\gamma})\times O(q_0^2pn^{-(k_1+k_2-2)+\lfloor(k_1+k_2)/2\rfloor}) = o(pn^{-\gamma}).
\end{aligned}$$

By noting that $pq_0^2 = o(n^2)$ and for $k_1+k_2\geq 7$, $-(k_1+k_2-2)+\lfloor(k_1+k_2)/2\rfloor\geq 2$, we can derive the last equation. In summary, we have $|T_{\gamma 1}| = o_p(n^{-\gamma/2}\sqrt{p})$.

Next, we discuss $1<v\leq\gamma$.

By Minkowski's inequality, we have

$$E[|T_{\gamma v}|]\leq C\sum_{j=1}^{p}E\Big[\Big|\Big(\frac{1}{n}\sum_{i=1}^{n}S_{ij}\Big)^{\gamma-v}\Big(\frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}-\hat{\mu}_{0i})X_{ij}\Big)^{v}\Big|\Big].$$

Then by Cauchy-Schwarz inequality,

$$E[|T_{\gamma v}|] \le C\sum_{j=1}^{p} E\Big[\big(\frac{1}{n}\sum_{i=1}^{n} S_{ij}\big)^{2(\gamma-v)}\Big]^{1/2} E\Big[\big(\frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}-\hat{\mu}_{0i})X_{ij}\big)^{2v}\Big]^{1/2}$$
$$\le O(n^{-(\gamma-v)/2})\sum_{j=1}^{p} E\Big[\big(\frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}-\hat{\mu}_{0i})X_{ij}\big)^{2v}\Big]^{1/2}.$$

Next, we derive the order of $T_{2v,j} = \big(\frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}-\hat{\mu}_{0i})X_{ij}\big)^{2v}$ for any positive integer v. For $j \in P_0$, we have

$$E[T_{2v,j}|\tilde{\mathbb{Z}}] = \frac{1}{n^{2v}}\sum_{i_1,i_2,\ldots,i_{4v}} E[X_{i_1 j}b_{i_1 i_{2v+1}}\epsilon_{0i_{2v+1}}X_{i_2 j}b_{i_2 i_{2v+2}}\epsilon_{0i_{2v+1}}\times\cdots\times X_{i_{2v}j}b_{i_{2v}i_{4v}}\epsilon_{0i_{4v}}|\tilde{\mathbb{Z}}]$$
$$\le \frac{Cq_0^{2v}}{n^{4v}}\sum_{i_1,i_2,\ldots,i_{3v}} E[X_{i_1 j}X_{i_2 j}\times\cdots\times X_{i_{2v}j}\times\epsilon^2_{0i_{2v+1}}\epsilon^2_{0i_{2v+2}}\times\cdots\times\epsilon^2_{0i_{3v}}|\tilde{\mathbb{Z}}]$$
$$= O(q_0^{2v}n^{-v}).$$

Note that for $j \notin P_0$, $E[X_{ij}|\tilde{\mathbb{Z}}] = 0$. Then for For $j \notin P_0$,

$$E[T_{2v,j}|\tilde{\mathbb{Z}}] = \frac{1}{n^{2v}}\sum_{i_1,i_2,\ldots,i_{4v}} E[X_{i_1 j}b_{i_1 i_{2v+1}}\epsilon_{0i_{2v+1}}X_{i_2 j}b_{i_2 i_{2v+2}}\epsilon_{0i_{2v+1}}\times\cdots\times X_{i_{2v}j}b_{i_{2v}i_{4v}}\epsilon_{0i_{4v}}|\tilde{\mathbb{Z}}]$$
$$\le \frac{Cq_0^{2v}}{n^{4v}}\sum_{i_1,i_2,\ldots,i_{2v}} E[X^2_{i_1 j}X^2_{i_2 j}\times\cdots\times X^2_{i_v j}\times\epsilon^2_{0i_{v+1}}\epsilon^2_{0i_{v+2}}\times\cdots\times\epsilon^2_{0i_{2v}}|\tilde{\mathbb{Z}}]$$
$$= O(q_0^{2v}n^{-2v}).$$

In summary, $E[T_{2v,j}] = O(q_0^{2v}n^{-v})$ if $j \in P_0$ and $E[T_{2v,j}] = O(q_0^{2v}n^{-2v})$ if $j \notin P_0$. Then we have $\sum_{j=1}^{p} E\big[\big(\frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}-\hat{\mu}_{0i})X_{ij}\big)^{2v}\big]^{1/2} = O(p_0 q_0^v n^{-v/2} + pq_0^v n^{-v})$. This leads to

$$E[|T_{\gamma v}|] \le O(n^{-(\gamma-v)/2})\times O(p_0 q_0^v n^{-v/2} + pq_0^v n^{-v})$$
$$= O(p_0 q_0^v n^{-\gamma/2} + \sqrt{p}n^{-\gamma/2}\sqrt{p}n^{-v/2}q_0^v).$$

Note that by assumption C5, $pq_0^4 = o(n^2)$, $v \ge 2$, and by assumption C4, $p_0 = O(p^{\eta_1})$, $q_0 = O(n^{\eta_2})$ for some small positive η_1 and η_2, we have $E[|T_{\gamma v}|] = o(\sqrt{p}n^{-\gamma/2})$, leading to $|T_{\gamma v}| = o_p(\sqrt{p}n^{-\gamma/2})$.

In summary, we have proved for any finite γ,

$$[\{L(\gamma,\hat{\mu}_0)-\psi(\gamma)\}/\omega(\gamma)]'_{\gamma\in\Gamma'} = [\{L(\gamma,\mu_0)-\psi(\gamma)\}/\omega(\gamma)]'_{\gamma\in\Gamma'} + o_p(1).$$

(ii) Asymptotic null distribution for iSPU(∞). Define

$$\tilde{V}_{ij} = (Y_i - \hat{\mu}_{0i})X_{ij}/\sqrt{\sigma_{jj}}, \qquad 1 \le i \le n, 1 \le j \le p.$$

Let $\tilde{W}_j = \sum_{i=1}^{n}\tilde{V}_{ij}/\sqrt{n}$. We discuss two cases: $j \in P_0$ and $j \notin P_0$.

For case $j \in P_0$: we define ϵ is a small constant. Note that

$$\begin{aligned}
&Pr\big(\max_{j\in P_0}\tilde{W}_j^2 > \epsilon\log p\big)\\
\leq& Pr\big(\max_{j\in P_0}|\tilde{W}_j| > \epsilon(\log p)^{1/2}\big)\\
\leq& Pr\Big(\max_{j\in P_0}\Big|\frac{\sum_{i=1}^n(Y_i-\mu_{0i}+\mu_{0i}-\hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\Big| > (\epsilon\log p)^{1/2}\Big)\\
\leq& Pr\Big(\max_{j\in P_0}\Big|\frac{\sum_{i=1}^n(Y_i-\mu_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\Big| > \frac{(\epsilon\log p)^{1/2}}{2}\Big)\\
&+ Pr\Big(\max_{j\in P_0}\Big|\frac{\sum_{i=1}^n(\mu_{0i}-\hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\Big| > \frac{(\epsilon\log p)^{1/2}}{2}\Big).
\end{aligned}$$

For the first term, we have

$$Pr\Big(\max_{j\in P_0}\Big|\frac{\sum_{i=1}^n(Y_i-\mu_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\Big| > \frac{(\epsilon\log p)^{1/2}}{2}\Big) \leq p_0 Pr\Big(\Big|\frac{\sum_{i=1}^n S_{ij}}{\sqrt{\sigma_{jj}n}}\Big| > \frac{(\epsilon\log p)^{1/2}}{2}\Big).$$

Note that S_{ij} follows a sub-Gaussian distribution (C2) and S_{i_1j} and S_{i_2j} are independent for $i_1 \neq i_2$. Suppose $S_{1j},\ldots,S_{nj}$ be n independent random variables such that S_{ij} follows sub-Gaussian distribution subG$(0,\sigma^2)$. Then for any $a\in\mathbb{R}^n$, using a Chernoff bound, we have $Pr\left(|\sum_{i=1}^n a_iS_{ij}| > t\right) \leq 2\exp\left(-t^2/(2\sigma^2||a||_2^2)\right)$. Similarly, we have

$$\begin{aligned}
&Pr\Big(\max_{j\in P_0}\Big|\frac{\sum_{i=1}^n(Y_i-\mu_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\Big| > \frac{(\epsilon\log p)^{1/2}}{2}\Big)\\
\leq& p_0\times 2\exp\Big(-\frac{\epsilon\log p/4}{2}\Big) = 2p_0p^{-\epsilon/8} = o(1).
\end{aligned}$$

By noting that $p_0 = p^{\eta_1}$, where η_1 is a small constant, we have the last equation.

For the second term, we have

$$\begin{aligned}
&Pr\Big(\max_{j\in P_0}\Big|\frac{\sum_{i=1}^n(\mu_{0i}-\hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\Big| > \frac{(\epsilon\log p)^{1/2}}{2}\Big)\\
\leq& \Pr\Big(\max_{j\in P_0}\Big|\frac{\sum_{i_1,i_2}X_{i_1j}\epsilon_{0i_2}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}}\Big| > \frac{(\epsilon\log p)^{1/2}}{2}\Big)\\
\leq& Pr\Big(\max_{j\in P_0}\Big|\sum_{i_2=1}^n\epsilon_{0i_2}\Big(\frac{\sum_{i_1}X_{i_1j}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}}\Big)\Big| > \frac{(\epsilon\log p)^{1/2}}{2}\Big|\max_{i_2}\frac{\sum_{i_1}X_{i_1j}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}} < \frac{C}{\sqrt{\sigma_{jj}n}}\Big)\\
&+ Pr\Big(\max_{i_2}\frac{\sum_{i_1}X_{i_1j}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}} \geq \frac{C}{\sqrt{\sigma_{jj}n}}\Big).
\end{aligned}$$

We discuss these two terms separately. For the first term, we have

$$\begin{aligned}
&Pr\Big(\max_{j\in P_0}\big|\sum_{i_2=1}^{n}\epsilon_{0i_2}\big(\frac{\sum_{i_1}X_{i_1j}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}}\big)\big|>\frac{(\epsilon\log p)^{1/2}}{2}\Big|\max_{i_2}\frac{\sum_{i_1}X_{i_1j}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}}<\frac{C}{\sqrt{\sigma_{jj}n}}\Big)\\
\le& p_0 Pr\Big(\big|\sum_{i_2=1}^{n}\epsilon_{0i_2}\big(\frac{\sum_{i_1}X_{i_1j}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}}\big)\big|>\frac{(\epsilon\log p)^{1/2}}{2}\Big|\max_{i_2}\frac{\sum_{i_1}X_{i_1j}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}}<\frac{C}{\sqrt{\sigma_{jj}n}}\Big)\\
\le& p_0 E\Big[Pr\Big(\big|\sum_{i_2=1}^{n}\epsilon_{0i_2}\big(\frac{\sum_{i_1}X_{i_1j}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}}\big)\big|>\frac{(\epsilon\log p)^{1/2}}{2}\Big|\max_{i_2}\frac{\sum_{i_1}X_{i_1j}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}}<\frac{C}{\sqrt{\sigma_{jj}n}},\mathbb{X},\tilde{\mathbb{Z}}\Big)\Big].
\end{aligned}$$

By noting that ϵ_{0i} follows a sub-Gaussian distribution, we have

$$\begin{aligned}
&Pr\Big(\max_{j\in P_0}\big|\sum_{i_2=1}^{n}\epsilon_{0i_2}\big(\frac{\sum_{i_1}X_{i_1j}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}}\big)\big|>\frac{(\epsilon\log p)^{1/2}}{2}\Big|\max_{i_2}\frac{\sum_{i_1}X_{i_1j}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}}<\frac{C}{\sqrt{\sigma_{jj}n}},\mathbb{X},\tilde{\mathbb{Z}}\Big)\\
\le& p_0\times 2\exp\Big(-\frac{\epsilon\log p/4}{2C^2}\Big)=2p_0p^{-\epsilon/(8C^2)}=o(1).
\end{aligned}$$

For the second term, we have

$$Pr\Big(\max_{i_2}\frac{\sum_{i_1}X_{i_1j}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}}\ge\frac{C}{\sqrt{\sigma_{jj}n}}\Big)\le nE\Big[Pr\Big(\frac{\sum_{i_1}X_{i_1j}}{n}\ge C/q_0\big|\tilde{\mathbb{Z}}\Big)\Big].$$

By central limit theorem and the Gaussian tail inequality, we have

$$Pr\Big(\max_{i_2}\frac{\sum_{i_1}X_{i_1j}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}}\ge\frac{C}{\sqrt{\sigma_{jj}n}}\Big)\le n\times\frac{2\exp(-(Cn^{1/2}/q_0)^2/2)}{Cn^{1/2}/q_0}.$$

By noting that $q_0=n^{\eta_2}$ for a small positive η_2, we have

$$Pr\Big(\max_{i_2}\frac{\sum_{i_1}X_{i_1j}b_{i_1i_2}}{\sqrt{\sigma_{jj}n}}\ge\frac{C}{\sqrt{\sigma_{jj}n}}\Big)\le Cn^{1/2+\eta_2}\times\exp(-n^{1-2\eta_2}/2)=o(1).$$

In summary, as $n,p\to\infty$, $Pr\big(\max_{j\in P_0}\tilde{W}_j^2>\epsilon\log p\big)=o(1)$. Then we focus on the second situation. Define $V_{ij}=(Y_i-\mu_{0i})X_{ij}/\sqrt{\sigma_{jj}}$, $\hat{V}_{ij}=V_{ij}I(|V_{ij}|\le\tau_n)$ for $i=1,\ldots,n$ and $j=1,\ldots,p$, where $\tau_n=2\eta^{-0.5}\sqrt{\log(p+n)}$. Further define $W_j=\sum_{i=1}^n(Y_i-\mu_{0i})X_{ij}/\sqrt{\sigma_{jj}n}$ and $\hat{W}_j=\sum_{i=1}^n\hat{V}_{ij}/\sqrt{n}$. Then we have

$$\begin{aligned}
&Pr\big(\max_{j\notin P_0}|\tilde{W}_j-W_j|\ge\frac{1}{\log p}\big)\\
&\le np\max_{j\notin P_0}Pr(|V_{1j}|\ge\tau_n)+Pr\Big(\max_{j\notin P_0}\big|\sum_{i=1}^{n}\frac{(\mu_{0i}-\hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}\big|\ge\frac{1}{\log p}\Big).
\end{aligned}$$

From the proof of Lemma 1, the first term is $O(1/p+1/n)$ and thus we only need discuss the second term. By the Markov inequality and the Jensen's inequality,

$$
\begin{aligned}
&Pr\Big(\max_{j\notin P_0}|\sum_{i=1}^{n}\frac{(\mu_{0i}-\hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}}|\geq\frac{1}{\log p}\Big)\\
&\leq Pr\Big(\max_{j\notin P_0}(\sum_{i=1}^{n}\frac{(\mu_{0i}-\hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}})^{32}\geq\frac{1}{(\log p)^{32}}\Big)\\
&\leq pPr\Big((\sum_{i=1}^{n}\frac{(\mu_{0i}-\hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}})^{32}\geq\frac{1}{(\log p)^{32}}\Big)\\
&\leq p\log pE\Big[(\sum_{i=1}^{n}\frac{(\mu_{0i}-\hat{\mu}_{0i})X_{ij}}{\sqrt{\sigma_{jj}n}})^{32}\Big]\\
&\leq p\log p\times O(n^{-16}q_0^{16})=o(1).
\end{aligned}
$$

Thus, we have $Pr\big(\max_{1\leq j\leq p}|\tilde{W}_j-W_j|\geq 1/\log p\big)=o(1)$ as $n,p\to\infty$. Further note that

$$
|\max_{j\notin P_0}W_j^2-\max_{j\notin P_0}\tilde{W}_j^2|\leq 2\max_{j\notin P_0}|W_j|\max_{j\notin P_0}|W_j-\tilde{W}_j|+\max_{j\notin P_0}|W_j-\tilde{W}_j|^2.
$$

The above two inequalities indicate that when $n,p\to\infty$, $|\max_{j\notin P_0}W_j^2-\max_{j\notin P_0}\tilde{W}_j^2|\to 0$. By Cai et al. (2014), we have

$$
Pr\big(\max_{j\notin P_0}\tilde{W}_j^2-2\log p+\log\log p\leq x\big)\to\exp\{-\pi^{-1/2}\exp(-x/2)\}.
$$

Note that

$$
\max_{1\leq j\leq p}\tilde{W}_j^2=\max\big(\max_{j\in P_0}\tilde{W}_j^2,\max_{j\notin P_0}\tilde{W}_j^2\big)=\max_{j\notin P_0}\tilde{W}_j^2.
$$

Thus,

$$
Pr\big(\max_{1\leq j\leq p}\tilde{W}_j^2-2\log p+\log\log p\leq x\big)\to\exp\{-\pi^{-1/2}\exp(-x/2)\}.
$$

Note that $\hat{\sigma}_{jj}=(1+o(1))\sigma_{jj}$ and by Slutsky's theorem, we have

$$
Pr\big(\max_{1\leq j\leq p}\frac{n(\frac{1}{n}\sum_{i=1}^{n}U_{ij})^2}{\check{\sigma}_{jj}}-2\log p+\log\log p\leq x\big)\to\exp\{-\pi^{-1/2}\exp(-x/2)\}.
$$

(iii) By proof in (i) and (ii), we have

$$
[\{L(\gamma,\hat{\mu}_0)-\psi(\gamma)\}/\omega(\gamma)]'_{\gamma\in\Gamma'}=[\{L(\gamma,\mu_0)-\psi(\gamma)\}/\omega(\gamma)]'_{\gamma\in\Gamma'}+o_p(1)
$$

and $L(\infty,\hat{\mu}_0)=L(\infty,\mu_0)+o_p(1)$. By Lemma 1, $[\{L(\gamma,\mu_0)-\psi(\gamma)\}/\omega(\gamma)]'_{\gamma\in\Gamma'}$ is asymptotically independent with $L(\infty,\mu_0)$. Note that $o_p(1)$ is asymptotic independent with $L(\infty,\mu_0)$ and $[\{L(\gamma,\mu_0)-\psi(\gamma)\}/\omega(\gamma)]'_{\gamma\in\Gamma'}$, thus $[\{L(\gamma,\hat{\mu}_0)-\psi(\gamma)\}/\omega(\gamma)]'_{\gamma\in\Gamma'}$ is asymptotically independent with $L(\infty,\hat{\mu}_0)$. This completes the proof. ■

Appendix D. Details of asymptotic power analysis

To derive some propositions, we define some additional notation. Under an alternative H_A : $\beta \neq 0$, we denote the mean and variance of $L(\gamma, \mu_0)$ with $\gamma < \infty$ by $\psi_A(\gamma) = \sum_{j=1}^{p} \psi_A^{(j)}(\gamma)$ with $\psi_A^{(j)}(\gamma) = E[L^{(j)}(\gamma, \mu_0)|H_A]$ for $1 \leq j \leq p$, and by $\omega_A^2(\gamma) = \text{var}[L(\gamma, \mu_0)|H_A]$, respectively. Define $\tilde{S}_{ij} \equiv (Y_i - \mu_{0i}^A)X_{ij}$ for $1 \leq i \leq n$ and $1 \leq j \leq p$ and $\tilde{\sigma}_{kj} = \text{cov}[\tilde{S}_{1k}, \tilde{S}_{1j}]$ for $1 \leq k, j \leq p$.

Next, we introduce Propositions 1 and 2 to calculate the $\psi_A(\gamma)$ and $\omega_A^2(\gamma)$, respectively.

Proposition 1. *Under assumptions C8–C9 and $H_A : \beta \neq 0$, we have*

$$\psi_A^{(j)}(\gamma) \sim \tilde{\psi}^{(j)}(\gamma) + \sum_{c=1}^{\gamma} \binom{\gamma}{c} \Delta_j^c \tilde{\psi}^{(j)}(\gamma - c),$$

where $\sim$ stands for the two sides are in the same order, $\tilde{\psi}^{(j)}(1) = 0$, $\tilde{\psi}^{(j)}(\gamma) = \frac{\gamma!}{d!2^d} n^{-d} \tilde{\sigma}_{jj}^d + o(n^{-d})$ if $\gamma = 2d$, and $\tilde{\psi}^{(j)}(\gamma) = o(n^{-(d+1)})$ if $\gamma = 2d + 1$. In particular, $\mu_A^{(j)}(1) = \Delta_j$ and $\mu_A(1) = \sum_{j=1}^{p} \Delta_j$.

Proof of Proposition 1. Under the H_A, the true conditional mean of Y_i is μ_{0i}^A. Then similarly to Theorem 1, we can calculate $\tilde{\psi}^{(j)}(\gamma)$ accordingly.

Under the alternative H_A, it is trivial to find $\mu_A(1) = \sum_{i=1}^{p} \Delta_i$ since $\tilde{\psi}(1) = 0$. Next, we focus on $\gamma \geq 2$. The mean function of $L^{(j)}(\gamma)$ under H_A equals

$$\begin{aligned}
E[L^{(j)}(\gamma)] =& E\Big[\big(\frac{1}{n}\sum_{i=1}^{n} U_{ij}\big)^{\gamma}\Big] \\
=& E\Big[\big(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_{0i}^A)X_{ij} + \frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}^A - \mu_{0i})X_{ij}\big)^{\gamma}\Big] \\
=& \sum_{0 \leq a \leq \gamma} \binom{\gamma}{a} E\Big[\big(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_{0i}^A)X_{ij}\big)^{a}\Big] E\Big[\big(\frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}^A - \mu_{0i})X_{ij}\big)^{\gamma - a}\Big].
\end{aligned}$$

Note that under the local alternative considered here, $\Delta_j = O(n^{-1/2}(\log p)^{\kappa})$ with $\kappa > 0$ for $1 \leq j \leq p$. Then $E\big[\big|(\mu_{0i}^A - \mu_{0i})X_{1j}\big|\big] \leq C|\Delta_j|$ for any positive integer γ, where C is some constant. Thus $E\big[(\frac{1}{n}\sum_{i=1}^{n}(\mu_{0i}^A - \mu_{0i})X_{ij})^a\big] = \Delta_j^a(1 + o(1/n))$. Then we have

$$\psi_A^{(j)}(\gamma) = \tilde{\psi}^{(j)} + \sum_{c=1}^{\gamma} \binom{\gamma}{c} \Delta_j^c \tilde{\psi}^{(j)}(\gamma - c)[1 + o(1/n)].$$

Then Proposition 1 follows directly from the above equation. ■

Proposition 2. *Under assumptions C8–C9 and $H_A : \beta \neq 0$, we have*

$$\omega_A^2(\gamma) \sim \psi_A(2\gamma) - \sum_{j=1}^{p} \psi_A^{(j)}(\gamma)^2 + \sum_{k \neq j} \sum_{h=0}^{\gamma} \sum_{l=0}^{\gamma} \binom{\gamma}{h} \binom{\gamma}{l} \Delta_k^h \Delta_j^l r_{kj}(\gamma - h, \gamma - l),$$

where

$$r_{kj}(h,l)=\begin{cases}\frac{1}{n^c}\sum_{\substack{2c_1+c_3=h\\2c_2+c_3=l\\c_3>0}}\frac{h!l!}{c_3!c_1!c_2!2^{c_1+c_2}}\tilde{\sigma}_{kk}^{c_1}\tilde{\sigma}_{jj}^{c_2}\tilde{\sigma}_{kj}^{c_3}+o(n^{-c}) & \text{if } h+l=2c;\\ \frac{1}{n^{c+1}}\sum_{\substack{a+b=3\\2c_1+c_3=h-a\\2c_2+c_3=l-b}}\frac{h!l!}{a!b!c_3!c_1!c_2!2^{c_1+c_3}}\tilde{\sigma}_{kk}^{c_1}\tilde{\sigma}_{jj}^{c_2}\tilde{\sigma}_{kj}^{c_3}m_{k^aj^b}+o(n^{-(c+1)}) & \text{if } h+l=2c+1,\end{cases}$$

with $m_{k^aj^b}=E\big[\big((Y_1-\mu_{01}^A)X_{1k}\big)^a\big((Y_1-\mu_{01}^A)X_{1j}\big)^b\big]$.

Proof of Proposition 2. Under the local alternative,

$$\begin{aligned}\omega_A^2(\gamma)=&E\Big[\Big\{\sum_{j=1}^p\big(\frac{1}{n}\sum_{i=1}^n(Y_i-\mu_{0i})X_{ij}\big)^\gamma\Big\}^2\Big]-E\Big[\sum_{j=1}^p\Big(\frac{1}{n}\sum_{i=1}^n(Y_i-\mu_{0i})X_{ij}\Big)^\gamma\Big]^2\\ =&\psi_A(2\gamma)-\sum_{i=1}^p\{\psi_A^{(i)}(\gamma)\}^2-\sum_{k\neq j}\psi_A^{(k)}(\gamma)\psi_A^{(j)}(\gamma)\\ &+E\Big[\sum_{k\neq j}\big(\frac{1}{n}\sum_{i=1}^n(Y_i-\mu_{0i})X_{ik}\big)^\gamma\big(\frac{1}{n}\sum_{i=1}^n(Y_s-\mu_{0i})X_{ij}\big)^\gamma\Big]\\ =&\psi_A(2\gamma)-\sum_{i=1}^p\{\psi_A^{(i)}(\gamma)\}^2\\ &-\sum_{k\neq j}\Big[\sum_{c=0}^\gamma\binom{\gamma}{c}\Delta_k^c\tilde{\psi}^{(k)}(\gamma-c)(1+o(1/n))\Big]\Big[\sum_{c=0}^\gamma\binom{\gamma}{c}\Delta_j^c\tilde{\psi}^{(j)}(\gamma-c)(1+o(1/n))\Big]\\ &+\sum_{k\neq j}\sum_{h=0}^\gamma\sum_{l=0}^\gamma\binom{\gamma}{h}\binom{\gamma}{l}\Delta_k^h\Delta_j^lE\Big[\big(\frac{1}{n}\sum_{i=1}^n\tilde{S}_{ik}\big)^{\gamma-h}\big(\frac{1}{n}\sum_{i=1}^n\tilde{S}_{ij}\big)^{\gamma-l}\Big](1+o(1/n)).\end{aligned}$$

By the derivation of Proposition 3 in Wu et al. (2019), the last two terms in the above equation can be simplified as

$$\begin{aligned}&\sum_{k\neq j}\sum_{h=0}^\gamma\sum_{l=0}^\gamma\binom{\gamma}{h}\binom{\gamma}{l}\Delta_k^h\Delta_j^l\Big(E\Big[\big(\frac{1}{n}\sum_{i=1}^n\tilde{S}_{ik}\big)^{\gamma-h}\big(\frac{1}{n}\sum_{i=1}^n\tilde{S}_{ij}\big)^{\gamma-l}\Big]-\tilde{\psi}(k)(\gamma-h)\tilde{\psi}^{(j)}(\gamma-l)\Big)\\ &\sim\sum_{k\neq j}\sum_{h=0}^\gamma\sum_{l=0}^\gamma\binom{\gamma}{h}\binom{\gamma}{l}\Delta_k^h\Delta_j^lr_{kj}(\gamma-h,\gamma-l).\end{aligned}$$

This completes the proof. ■

Proof of iSPU(1) is more powerful when Δ_j is fixed at the same level. We further assume Δ_j equals to Δ for $j\in\mathcal{S}_\eta$ under the alternative. Note that the asymptotic power of iSPU(γ) goes to 1 if $\big(\mu_A(\gamma)-\mu_0(\gamma)\big)n^{\gamma/2}p^{-1/2}\to\infty$, which implies that for any finite γ, a sufficient condition for the asymptotic power of iSPU(γ) goes to 1 is

$$\frac{\Delta}{n^{-1/2}p^{(2\eta-1)/(2\gamma)}}\to\infty,\qquad\text{as } p,n\to\infty.$$

This sufficient condition comes from the fact that $\Delta = O(n^{-1/2}(\log p)^{\kappa})$ and $\mu_A(\gamma) - \mu_0(\gamma) \sim \sum_{i=1}^{p}\sum_{c=1}^{\gamma}\Delta^c O(n^{-(\gamma-c)/2}) \sim p^{1-\eta}\Delta^{\gamma}$. Since $0 < \eta < 1/2$, $p^{(2\eta-1)/(2\gamma)} \to 0$ as $p \to \infty$. Thus, to compare the asymptotic powers of different iSPU(γ), we focus on the local alternative such that $n^{1/2}\Delta \to 0$, as $p, n \to \infty$. Equivalently, we write $\Delta = n^{-1/2}r^{1/2}$, where $r \to 0$ as $p, n \to \infty$.

Then we calculate $\psi_A(\gamma) - \tilde{\psi}(\gamma)$. When γ is odd, by Proposition 1,

$$\begin{aligned}
\psi_A(\gamma) - \tilde{\psi}(\gamma) &= \sum_{j=1}^{p}\sum_{c=1}^{\gamma}\binom{\gamma}{c}\Delta_j^c\tilde{\psi}^{(j)}(\gamma-c)(1+o(1/n)) \\
&\sim \gamma\sum_{j=1}^{p}\Delta_j\tilde{\psi}^{(j)}(\gamma-1) \\
&\sim \gamma\sum_{j=1}^{p}\Delta_j\frac{(\gamma-1)!}{((\gamma-1)/2)!2^{(\gamma-1)/2}}n^{-(\gamma-1)/2} \\
&\sim p^{1-\eta}\Delta\frac{\gamma!}{((\gamma-1)/2)!2^{(\gamma-1)/2}}n^{-\gamma/2}n^{1/2} \\
&\sim \frac{\gamma!}{((\gamma-1)/2)!2^{(\gamma-1)/2}} \times r^{1/2}p^{1-\eta}n^{-\gamma/2}.
\end{aligned}$$

Similarly, when γ is even

$$\begin{aligned}
\psi_A(\gamma) - \tilde{\psi}(\gamma) &\sim \gamma\sum_{j=1}^{p}\Delta_j\tilde{\psi}^{(j)}(\gamma-1) + \binom{\gamma}{2}\sum_{j=1}^{p}\Delta_j^2\tilde{\psi}^{(j)}(\gamma-2) \\
&\sim p^{1-\eta}n^{-1/2}r^{1/2}o\big(n^{-\gamma/2}\big) + p^{1-\eta}nrO(n^{-(\gamma+2)/2}) \\
&\sim o(r^{1/2}p^{1-\eta}n^{-\gamma/2}).
\end{aligned}$$

Further, under this local alternative, $\omega_A(\gamma) \sim \tilde{\omega}(\gamma) \sim c_\gamma p^{1/2}n^{-\gamma/2}$, where c_γ is some constant and can be calculated by Proposition 2. Next, we study $\{\psi_A(\gamma) - \tilde{\psi}(\gamma)\}/\omega_A(\gamma)$, which determines the power. As $n, p \to \infty$,

$$\begin{aligned}
&\frac{\psi_A(\gamma) - \tilde{\psi}(\gamma)}{\omega_A(\gamma)} \sim \frac{\gamma!}{((\gamma-1)/2)!2^{(\gamma-1)/2}c_\gamma} \times r^{1/2}p^{1/2-\eta}, && \gamma \text{ is odd}, \\
&\frac{\psi_A(\gamma) - \tilde{\psi}(\gamma)}{\omega_A(\gamma)} \sim o(1) \times r^{1/2}p^{1/2-\eta}, && \gamma \text{ is odd}.
\end{aligned}$$

The above results show that the asymptotic power of iSPU(γ) does not converge to 1 if $r^{1/2}p^{1/2-\eta} < \infty$. Thus we focus on the local alternative when $r \to 0$ and $r^{1/2}p^{1/2-\eta} \to \infty$. Then the asymptotic power of iSPU with odd γ goes to 1 while iSPU with even γ does not, that is, under the considered alternative, iSPU with odd γ is more powerful than iSPU with even γ. Therefore, we only need to focus on odd γ's and compare their power. To find which odd γ yields an asymptotically more powerful test, we only need to find which γ maximizes $\gamma!/\big(((\gamma-1)/2)!2^{(\gamma-1)/2}c_\gamma\big)$. To simplify our discussion, we first consider the situation where $\tilde{\sigma}_{ij} = 0$ for $i \neq j$. In this case

$$\frac{\gamma!}{((\gamma-1)/2)2^{(\gamma-1)/2}c_\gamma} = \frac{\gamma!}{((\gamma-1)/2)!2^{(\gamma-1)/2}\sqrt{(2\gamma)!/(\gamma!2^\gamma) - (\gamma!)^2/([(\gamma/2)!]^2 2^\gamma)}},$$

which has maximum value 1.66 at $\gamma = 1$. More generally, under the situation where $\tilde{\sigma}_{ij} \geq 0$, a similar calculation gives that iSPU(1) is asymptotically more powerful than iSPU test with other γ. This completes the proof. ■

Proof of iSPU(2) is more powerful when the absolute values of Δ_j are the same but half being positive while the other half being negative. We assume $|\Delta_j|$ equals to Δ for $j \in \mathcal{S}_\eta$ under the alternative. Like previous subsection, we consider the local alternative with $\Delta = n^{-1/2}r^{1/2}$, where $r \to 0$ as $p, n \to \infty$. Similarly, we calculate $\psi_A(\gamma) - \tilde{\psi}(\gamma)$ for both odd and even γ.

For $\gamma = 1$, we have

$$\psi_A(1) - \tilde{\psi}(1) \sim \sum_{j=1}^{p} \Delta_j \sim 0.$$

When $\gamma = 3$, by noting that $\tilde{\psi}^{(j)}(1) = 0$ for $1 \leq j \leq p$, we have

$$\begin{aligned}
\psi_A(3) - \tilde{\psi}(3) &= \sum_{j=1}^{p} \sum_{c=1}^{3} \binom{3}{c} \Delta_j^c \tilde{\psi}^{(j)}(3-c)(1+o(1/n)) \\
&\sim \sum_{j=1}^{p} \left(\Delta_j^3 + \Delta_j \tilde{\psi}^{(j)}(2)\right) \\
&\sim 0.
\end{aligned}$$

For odd $\gamma > 3$, we have

$$\begin{aligned}
\psi_A(\gamma) - \tilde{\psi}(\gamma) &= \sum_{j=1}^{p} \sum_{c=1}^{\gamma} \binom{\gamma}{c} \Delta_j^c \tilde{\psi}^{(j)}(\gamma-c)(1+o(1/n)) \\
&\sim \binom{\gamma}{2} \sum_{j=1}^{p} \Delta_j^2 \tilde{\psi}^{(j)}(\gamma-2) \\
&\sim p^{1-\eta} nr \times o(n^{-(\gamma-1)/2}) \\
&\sim o(rp^{1-\eta} n^{-(\gamma+1)/2}).
\end{aligned}$$

Similarly, for even $\gamma \geq 2$, we have

$$\begin{aligned}
\psi_A(\gamma) - \tilde{\psi}(\gamma) &= \sum_{j=1}^{p} \sum_{c=1}^{\gamma} \binom{\gamma}{c} \Delta_j^c \tilde{\psi}^{(j)}(\gamma-c)(1+o(1/n)) \\
&\sim \binom{\gamma}{2} \sum_{j=1}^{p} \Delta_j^2 \tilde{\psi}^{(j)}(\gamma-2) \\
&\sim \frac{\gamma(\gamma-1)}{2} p^{1-\eta} \Delta^2 \frac{(\gamma-2)!}{((\gamma-2)/2)! 2^{(\gamma-2)/2}} n^{-(\gamma-2)/2} \\
&\sim \frac{\gamma!}{((\gamma-2)/2)! 2^{\gamma/2}} p^{1-\eta} r n^{-\gamma/2}.
\end{aligned}$$

Again, under this local alternative, $\omega_A(\gamma) \sim \tilde{\omega}(\gamma) \sim c_\gamma p^{1/2} n^{-\gamma/2}$, where c_γ is some constant. Next, we study $\{\mu_A(\gamma) - \mu_0(\gamma)\}/\sigma_A(\gamma)$, which determines the power. As $n, p \to \infty$,

$$\begin{aligned} \frac{\psi_A(\gamma) - \tilde{\psi}(\gamma)}{\omega_A(\gamma)} &\sim o(rp^{1/2-\eta} n^{-1/2}), && \gamma \text{ is odd}, \\ \frac{\psi_A(\gamma) - \tilde{\psi}(\gamma)}{\omega_A(\gamma)} &\sim \frac{\gamma!}{c_\gamma((\gamma-2)/2)!2^{\gamma/2}} rp^{1/2-\eta}, && \gamma \text{ is even}. \end{aligned}$$

These results show that if $rp^{1/2-\eta} < \infty$, the asymptotic power of iSPU(γ) does not converge to 1. Thus we discuss the local alternative when $r \to 0$ and $rp^{1/2-\eta} \to \infty$. Then the asymptotic power of iSPU with even γ goes to 1 while iSPU with odd γ does not. In other words, under the considered alternative here, iSPU with even γ is more powerful than iSPU with odd γ. Therefore, we only need to focus on iSPU with even γs and compare their power. To find which even γ yields an asymptotically more powerful test, we need to find which γ maximizes $\gamma!/\big(c_\gamma((\gamma-2)/2)!2^{\gamma/2}\big)$. We first consider the situation where $\tilde{\sigma}_{ij} = 0$ for $i \neq j$. In this case

$$\frac{\gamma!}{c_\gamma((\gamma-2)/2)2^{\gamma/2}} = \frac{\gamma!}{((\gamma-2)/2)!2^{\gamma/2}\sqrt{(2\gamma)!/(\gamma!2^\gamma) - (\gamma!)^2/([(\gamma/2)!]^2 2^\gamma)}},$$

which has maximum value $\sqrt{2}$ at $\gamma = 2$. More generally, under the situation where $\tilde{\sigma}_{ij} \geq 0$, a similar calculation yields that iSPU(2) is asymptotically more powerful than iSPU test with other γ. This completes the proof. ■

Appendix E. Supplementary Tables and Figures

Table S1: Empirical Type I error rates of various tests under $G \times E$ interaction simulations with $n = 2000$, $p = 200$ and various $q_1 = q_2$. n, p, q_1, and q_2 stand for the sample size, number of coefficients in $G \times E$ interaction effects, number of positive coefficients in main genetics effects, and number of negative coefficients in main genetics effects, respectively. * Conservative Type I error rates.

$q_1 = q_2$	2	5	7	10	20	30
GESAT	0.098	0.105	0.108	0.094	0.095	0.095
aiSPU(Oracle)	0.060	0.053	0.067	0.056	0.052	0.055
aiSPU(Lasso)	0.054	0.045	0.056	0.046	0.031	0.030*
aiSPU(Ridge)	0.052	0.044	0.058	0.044	0.037	0.029*
aiSPU(TLP)	0.058	0.056	0.067	0.059	0.056	0.064
aiSPU(Full)	0.085	0.104	0.107	0.097	0.084	0.093

Table S2: Empirical Type I error rates of various tests in rare variants simulations with $n = 2000$, $q_1 = 2$, $q_2 = 0$, and various p. n, p, q_1, and q_2 stand for the sample size, number of terms in $G \times E$ interaction, number of the positive genetic main effects, and number of the negative genetic main effects, respectively. * Inflated Type I error rates.

p	25	50	70	100	200	300	400	500
iSKAT	0.050	0.077	0.079	0.114*	0.229*	0.560*	0.909*	0.999*
MiSTi	0.051	0.060	0.085	0.088	0.201*	0.514*	0.881*	0.995*
Full	0.043	0.054	0.080	0.089	0.197*	0.557*	0.953*	1.000*
aiSPU(Oracle)	0.037	0.041	0.066	0.066	0.048	0.060	0.060	0.049
aiSPU(TLP)	0.043	0.039	0.060	0.064	0.043	0.053	0.058	0.049

Table S3: Empirical Type I error rates (in percentage) of various tests in rare variants simulations with $n = 2000$, $p = 300$ and various $q_1 = q_2$. n, p, q_1, and q_2 stand for the sample size, number of terms in $G \times E$ interaction, number of the positive genetic main effects, and number of the negative genetic main effects, respectively. * Inflated Type I error rates.

$q_1 = q_2$	2	5	7	10	20	30	50
iSKAT	0.535*	0.534*	0.53*	0.545*	0.550*	0.573*	0.573*
MiSTi	0.508*	0.504*	0.502*	0.508*	0.509*	0.503*	0.536*
Full	0.538*	0.539*	0.535*	0.555*	0.550*	0.573*	0.602*
aiSPU(Oracle)	0.046	0.041	0.052	0.051	0.061	0.071	0.070
aiSPU(TLP)	0.047	0.042	0.054	0.047	0.058	0.071	0.061

Table S4: Empirical Type I errors and power (in percentage) of various tests under $G \times E$ interactions with $p = 1000$ and $n = 200$. Zero signal strength $c = 0$ represents Type I errors, while $c \neq 0$ represents powers. The sparsity level was $s = 0.001$, leading to 1 non-zero elements in β. The results outside and inside parentheses were calculated from parametric bootstrap- and asymptotics-based methods, respectively.

c	0	1	2	3	4	5
iSPU(1)	4.9 (4.7)	5.7 (5.5)	4.5 (4.5)	4.9 (4.6)	6.2 (5.7)	5.9 (5.8)
iSPU(2)	2.7 (5.4)	2.3 (5.8)	6.3 (9.9)	11.6 (16.8)	20.6 (25.9)	26.9 (31.9)
iSPU(3)	5.8 (5.6)	6.8 (6.4)	15.6 (14.9)	31.7 (31.1)	42.7 (42.4)	52.4 (52.5)
iSPU(4)	3.1 (4.1)	5.9 (7.1)	23.4 (24.4)	42.6 (43.7)	54.4 (55.3)	61.4 (62.9)
iSPU(5)	5.8 (4.9)	9.1 (8.2)	29 (28.1)	47.6 (46.5)	59.8 (58.5)	67.6 (67)
iSPU(6)	3.8 (3.5)	8.9 (8)	30.8 (28.9)	51 (49.7)	60.1 (59.4)	69.7 (69)
iSPU(∞)	9 (7.6)	15.1 (13.1)	43.6 (40.8)	63.2 (61.8)	70.1 (69.3)	76.6 (75.9)
aiSPU	5.8 (6.3)	10 (10.6)	37.5 (38.6)	58.5 (58.3)	67.6 (68.1)	74.6 (74.8)

Table S5: Empirical Type I errors and power (in percentage) of various tests under $G \times E$ interactions with $p = 1000$ and $n = 200$. Zero signal strength $c = 0$ represents Type I errors, while $c \neq 0$ represents powers. The sparsity level was $s = 0.01$, leading to 10 non-zero elements in β. The results outside and inside parentheses were calculated from parametric bootstrap- and asymptotics-based methods, respectively.

c	0	1	2	3	4	5
iSPU(1)	4.8 (4.7)	4.1 (3.9)	5.1 (4.9)	5 (4.9)	5.7 (5.4)	5.7 (5.7)
iSPU(2)	2.6 (5.3)	11.7 (17.7)	40.9 (44.3)	62.7 (65.3)	73.1 (73.9)	78.5 (78.5)
iSPU(3)	5.9 (5.7)	9.2 (8.5)	28.8 (28)	44.9 (43)	50.3 (49.5)	55 (53.3)
iSPU(4)	3 (4)	25.4 (26.7)	82.5 (82.9)	95.1 (95.6)	98.2 (98.2)	99.1 (99.2)
iSPU(5)	5.9 (5)	19 (18.1)	64.2 (62)	79.3 (78.2)	84.6 (83.9)	86.9 (86.2)
iSPU(6)	3.7 (3.3)	30.1 (27.7)	89.3 (87.9)	97.5 (97.3)	98.9 (98.9)	99.6 (99.3)
iSPU(∞)	9 (7.5)	32.4 (28.7)	91.7 (89.1)	98.7 (98.3)	99.5 (99.5)	99.9 (99.9)
aiSPU	5.8 (6.2)	27.2 (29.7)	89.3 (89.4)	98.1 (98.3)	99.4 (99.4)	99.8 (99.8)

Table S6: Empirical Type I errors and power (in percentage) of various tests under $G \times E$ interactions with $p = 1000$ and $n = 200$. Zero signal strength $c = 0$ represents Type I errors, while $c \neq 0$ represents powers. The sparsity level was $s = 0.05$, leading to 50 non-zero elements in β. The results outside and inside parentheses were calculated from parametric bootstrap- and asymptotics-based methods, respectively.

c	0	1	2	3	4	5
iSPU(1)	5.8 (5.5)	4.8 (5.5)	6.2 (5.4)	6.1 (6)	7.3 (6.9)	6.9 (7.3)
iSPU(2)	2.3 (5.4)	45.3 (48.7)	75.4 (76.6)	84.8 (83.6)	86.2 (85.6)	86.9 (86.1)
iSPU(3)	5.4 (5.2)	11.4 (11.6)	17.8 (15.8)	19.8 (18.7)	20.3 (19.2)	21.8 (19.4)
iSPU(4)	2.7 (4.1)	56.7 (55.9)	88.5 (86.8)	93.7 (91.8)	95 (93.8)	95.4 (94.8)
iSPU(5)	6.1 (5)	25 (22.4)	37 (34.5)	40.6 (37.4)	43.6 (40.8)	45.5 (41.8)
iSPU(6)	4.1 (3.9)	53.7 (48.6)	83.7 (79.8)	90.8 (88.2)	91.5 (88.6)	92.3 (90.6)
iSPU(∞)	8.5 (7.3)	34.2 (27.5)	61.7 (52.1)	69.3 (59)	75 (63.4)	75.4 (64.1)
aiSPU	5.7 (6.5)	46.4 (46.9)	78.5 (78.9)	86.7 (86.2)	89.2 (88.2)	90.2 (89.8)

Table S7: Empirical Type I errors and power (in percentage) of various tests under high-dimensional linear models simulations. Zero signal strength $c = 0$ represents Type I errors, while $c \neq 0$ represents powers. Zero signal strength $c = 0$ represents Type I errors, while $c \neq 0$ represents powers. The sparsity level was $s = 0.2$, leading to 200 non-zero elements in β. The results outside and inside parentheses were calculated from parametric bootstrap- and asymptotics-based methods, respectively.

c	0	0.5	1	1.5	2
iSPU(1)	5.8 (5.5)	5.9 (5.1)	5.6 (5.5)	6.5 (6.1)	5.7 (5.9)
iSPU(2)	2.3 (5.4)	47.6 (51.6)	76.5 (75.5)	83 (81.1)	85.4 (85.3)
iSPU(3)	5.4 (5.2)	10.2 (9.1)	11.5 (10.5)	13.2 (11.5)	12.7 (10.8)
iSPU(4)	2.7 (4.1)	43.1 (42)	70.9 (67.2)	78.6 (71.7)	81 (77.1)
iSPU(5)	6.1 (5)	12.5 (11)	16.3 (14.4)	18.1 (15.7)	18.4 (16.5)
iSPU(6)	4.1 (3.9)	30.8 (25.2)	52.8 (43.1)	59.4 (52.6)	62.3 (53.8)
iSPU(Inf)	8.5 (7.4)	16.6 (11.7)	23.3 (15.5)	26.5 (19.4)	27.1 (19.4)
aiSPU	5.7 (6.5)	32.9 (34.4)	61.2 (58.6)	68.5 (65)	73.3 (71.4)

Table S8: Empirical Type I errors and power (in percentage) of various tests under $G \times E$ interactions with $p = 1000$ and $n = 200$. Zero signal strength $c = 0$ represents Type I errors, while $c \neq 0$ represents powers. The sparsity level was $s = 0.2$, leading to 200 non-zero elements in β. The informative variables in β was generated from a uniform distribution $U(0, c)$. The results outside and inside parentheses were calculated from parametric bootstrap- and asymptotics-based methods, respectively.

c	0	0.01	0.05	0.1	0.3	0.5
iSPU(1)	5.8 (5.5)	5.8 (5.7)	19.9 (19.2)	54.9 (53.1)	98.3 (98.4)	100 (99.9)
iSPU(2)	2.3 (5.4)	2.1 (5.8)	2.3 (7.6)	7.1 (12.7)	47.8 (52.8)	67.8 (70.2)
iSPU(3)	5.4 (5.2)	4.8 (5.1)	7.5 (7.9)	25.3 (23.8)	87.8 (86.2)	97.1 (96.5)
iSPU(4)	2.7 (4.2)	2.4 (3.1)	3.1 (4.3)	6.3 (7.5)	41.8 (39.1)	62.5 (56.3)
iSPU(5)	6.1 (5)	6.5 (5.4)	6.3 (5.4)	10 (8.9)	52.6 (48.3)	74.3 (70.1)
iSPU(6)	4.2 (4)	4.5 (4.1)	4.1 (3.6)	5.9 (5.2)	27.2 (22.2)	38.8 (32.4)
iSPU(∞)	8.5 (7.4)	9.2 (7.7)	10.5 (9.2)	10.2 (8.2)	21.2 (15.1)	21.9 (15.1)
aiSPU	5.7 (6.6)	6.2 (6.9)	13.8 (11.8)	34.3 (31.5)	96.3 (93.9)	99.3 (98.5)

Table S9: Empirical Type I errors and power (in percentage) of various tests under $G \times E$ interactions with $p = 1000$ and $n = 200$. Zero signal strength $c = 0$ represents Type I errors, while $c \neq 0$ represents powers. The informative variables in β was selected to be those with main effects and generated from a uniform distribution $U(-c, c)$. The results outside and inside parentheses were calculated from parametric bootstrap- and asymptotics-based methods, respectively.

c	0	1	2	3	4	5
iSPU(1)	4.9 (4.8)	5.6 (5.2)	6.4 (6)	6.1 (5.9)	6 (5.3)	6.1 (5.7)
iSPU(2)	2.7 (5.3)	6.9 (10.3)	19.4 (24.8)	37 (42)	51.2 (53.4)	58.5 (62)
iSPU(3)	5.9 (5.6)	6.3 (6.1)	27.3 (26.6)	50.4 (50.1)	62.3 (61.5)	69.7 (68.2)
iSPU(4)	3.1 (4)	12.4 (14.3)	58.1 (59.3)	85 (85.8)	94.5 (93.9)	97.2 (97.2)
iSPU(5)	5.9 (5)	13.6 (12.5)	58.7 (57.1)	80.7 (79.7)	88.6 (87.1)	92.6 (92.3)
iSPU(6)	3.8 (3.3)	18 (16.6)	70.6 (68.5)	91.8 (91.5)	96.9 (96.3)	98.6 (98.2)
iSPU(Inf)	9 (7.6)	24.1 (21.5)	82.6 (80.5)	96.5 (95.5)	98.7 (98.4)	99.7 (99.4)
aiSPU	5.8 (6.3)	17.3 (19.7)	76.2 (77.1)	94.5 (94.7)	98.4 (98)	99.1 (99.3)

Table S10: Empirical Type I errors and power (in percentage) of various tests under high-dimensional linear models simulations. Zero signal strength $c = 0$ represents Type I errors, while $c \neq 0$ represents powers. The sparsity level was $s = 0.001$, leading to 1 non-zero element in β. The results outside and inside parentheses were calculated from parametric bootstrap- and asymptotics-based methods, respectively.

c	0	0.3	0.5	0.7	1
iSPU(1)	5.6 (5.4)	6.7 (6.1)	6.6 (6.3)	7.5 (7.2)	8.9 (8.6)
iSPU(2)	3.6 (3.3)	4.2 (5.7)	6.6 (8.2)	15.3 (18.9)	32.2 (38.7)
iSPU(3)	5 (4.8)	6.4 (5.6)	14.6 (13.5)	41.7 (40.1)	64.2 (63.1)
iSPU(4)	3.8 (1.8)	9.1 (7.5)	29.5 (26.4)	54.6 (52.1)	71.3 (71.1)
iSPU(5)	5.5 (3.5)	16.5 (12.8)	36.1 (32.7)	57.7 (54.5)	72.1 (70.6)
iSPU(6)	4.9 (2.2)	18.2 (13.3)	38.8 (33.8)	61.9 (58.2)	73.7 (71.9)
iSPU(∞)	3.5 (4.6)	16.1 (18.3)	36.5 (38.7)	61.4 (61.9)	74.1 (74.5)
aiSPU	5.3 (4.1)	16.6 (16.5)	38.5 (38.3)	61.4 (60.1)	73.7 (73.7)

Table S11: Empirical Type I errors and power (in percentage) of various tests under high-dimensional linear models simulations. Zero signal strength $c = 0$ represents Type I errors, while $c \neq 0$ represents powers. The sparsity level was $s = 0.01$, leading to 10 non-zero elements in β. The results outside and inside parentheses were calculated from parametric bootstrap- and asymptotics-based methods, respectively.

c	0	0.1	0.2	0.3	0.4	0.5
iSPU(1)	5.2 (5.5)	7.1 (6)	6 (5.3)	7.7 (8.4)	9 (8.4)	8.7 (7.3)
iSPU(2)	4.1 (4.3)	4 (6.2)	10.4 (13.3)	24.4 (29.8)	42.6 (49.4)	52.9 (64.7)
iSPU(3)	5.1 (4.6)	5.7 (4.5)	10.2 (9)	21.2 (18.5)	35.7 (33.6)	47.3 (44)
iSPU(4)	5.6 (2.1)	5.9 (3.9)	19.5 (16.2)	55.3 (52.4)	84.4 (83.2)	95.3 (95.8)
iSPU(5)	5.2 (3.3)	5.6 (4)	18.3 (13)	40.8 (36.8)	68.4 (64.1)	81.4 (79.8)
iSPU(6)	5.9 (2.3)	6.6 (3.4)	24.7 (16.5)	67.9 (60.5)	93.9 (90)	98.8 (98.1)
iSPU(∞)	3.5 (4.6)	4.5 (5.2)	12.7 (16.2)	48.5 (52.8)	81.1 (83.6)	94.1 (96)
aiSPU	5.2 (4.5)	6.6 (5.4)	17.7 (16.9)	58.3 (55.7)	88.1 (86.7)	96.2 (96.5)

Table S12: Empirical Type I errors and power (in percentage) of various tests under high-dimensional linear models simulations. Zero signal strength $c = 0$ represents Type I errors, while $c \neq 0$ represents powers. The sparsity level was $s = 0.2$, leading to 200 non-zero elements in β. The results outside and inside parentheses were calculated from parametric bootstrap- and asymptotics-based methods, respectively.

c	0	0.05	0.07	0.1	0.2	0.3
iSPU(1)	6.3 (6)	5.5 (6.1)	7.4 (5.7)	3.4 (4)	5.3 (4.3)	6.4 (6.4)
iSPU(2)	2.8 (3.5)	13.9 (18.6)	23.3 (29.5)	42.7 (51.7)	85 (87.9)	95.7 (96.4)
iSPU(3)	5.4 (5.4)	6.1 (4.9)	6.7 (6.2)	8.4 (7.2)	12.1 (10.1)	15.7 (13.6)
iSPU(4)	3.7 (1.7)	12.4 (9.1)	22.8 (19.3)	41.7 (40.2)	84.5 (82.1)	95 (92.9)
iSPU(5)	6.1 (3.9)	7 (4)	6.7 (3.5)	8.7 (4.7)	23.2 (16.4)	27.1 (18.6)
iSPU(6)	4.7 (2)	9.9 (4.6)	14.9 (9.2)	33 (23.4)	77.8 (67.1)	85.7 (77.1)
iSPU(∞)	2.8 (4.3)	4.6 (5.5)	5.2 (6.4)	5.9 (7.8)	18.8 (20.8)	29.3 (27.9)
aiSPU	4.7 (3.7)	8.9 (8.9)	15.3 (15.3)	27.7 (33)	73.9 (79.2)	87.1 (89.3)

Table S13: Empirical Type I errors and power (in percentage) of various tests under high-dimensional linear models simulations. Zero signal strength $c = 0$ represents Type I errors, while $c \neq 0$ represents powers. The sparsity level was $s = 0.2$, leading to 200 non-zero elements in β. We generated informative variables in β from a uniform distribution $U(0, c)$. The results outside and inside parentheses were calculated from parametric bootstrap- and asymptotics-based methods, respectively.

c	0	0.01	0.02	0.03	0.04	0.05
iSPU(1)	6.2 (6.1)	24.6 (24)	59.6 (58.2)	86.2 (85.9)	96.1 (95.8)	98.7 (98.7)
iSPU(2)	3 (3.3)	4.5 (6.8)	18 (21.5)	40.6 (47)	65.8 (74.7)	82.2 (88.1)
iSPU(3)	5.2 (5.2)	19.3 (16.7)	51.7 (48.8)	80.4 (78.8)	93.8 (92.6)	98.7 (98)
iSPU(4)	3.6 (1.5)	5.6 (3.2)	16.6 (11.3)	38.2 (33.9)	59.8 (58.9)	78.2 (78.9)
iSPU(5)	5.5 (3.4)	12.1 (7.3)	32.1 (22.8)	60.1 (50.9)	80.4 (74.7)	92.4 (86.8)
iSPU(6)	4.4 (1.8)	7 (2.6)	13.3 (7.7)	26.7 (18.2)	46.7 (32.1)	63 (50.5)
iSPU(∞)	3.2 (4.1)	4.5 (6)	5.2 (7)	7.6 (8.8)	8.9 (11)	10.2 (13.9)
aiSPU	5.1 (3.9)	13.4 (11.4)	46.3 (41.8)	80.9 (77.4)	93.8 (92.9)	97.7 (97)

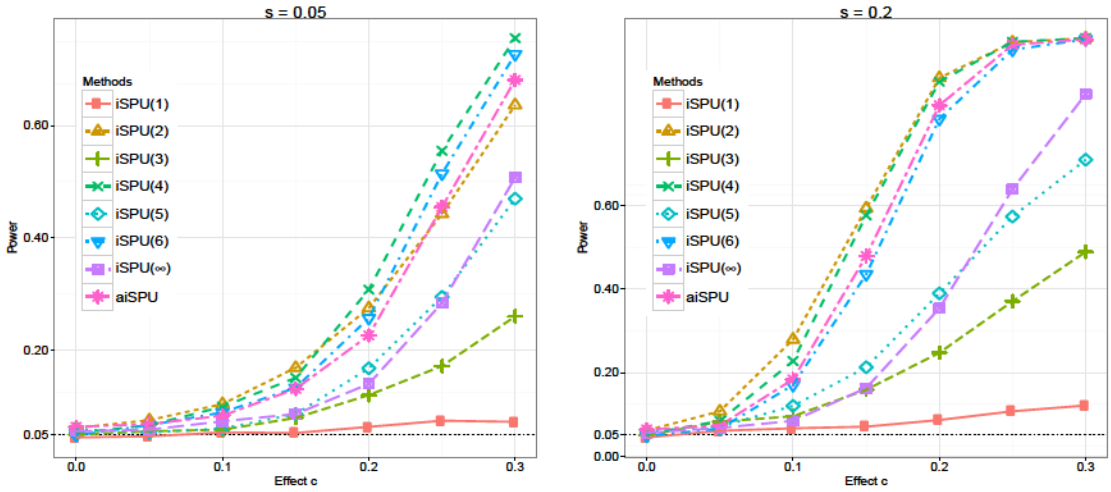

Figure S1: Power comparison for different methods under $G \times E$ interaction simulations with $n = 2000$, $p = 300$, and $q_1 = q_2 = 20$. n, p, q_1, and q_2 stand for the sample size, number of terms in $G \times E$ interaction, number of the positive genetic main effects, and number of the negative genetic main effects, respectively. All tests were based on TLP. We varied the sparsity level s.

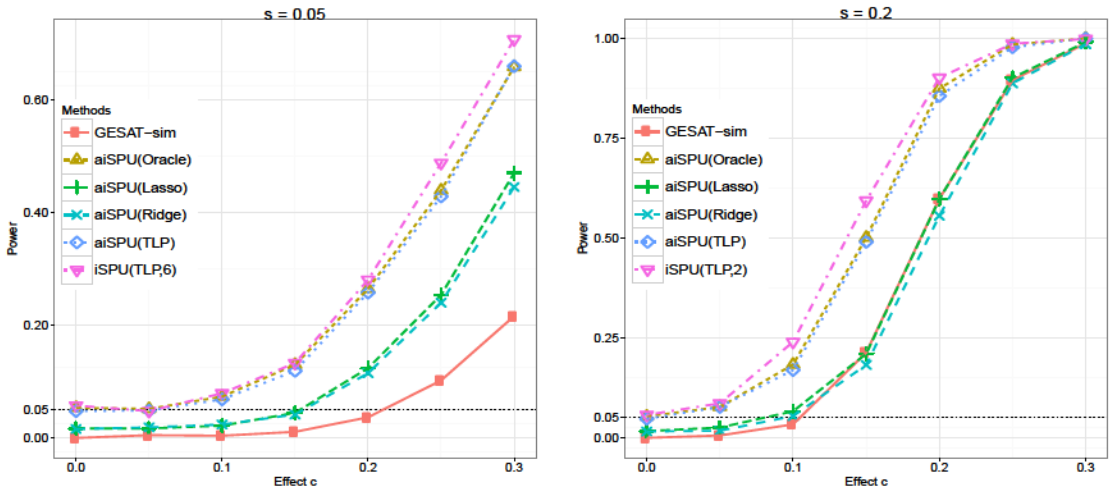

Figure S2: Power comparison for different methods under $G \times E$ interaction simulations with $n = 2000$, $p = 300$, and $q_1 = q_2 = 20$. n, p, q_1, and q_2 stand for the sample size, number of terms in $G \times E$ interaction, number of the positive genetic main effects, and number of the negative genetic main effects, respectively. The SNPs were correlated ($\rho = 0.3$). We varied the sparsity level s.

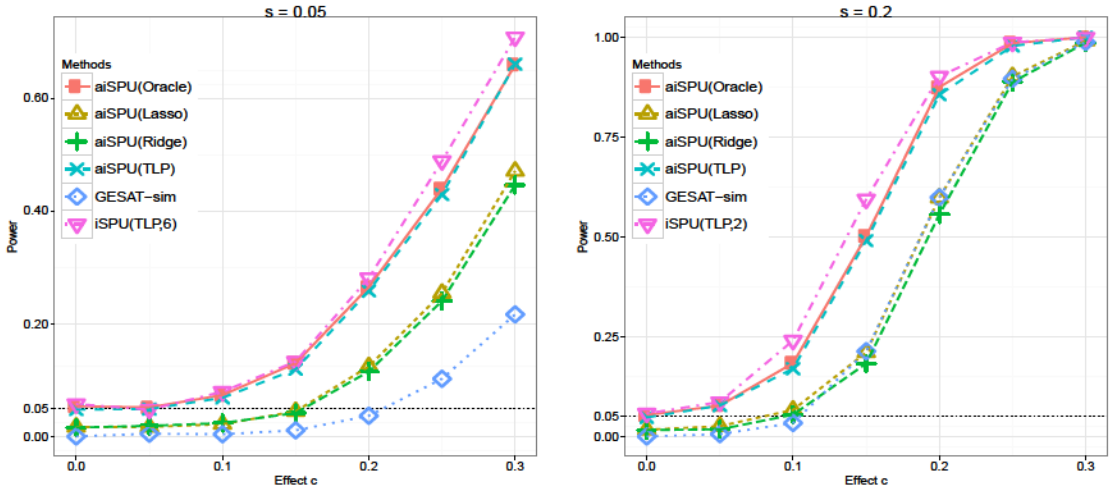

Figure S3: Power comparison for different methods under $G \times E$ interaction simulation with $n = 2000$, $p = 300$, and $q_1 = q_2 = 50$. n, p, q_1, and q_2 stand for the sample size, number of terms in $G \times E$ interaction, number of the positive genetic main effects, and number of the negative genetic main effects, respectively. The SNPs were correlated ($\rho = 0.3$). We varied the sparsity level s.

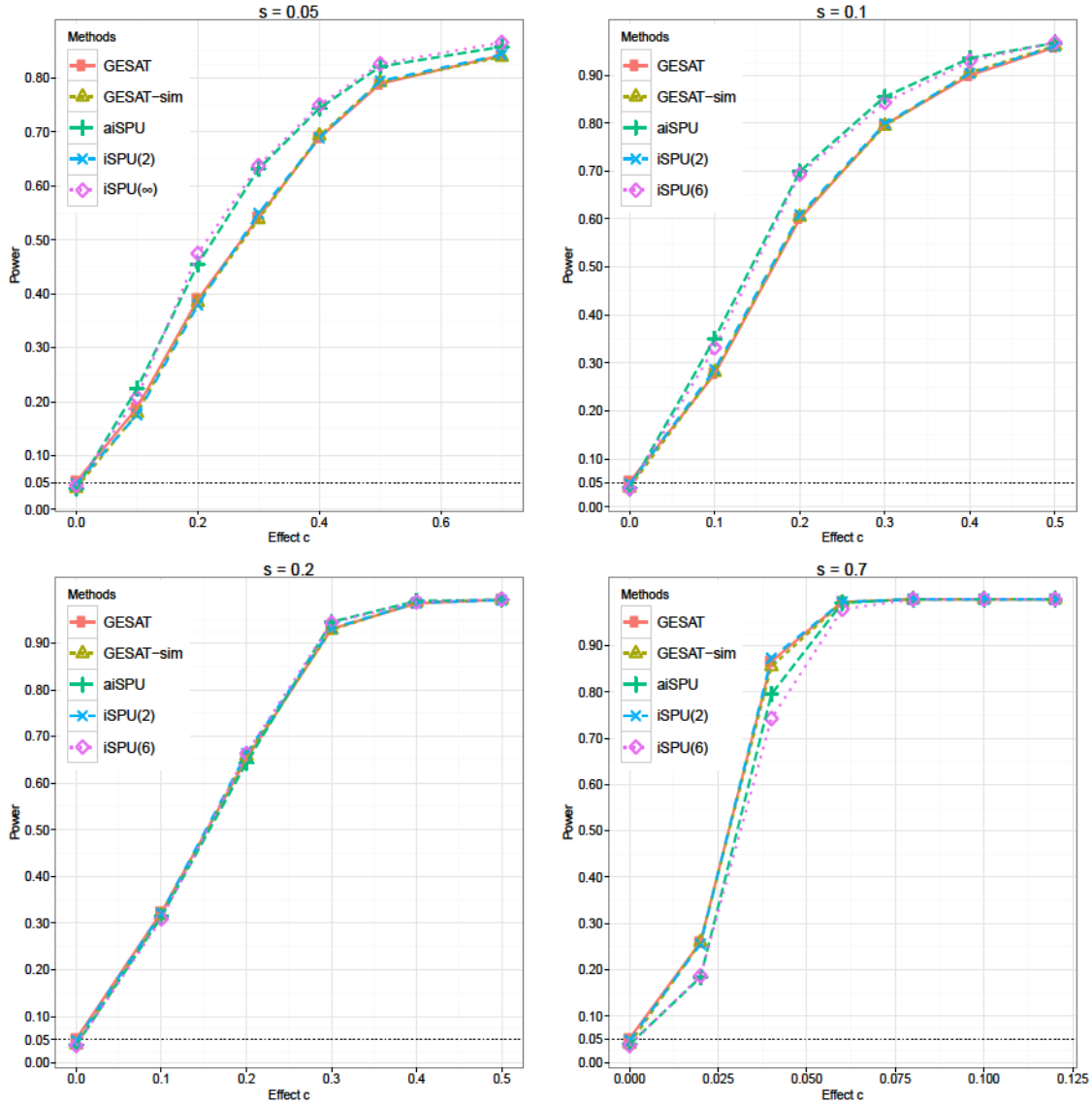

Figure S4: Power comparison for different methods under $G \times E$ interaction simulations with $n = 2000$, $q_1 = 2$, $q_2 = 0$, and $p = 25$. n, p, q_1, and q_2 stand for the sample size, number of coefficients in $G \times E$ interaction effects, number of positive coefficients in main genetics effects, and number of negative coefficients in main genetics effects, respectively. We varied the sparsity level s.

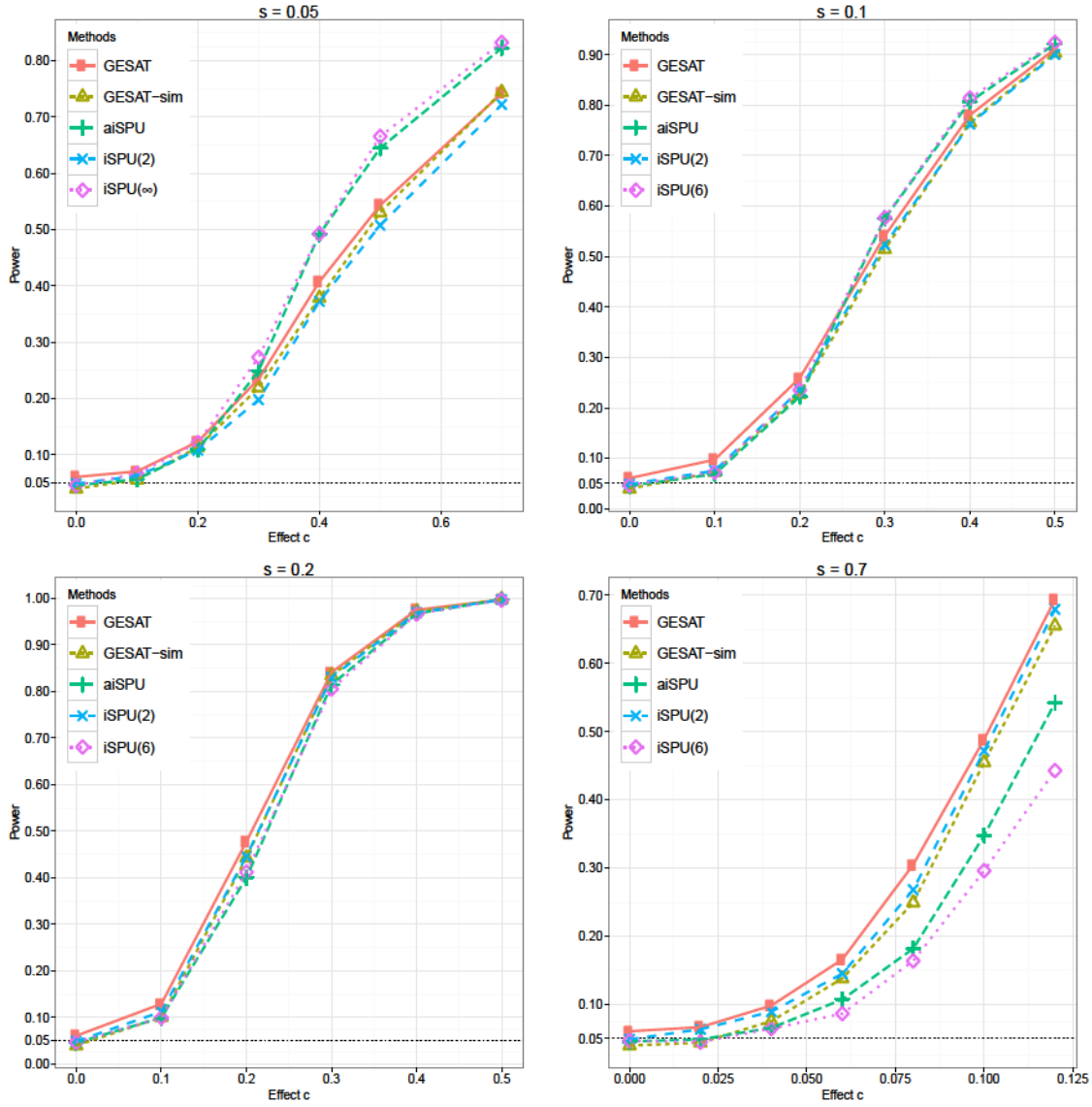

Figure S5: Power comparison for different methods under $G \times E$ interaction simulations with $n = 2000$, $q_1 = 2$, $q_2 = 0$, and $p = 50$. n, p, q_1, and q_2 stand for the sample size, number of coefficients in $G \times E$ interaction effects, number of positive coefficients in main genetics effects, and number of negative coefficients in main genetics effects, respectively. We varied the sparsity level s.

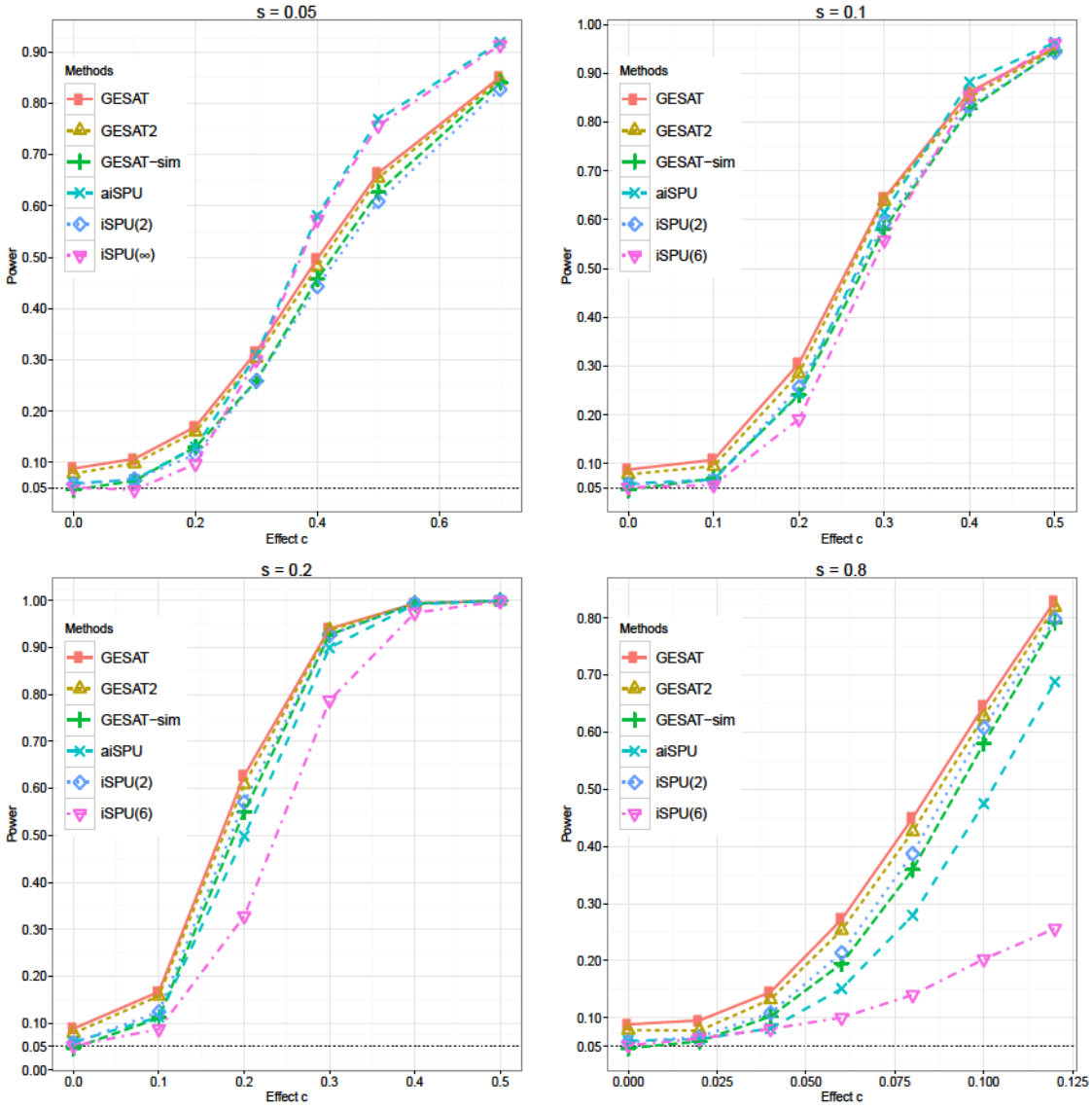

Figure S6: Power comparison for different methods under $G \times E$ interaction simulations with $n = 2000$, $q_1 = 2$, $q_2 = 0$, and $p = 75$. n, p, q_1, and q_2 stand for the sample size, number of coefficients in $G \times E$ interaction effects, number of positive coefficients in main genetics effects, and number of negative coefficients in main genetics effects, respectively. GESAT2 stands for the GESAT with much a larger searching region (from 1×10^{-6} to 44.7 (i.e. $\sqrt{n}$)) for tuning parameter λ. We varied the sparsity level s.

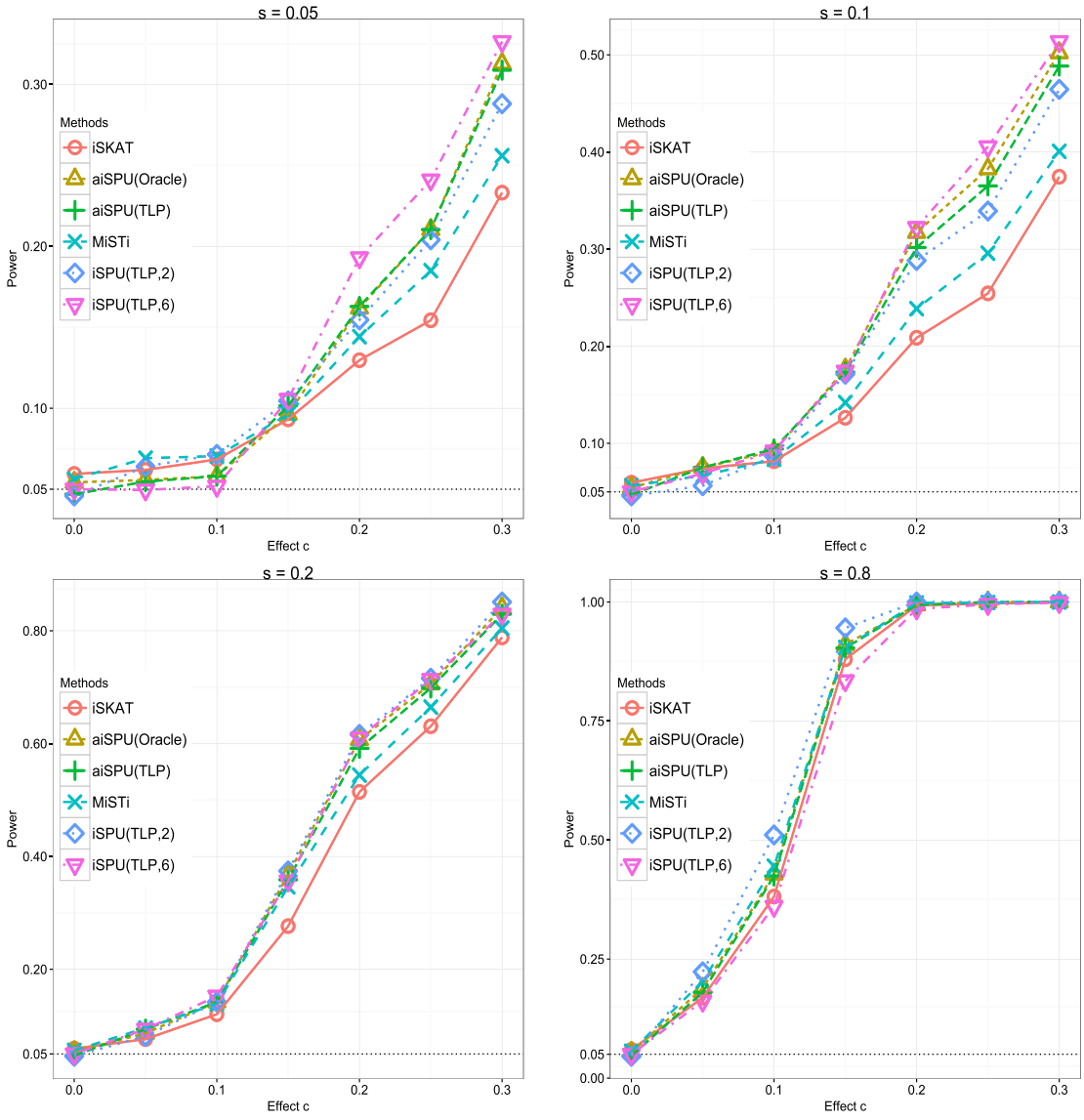

Figure S7: Power comparison for different methods in rare variants simulations with $n = 2000$, $q_1 = 2$, $q_2 = 0$, and $p = 25$. n, p, q_1, and q_2 stand for the sample size, number of terms in $G \times E$ interaction, number of the positive genetic main effects, and number of the negative genetic main effects, respectively. SNPs were generated with MAFs ranging from 0.005 to 0.05. We varied the sparsity level s.

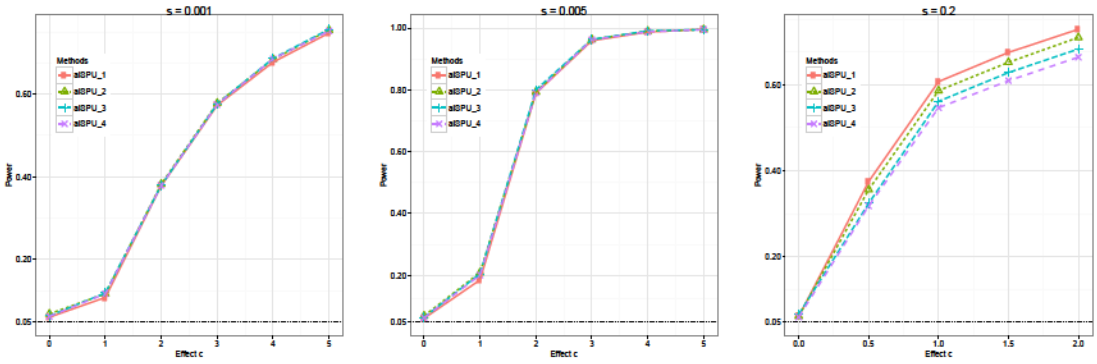

Figure S8: Empirical power of aiSPU with different Γ sets for $G \times E$ interactions with $n = 200$, $p = 1000$. aiSPU_1, aiSPU_2, aiSPU_3, aiSPU_4 represent aiSPU with $\Gamma_1 = \{1, 2, 3, 4; \infty\}$, $\Gamma_2 = \{1, 2, \ldots, 6, \infty\}$, $\Gamma_3 = \{1, \ldots, 8, \infty\}$, and $\Gamma_4 = \{1, 2, \ldots, 10, \infty\}$, respectively. We varied the sparsity level s.

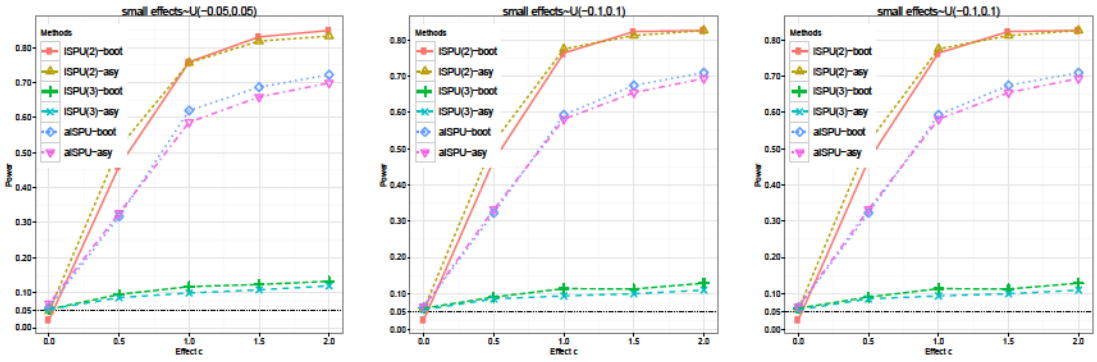

Figure S9: Empirical power of aiSPU under $G \times E$ interaction with $n = 200$, $p = 1000$, and sparsity level $s = 0.2$. We randomly selected 100 variables in $\mathbb{Z}$ and set the effect size followed a uniform distribution. -boot and -asy stand for the results based on bootstrap and asymptotics, respectively.

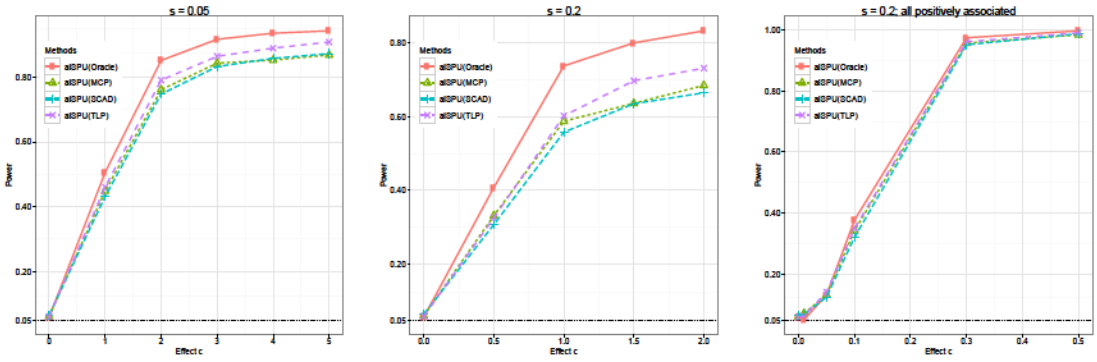

Figure S10: Empirical power of aiSPU with different non-convex penalities under $G \times E$ interaction with $n = 200$, $p = 1000$, and varied sparsity level s. For a fair comparison, all the results were based on the parametric bootstrap.

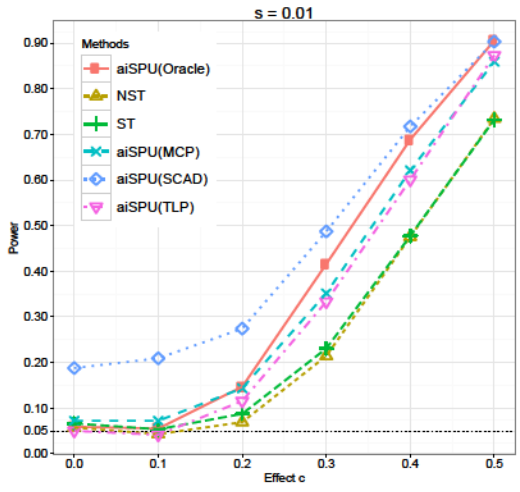

Figure S11: Empirical power of aiSPU with different non-convex penalities under high-dimensional linear models with $n = 200$, $p = 1000$, and sparsity level $s = 0.01$. For a fair comparison, all the aiSPU results were based on the parametric bootstrap.

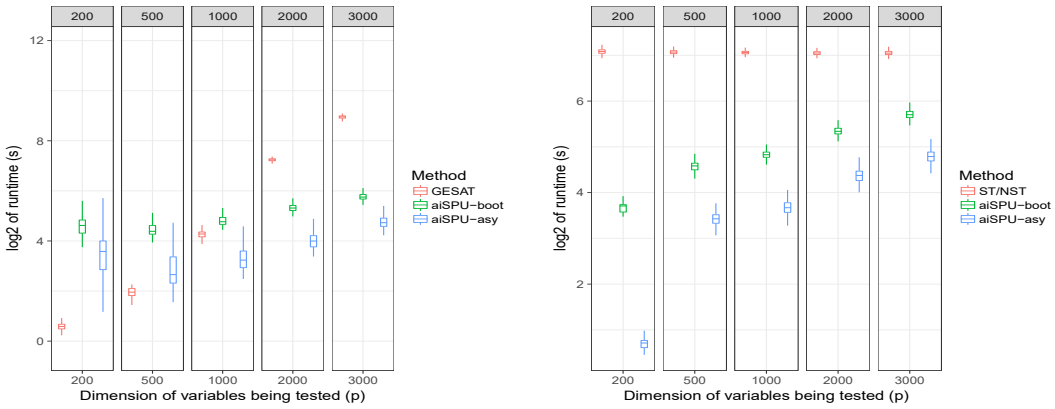

Figure S12: Computational time comparison under both $G \times E$ interaction models (left panel) and linear models (right panel) with $n = 200$, sparsity level $s = 0$. In Zhang and Cheng (2017), ST and NST have been calculated simultaneously; ST/NST stands for the runtime for ST plus NST. We varied the number of variables being tested, p.

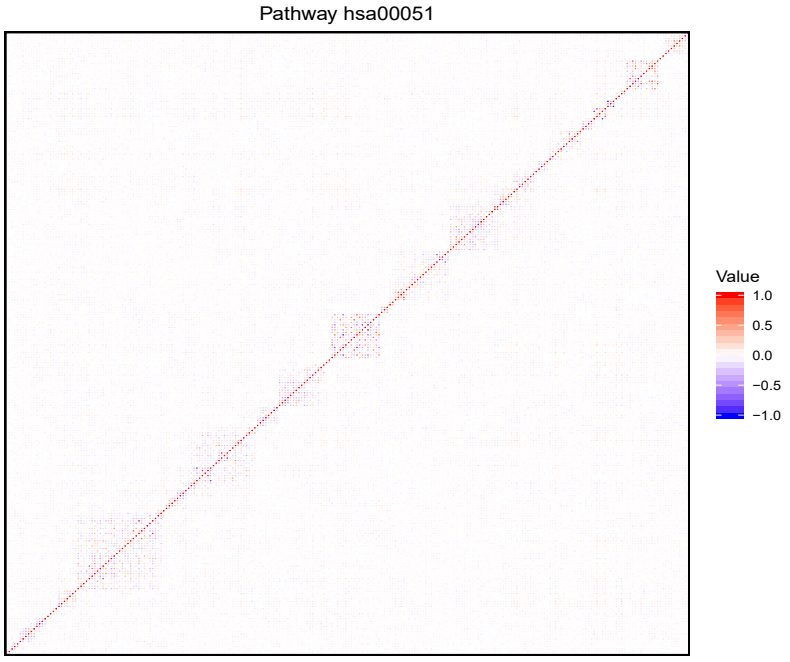

Figure S13: The correlation heatmap for SNPs used in pathway hsa00051. Pathway hsa00051 is the significant pathway identified by aiSPU.

References

Andre Altmann, Lu Tian, Victor W Henderson, and Michael D Greicius. Sex modifies the apoe-related risk of developing alzheimer disease. *Annals of Neurology*, 75(4):563–573, 2014.

Peter J Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36(1):199–227, 2008.

Annalisa Buniello, Jacqueline A L MacArthur, Maria Cerezo, Laura W Harris, James Hayhurst, Cinzia Malangone, Aoife McMahon, Joannella Morales, Edward Mountjoy, Elliot Sollis, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, 47(D1): D1005–D1012, 2018.

T Tony Cai, Weidong Liu, and Yin Xia. Two-sample test of high dimensional means under dependence. *Journal of the Royal Statistical Society: Series B*, 76(2):349–372, 2014.

Tony Cai and Weidong Liu. Adaptive thresholding for sparse covariance matrix estimation. *Journal of the American Statistical Association*, 106(494):672–684, 2011.

Heather J Cordell. Detecting gene–gene interactions that underlie human diseases. *Nature Reviews Genetics*, 10(6):392–404, 2009.

Sayantan Das, Lukas Forer, Sebastian Schönherr, Carlo Sidore, Adam E Locke, Alan Kwong, Scott I Vrieze, Emily Y Chew, Shawn Levy, Matt McGue, et al. Next-generation genotype imputation service and methods. *Nature Genetics*, 48(10):1284–1287, 2016.

Ruben Dezeure, Peter Bühlmann, and Cun-Hui Zhang. High-dimensional simultaneous inference with the bootstrap. *TEST*, 26(4):685–719, 2017.

Noureddine El Karoui and Elizabeth Purdom. Can we trust the bootstrap in high-dimensions? The case of linear models. *The Journal of Machine Learning Research*, 19(1):170–235, 2018.

Jianqing Fan. Test of significance based on wavelet thresholding and neyman's truncation. *Journal of the American Statistical Association*, 91(434):674–688, 1996.

Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.

Jianqing Fan and Rui Song. Sure independence screening in generalized linear models with np-dimensionality. *The Annals of Statistics*, 38(6):3567–3604, 2010.

Jianqing Fan, Yuan Liao, and Han Liu. An overview of the estimation of large covariance and precision matrices. *The Econometrics Journal*, 19(1):C1–C32, 2016.

Zhe Fei and Yi Li. Estimation and inference for high dimensional generalized linear models: A splitting and smoothing approach. *arXiv preprint arXiv:1903.04408*, 2019.

Eric R Gamazon, Heather E Wheeler, Kaanan P Shah, Sahar V Mozaffari, Keston Aquino-Michaels, Robert J Carroll, Anne E Eyler, Joshua C Denny, Dan L Nicolae, Nancy J Cox, et al. A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics*, 47(9):1091–1098, 2015.

Bin Guo and Song Xi Chen. Tests for high dimensional generalized linear models. *Journal of the Royal Statistical Society: Series B*, 78(5):1079–1102, 2016.

Alexander Gusev, Arthur Ko, Huwenbo Shi, Gaurav Bhatia, Wonil Chung, Brenda WJH Penninx, Rick Jansen, Eco JC De Geus, Dorret I Boomsma, Fred A Wright, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics*, 48(3):245–252, 2016.

Xavier Guyon. *Random fields on a network: modeling, statistics, and applications*. Springer Science & Business Media, 1995.

Yinqiu He, Gongjun Xu, Chong Wu, and Wei Pan. Asymptotically independent U-statistics in high-dimensional testing. *Annals of Statistics, to appear, arXiv:1809.00411*, 2020.

Zihuai He, Bin Xu, Seunggeun Lee, and Iuliana Ionita-Laza. Unified sequence-based association tests allowing for multiple functional annotations and meta-analysis of noncoding variation in metabochip data. *The American Journal of Human Genetics*, 101(3):340–352, 2017.

David J Hunter. Gene-environment interactions in human diseases. *Nature Reviews Genetics*, 6(4):287–298, 2005.

Il'dar Abdulovich Ibragimov and IUrii Vladimirovich Linnik. Independent and stationary sequences of random variables. 1971.

Clifford R Jack, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L Whitwell, Chadwick Ward, et al. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691, 2008.

Adel Javanmard and Andrea Montanari. Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research*, 15(1):2869–2909, 2014.

Adel Javanmard and Andrea Montanari. Debiasing the lasso: Optimal sample size for gaussian designs. *The Annals of Statistics*, 46(6A):2593–2622, 2018.

Minoru Kanehisa, Susumu Goto, Miho Furumichi, Mao Tanabe, and Mika Hirakawa. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research*, 38(1):D355–D360, 2009.

Keith Knight and Wenjiang Fu. Asymptotics for lasso-type estimators. *Annals of Statistics*, 28(5):1356–1378, 2000.

S Le Cessie and JC Van Houwelingen. A goodness-of-fit test for binary regression models, based on smoothing methods. *Biometrics*, 47(4):1267–1282, 1991.

Jason D Lee, Dennis L Sun, Yuekai Sun, and Jonathan E Taylor. Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 44(3):907–927, 2016.

Chunlin Li, Xiaotong Shen, and Wei Pan. Likelihood ratio tests for a large directed acyclic graph. *Journal of the American Statistical Association*, pages 1–36, 2019.

Xinyi Lin, Seunggeun Lee, David C Christiani, and Xihong Lin. Test for interactions between a genetic marker set and environment in generalized linear models. *Biostatistics*, 14(4):667–681, 2013.

Xinyi Lin, Seunggeun Lee, Michael C Wu, Chaolong Wang, Han Chen, Zilin Li, and Xihong Lin. Test for rare variants by environment interactions in sequencing association studies. *Biometrics*, 72(1):156–164, 2016.

Xuanyao Liu, Yang I Li, and Jonathan K Pritchard. Trans effects on gene expression can drive omnigenic inheritance. *Cell*, 177(4):1022–1034, 2019.

Rong Ma, T Tony Cai, and Hongzhe Li. Global and simultaneous hypothesis testing for high-dimensional logistic regression models. *Journal of the American Statistical Association*, pages 1–15, 2020.

Yiding Ma and Peng Wei. FunSPU: A versatile and adaptive multiple functional annotation-based association test of whole-genome sequencing data. *PLoS Genetics*, 15(4):e1008081, 2019.

Stephen B Manuck and Jeanne M McCaffery. Gene-environment interaction. *Annual Review of Psychology*, 65:41–70, 2014.

Nicolai Meinshausen, Lukas Meier, and Peter Bühlmann. P-values for high-dimensional regression. *Journal of the American Statistical Association*, 104(488):1671–1681, 2009.

Colm O'Dushlaine, Lizzy Rossin, Phil H Lee, Laramie Duncan, Neelroop N Parikshak, Stephen Newhouse, Stephan Ripke, Benjamin M Neale, Shaun M Purcell, Danielle Posthuma, et al. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nature Neuroscience*, 18(2):199–209, 2015.

Wei Pan, Junghi Kim, Yiwei Zhang, Xiaotong Shen, and Peng Wei. A powerful and adaptive association test for rare variants. *Genetics*, 197(4):1081–1095, 2014.

Wei Pan, Il-Youp Kwak, and Peng Wei. A powerful pathway-based adaptive test for genetic association with common or rare variants. *The American Journal of Human Genetics*, 97(1):86–98, 2015.

Xiaotong Shen, Wei Pan, and Yunzhang Zhu. Likelihood-based selection and sharp parameter estimation. *Journal of the American Statistical Association*, 107(497):223–232, 2012.

Chengchun Shi, Rui Song, Zhao Chen, Runze Li, et al. Linear hypothesis testing for high dimensional generalized linear models. *The Annals of Statistics*, 47(5):2671–2703, 2019.

Yu-Ru Su, Chong-Zhi Di, and Li Hsu. A unified powerful set-based test for sequencing data analysis of $G \times E$ interactions. *Biostatistics*, 18(1):119–131, 2017.

Pragya Sur and Emmanuel J Candès. A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences*, 116 (29):14516–14525, 2019.

Sara Van de Geer, Peter Bühlmann, Yaacov Ritov, and Ruben Dezeure. On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3):1166–1202, 2014.

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

Martin J Wainwright. Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. *IEEE Transactions on Information Theory*, 55(12):5728–5741, 2009.

Kai Wang. Boosting the power of the sequence kernel association test by properly estimating its null distribution. *The American Journal of Human Genetics*, 99(1):104–114, 2016.

Tao Wang and Robert C Elston. Improved power by use of a weighted score test for linkage disequilibrium mapping. *The American Journal of Human Genetics*, 80(2):353–360, 2007.

Larry Wasserman and Kathryn Roeder. High dimensional variable selection. *The Annals of Statistics*, 37(5A):2178–2201, 2009.

Chong Wu, Gongjun Xu, and Wei Pan. An adaptive test on high-dimensional parameters in generalized linear models. *Statistica Sinica*, 29:2163–2186, 2019.

Gongjun Xu, Lifeng Lin, Peng Wei, and Wei Pan. An adaptive two-sample test for high-dimensional means. *Biometrika*, 103(3):609–624, 2016.

Zhiyuan Xu, Chong Wu, Peng Wei, and Wei Pan. A powerful framework for integrating eqtl and gwas summary data. *Genetics*, 207(3):893–902, 2017.

Cun-Hui Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942, 2010.

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

Xianyang Zhang and Guang Cheng. Simultaneous inference for high-dimensional linear models. *Journal of the American Statistical Association*, 112(518):757–768, 2017.

Yunzhang Zhu, Xiaotong Shen, and Wei Pan. On high-dimensional constrained maximum likelihood inference. *Journal of the American Statistical Association*, pages 1–14, 2019.