An Adaptive Test on High-dimensional Parameters in Generalized Linear Models

Chong Wu[1*], Gongjun Xu[2], Wei Pan[1*],
for the Alzheimer's Disease Neuroimaging Initiative[†]

[1]*Division of Biostatistics, University of Minnesota*
[2]*Department of Statistics, University of Michigan*

Abstract: Significance testing for high-dimensional generalized linear models (GLMs) has been increasingly needed in various applications, however, existing methods are mainly based on a sum of squares of the score vector and only powerful under certain alternative hypotheses. In practice, depending on whether the true association pattern under an alternative hypothesis is sparse or dense or between, the existing tests may or may not be powerful. In this paper, we propose an adaptive test on a high-dimensional parameter of a GLM (in the presence of a low-dimensional nuisance parameter), which can maintain high power across a wide range of scenarios. To evaluate its p-value, its asymptotic null distribution is derived. We conduct simulations to demonstrate the superior performance of the proposed test. In addition, we apply it and other existing tests to an Alzheimer's Disease Neuroimaging Initiative (ADNI)

∗ Correspondence: wuxx0845@umn.edu (C.W.), weip@biostat.umn.edu (W.P.)

† Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or 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

data set, detecting possible associations between Alzheimer's disease and some gene pathways with a large number of single nucleotide polymorphisms (SNPs). We also implemented the proposed method in R package *GLMaSPU* that is publicly available on GitHub and CRAN.

Key words and phrases: Adaptive tests, Generalized linear models, High-dimensional testing, Power

1. Introduction

Generalized linear models (GLMs; McCullagh and Nelder, 1989) have been increasingly used in high-dimensional settings due to the surge of high-dimensional data in many fields, ranging from business to genetics. One topic of intensive interest is significance testing on regression coefficients in high-dimensional GLMs. For example, genome-wide association studies (GWASs) have led to the discovery of many genetic variants, mostly single nucleotide polymorphisms (SNPs), associated with common and complex diseases. Given the number of SNPs tested in GWASs, a univariate test must meet a stringent threshold for statistical significance (with p-value $< 5 \times 10^{-8}$) and thus is often underpowered. When failing to identify any or a sufficient number of associated SNPs based on the univariate test, one may be interested in directly testing a genetic marker set with possibly a large number of SNPs to both gain statistical power and enhance biological interpretation.

In these applications, the dimension of the parameters to be tested, p, is often

close to or higher than the sample size, n. For low dimensional situations with $p \ll n$, traditional multivariate tests, such as the likelihood ratio test and the Wald test, have been widely used (McCullagh and Nelder, 1989); however, the power of both the Wald test and the likelihood ratio test tend to diminish quite rapidly as p increases (Goeman et al., 2006). These tests even break down completely when $p > n$ since the maximum likelihood estimates (MLEs) of the parameters are not uniquely determined. To deal with these difficulties, several tests for high-dimensional data have been proposed accordingly (e.g., Goeman et al., 2006, 2011; Zhong and Chen, 2011; Lan et al., 2014; Guo and Chen, 2016). In particular, Zhong and Chen (2011) proposed a modified F-test in high-dimensional linear regression models, allowing $p \to \infty$ as $n \to \infty$; Lan et al. (2014) extended the test to GLMs with a general random design matrix. Meanwhile, Goeman et al. (2006) proposed a test statistic for high-dimensional linear models and Goeman et al. (2011) derived its asymptotic distribution for a fixed p in GLMs. Guo and Chen (2016) further modified Goeman's test statistic (Goeman et al., 2011) to a simpler form and allowed both n and $p \to \infty$. In a penalized regression framework, several inference methods for a low-dimensional sub-vector of a high-dimensional regression coefficient vector have been developed (Van de Geer et al., 2014; Zhang and Zhang, 2014; Voorman et al., 2014), which however differs from the goal of testing on a high-dimensional parameter here and thus will not be further discussed.

The existing methods are mainly based on the sum-of-squares of the score vector for the parameters of interest and are usually powerful against alternative hypotheses with moderately dense signals/association patterns, where there is a relatively large proportion of associated (i.e. non-null) parameters. In contrast, if the nonzero associations are strong but sparse, the sum-of-squares-type tests lose substantial power while a test based on the supremum of the score vector is more powerful. Importantly, as to be shown in the simulation section, there are some intermediate situations in which neither type of the above tests is powerful. In practice, it is often unclear which type of tests should be applied since the underlying truth is unknown.

In this paper, we develop an adaptive test that would yield high statistical power under various high-dimensional scenarios, ranging from highly dense to highly sparse signal situations. The main idea is that, since we do not know which and how many parameters being tested are associated with the response, we first construct a class of sum of *powered* score tests such that hopefully at least one of them would be powerful for a given situation. The proposed adaptive test then selects the one with the most significant testing result with a proper adjustment for multiple testing. To apply the proposed test, we establish its asymptotic null distribution. In particular, we derive the joint null distribution of the individual powered score test statistics, which converge to either a multivariate normal distribution or an extreme value distribution. The joint asymptotic null distribution for the proposed tests is used

to calculate asymptotics-based p-values, a more convenient and faster alternative to other computing-intensive resampling methods such as the bootstrap.

The rest of the paper is organized as follows. In Section 2, we review some existing tests. In Section 3, we propose the new adaptive test and study its asymptotic properties in the contexts with and without nuisance parameters, respectively. Results for simulation studies and real data analyses are presented in Section 4. All technical details for proofs and more extensive simulation results are relegated to the online supplementary material. An R package *GLMaSPU* implementing the proposed test is also publicly available on GitHub and CRAN.

2. Some Existing Tests

Suppose n identical and independently distributed (i.i.d.) samples $\{(Y_i, Z_i, X_i) : i = 1, 2, \ldots, n\}$ have been collected, for which we have an n-vector response (outcome of interest) Y, an $n \times q$ matrix $\mathbb{Z}$ for q covariates, and an $n \times p$ matrix $\mathbb{X}$ for p variables of interest. For subject i, let $Z_i = (Z_{i1}, \ldots, Z_{iq})$ be the q covariates, such as age, gender, and other clinical variables that we want to adjust for, and $X_i = (X_{i1}, \ldots, X_{ip})$ be the p-dimensional variables of interest. Without loss of generality, we assume that $E(\mathbb{X}) = 0$ as otherwise $\mathbb{X}$ can be re-centered by its mean. Assuming a generalized

linear model, we have

$$E(Y|\mathbb{X},\mathbb{Z}) = g^{-1}(\mathbb{X}\beta + \mathbb{Z}\alpha), \tag{2.1}$$

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

$$H_0 : \beta = \beta_0 \qquad \text{versus} \qquad H_1 : \beta \neq \beta_0, \tag{2.2}$$

while treating α as the nuisance parameter. We target the situation with "small q, large p and large n".

The best-known tests for low-dimensional data are the Wald test and the likelihood ratio test; however, the power of both the Wald test and the likelihood ratio test diminishes quite rapidly as the dimension p increases (Goeman et al., 2006). More importantly, in a high-dimensional situation with $p > n$, these tests break down completely since the MLEs for the parameters no longer exist uniquely. Goeman et al. (2006) derived the following test statistic for testing hypothesis (2.2) based on the score vector

$$T_{\text{Goe}} = U^\intercal U - \text{trace}(\mathcal{I}),$$

where U and $\mathcal{I}$ are the score vector and observed information matrix for β under the null hypothesis, respectively. Ignoring some constant, T_{Goe} equals to

$$T_{\text{Goe2}} = n^{-1}(Y - \mu_0)^\intercal \mathbb{X}\mathbb{X}^\intercal (Y - \mu_0),$$

where μ_0 is the expectation of Y under the null hypothesis. Goeman et al. (2006) calculated the p-value of this test statistic via permutations or moment matching. Goeman et al. (2011) modified T_{Goe} with the following statistic

$$T_{\text{GT}} = \frac{(Y - \hat{\mu}_0)^\intercal \mathbb{X}\mathbb{X}^\intercal (Y - \hat{\mu}_0)}{(Y - \hat{\mu}_0)^\intercal \mathbb{D} (Y - \hat{\mu}_0)},$$

where $\hat{\mu}_0$ and $\mathbb{D}$ are the maximum likelihood estimate of μ_0 under the null hypothesis and a diagonal $n \times n$ matrix equal to the diagonal of $\mathbb{X}\mathbb{X}^\intercal$, respectively. Goeman et al. (2011) derived its asymptotic null distribution for fixed p. Since the denominator of T_{GT} increases the variance and thus adversely affects the power, Guo and Chen (2016) proposed the following test statistic

$$T_{\text{HDGLM}} = n^{-1}(Y - \hat{\mu}_0)^\intercal (\mathbb{X}\mathbb{X}^\intercal - \mathbb{D})(Y - \hat{\mu}_0),$$

and further derived the asymptotic normal distribution of T_{HDGLM} for diverging $p \to \infty$ as $n \to \infty$ under some assumptions.

Remark 1. To our knowledge, most high-dimensional tests are based on a sum-of-squared score vector, which have also been used in GWASs with large n and small p. For instance, Pan (2009) proposed a sum-of-squared-score test (similar to T_{Goe2}) for testing the association between multiple SNPs and the outcome of interest in GLMs. Another similar test is SKAT (Wu et al., 2011).

3. New Method

For the purpose of presentation, we first consider the case without nuisance parameters, then the case with nuisance parameters.

3.1 Testing Without Nuisance Parameters

In this subsection, we assume the GLM (2.1) with $\alpha = 0$. Many existing tests are based on the score vector $U = (U_1, \ldots, U_p)^\intercal$ for β, which, up to some constant, has elements

$$U_j = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \mu_{0i}) X_{ij}, \qquad 1 \le j \le p,$$

with $\mu_{0i} = g^{-1}(X_i \beta_0)$.

For notation simplicity, we write $S_{ij} = (Y_i - \mu_{0i}) X_{ij}$ for $1 \le i \le n$ and $1 \le j \le p$. As to be demonstrated later, depending on the unknown association effects β to be

tested, different tests may be more powerful. Inspired by Pan et al. (2014), we would use U to construct some weights to upweight more informative components of the score vector, proposing a sum of powered score (SPU) test statistic with power index $0 < \gamma < \infty$ as

$$L(\gamma, \mu_0) = \sum_{j=1}^{p} w_j U_j = \sum_{j=1}^{p} U_j^{\gamma-1} U_j = \sum_{j=1}^{p} U_j^{\gamma} = \sum_{j=1}^{p} \Big(\frac{1}{n} \sum_{i=1}^{n} S_{ij}\Big)^{\gamma},$$

where $w_j = U_j^{\gamma-1}$ can be considered as a data-dependent weight.

Note that $\gamma = 2$ yields a sum-of-squares-type test statistic, which is similar to the existing tests reviewed in the previous section. As an even integer $\gamma \to \infty$, we have $L(\gamma, \mu_0) \propto L(\gamma, \mu_0)^{1/\gamma} \to \max_{1\le j\le p} \Big(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \mu_{0i})X_{ij}\Big)$ thus we define $L(\infty, \mu_0)$ as

$$L(\infty, \mu_0) = \max_{1\le j\le p} \frac{n \left(\frac{1}{n}\sum_{i=1}^{n} S_{ij}\right)^2}{\sigma_{jj}},$$

where $\mathbf{\Sigma} = (\sigma_{kj})_{p\times p}$, and $\sigma_{kj} = \mathrm{Cov}[S_{ik}, S_{ij}]$ for $1 \le k, j \le p$. Note that here the covariance matrix $\mathbf{\Sigma}$ is defined *unconditionally* on the covariates and consequently it does not depend on the subject index i. See Remark 7 for more discussion.

The class of the SPU tests cover several tests used in GWASs as special cases. For example, for large n and small p, $L(2, \mu_0)$ is like SKAT with a linear kernel (Wu et al., 2011); $L(1, \mu_0)$ is a burden test in genetic rare variant assoication analysis

(Morgenthaler and Thilly, 2007). As to be shown in simulations, if most variables of $\mathbb{X}$ are associated with the response Y with similar effect sizes and the same association direction, then a burden test like $L(1, \mu_0)$ would yield high statistical power. In contrast, in a situation with only moderately dense signals or with different association directions, $L(\gamma, \mu_0)$ with an even integer $\gamma \geq 2$ would be more powerful. In particular, the supremum based test statistic, $L(\infty, \mu_0)$ yields high statistical power if only few variables are strongly associated with Y (i.e. a highly sparse non-zero components of β). In short, the power of $L(\gamma, \mu_0)$ depends on the unknown true association pattern (i.e. value of β), such as signal sparsity and magnitudes. To choose the most powerful test automatically, we propose the following adaptive test to combine the multiple tests accordingly:

$$T_{\text{aSPU}} = \min_{\gamma \in \Gamma} P_{\text{SPU}(\gamma, \mu_0)},$$

where $P_{\text{SPU}(\gamma,\mu_0)}$ is the p-value of $L(\gamma, \mu_0)$ test. For simplicity, we write $L(\gamma, \mu_0)$, $\text{SPU}(\gamma, \mu_0)$ and $\text{SPU}(\gamma)$ exchangeably. Taking the minimum p-value is a simple and effective way to approximate the most powerful test (Pan et al., 2014). Note that T_{aSPU} is no longer a genuine p-value and we need to derive its asymptotic null distribution to facilitate calculating its p-value.

Remark 2. The optimal value of γ for the test statistic $L(\gamma)$ to achieve the highest

power depends on the specific alternative. We aim to choose a Γ set to maintain high power of the aSPU test under a wide range of scenarios. The supremum based test statistic for high-dimensional two-sample testing has been studied in Cai et al. (2014); from their Theorem 2, the power of the supremum based test converges to 1 if the signal is strong with a high sparsity level; see also related discussions in Donoho and Jin (2015) and Jin and Ke (2014). When the signal is dense with a constant effect size, $L(1)$ is most powerful (Xu et al., 2016). $L(2)$ is a sum-of-squares-type test that has been widely used and studied. By default, we recommend include $\gamma = 1, 2, \infty$ and a small subset of moderate values of γ in Γ. More generally, as recommended in Xu et al. (2016), we use $\Gamma = \{1, 2, \ldots, \gamma_u, \infty\}$ with a γ_u such that $L(\gamma_u)$ gives similar results to that of $L(\infty)$; we find in the simulation studies that often $\gamma_u = 6$ or 8 suffices and the performance of the aSPU test is robust to such a choice of γ_u.

Remark 3. Our proposed test is an extension of the original aSPU test (Pan et al., 2014) to high-dimensional GLMs; the original aSPU test was proposed for analysis of rare variants with large n and small p. For simplicity, we use the same name "aSPU" for our proposed test here. Since the asymptotic properties of the adaptive aSPU test for GLMs have not been studied, we derive its asymptotic null distribution in a high-dimensional setting, based on which the asymptotic p-values of $L(\gamma, \mu_0)$ and T_{aSPU} can be calculated.

Next we derive the asymptotic properties under the null hypothesis. For two

sequences of real numbers $\{a_n\}$ and $\{b_n\}$, we write $a_n = O(b_n)$ if there exists some constant C such that $|a_n| \leq C|b_n|$ holds for all $n \geq N$, and write $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$. Under $H_0 : \beta = \beta_0$, we first derive some asymptotic approximations to the mean and the variance of $L(\gamma, \mu_0)$ for $\gamma < \infty$, and then establish the asymptotic distribution of $L(\gamma, \mu_0)$. The following assumptions are needed.

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; in other words, $\max_{1\leq i\neq j\leq p} |\sigma_{ij}|/\sqrt{\sigma_{ii}\sigma_{jj}} < 1 - \xi$ for some constant $\xi > 0$.

C2. Given a set of multivariate random vectors $W = \{W^{(j)} : j \geq 1\}$, for integers $a < b$, let χ_a^b be the σ-algebra generated by $\{W^{(m)} : m \in [a, b]\}$. The α-mixing coefficient $\alpha_W(s)$ is defined as $\sup\{|Pr(A \cap B) - Pr(A)Pr(B)| : 1 \leq t < p, A \in \chi_1^t, B \in \chi_{t+s}^{\infty}\}$. We assume $W = \{W^{(j)} = (S_{ij}, i = 1, \ldots, n) : j \geq 1\}$ is α-mixing such that $\alpha_W(s) \leq M\delta^s$, where $\delta \in (0, 1)$ and M is some constant.

C3. Under $H_0 : \beta = \beta_0$, $E\left[(S_{ij})^3\right] = 0$ for $1 \leq j \leq p$.

C4. $(\log p)/n^{1/4} = o(1)$.

C5. There exist some constants η and $K > 0$ such that $E\left[\exp\left\{\eta(S_{ij})^2/\sigma_{jj}\right\}\right] \leq K$ for $1 \leq j \leq p$.

Remark 4. Assumptions C1, C4, and C5 are mild conditions, and are used to establish the weak convergence of $L(\infty, \mu_0)$. Cai et al. (2014) used exactly the same assumptions (C1, C4, and C5) when deriving the limiting distribution of a supremum-type test statistic for high-dimensional two-sample mean testing. Assumption C2 assumes an α-mixing-type weak dependence structure of the data, which has been widely used in spatial statistics and time series. For high-dimensional two-sample mean testing, a similar mixing condition has been used in Xu et al. (2016) and Chen et al. (2014). Alternatively, we may consider the weak dependence structure adopted in Guo and Chen (2016), where a factor-type model for $S_i = (S_{i1}, \ldots, S_{ip})^\intercal$ is assumed. Intuitively, many random vectors, e.g., any ergodic and aperiodic Markov chain, meet the α-mixing weak dependence condition. Another example is for random vectors $X = (X_1, X_2, \ldots)^\intercal$, where X_i and X_j are independent with $|i - j| > C$ for some constant C; then $\alpha_X(s) = 0$ if $s > C$, satisfying the α-mixing assumption as well. This type of structure has also been used for estimating a high-dimensional covariance matrix (Bickel and Levina, 2008). In addition, because the correlations among variables (i.e. SNPs) in our motivating genome-wide association study data decay to zero as their physical distances on the same chromosome increase (while the SNPs from different chromosomes are usually independent), the α-mixing assumption fits the application well and thus will be used in this paper.

We write $L(\gamma, \mu_0) = \sum_{j=1}^{p} L^{(j)}(\gamma, \mu_0)$ with $L^{(j)}(\gamma, \mu_0) = \frac{1}{n} \sum_{i=1}^{n} S_{ij}|^{\gamma}$, then de-

note $\mu(\gamma) = \sum_{j=1}^{p} \mu^{(j)}(\gamma)$ with $\mu^{(j)}(\gamma) = E\ \ L^{(j)}(\gamma, \mu_0)|$, and $\sigma^2(\gamma) = \text{var}(L(\gamma, \mu_0))$.

PROPOSITION 1. *Under assumptions C1, C3, and* $H_0 : \beta = \beta_0$, $\mu(1) = 0$ *and*

$$\mu(\gamma) = \begin{cases} \frac{\gamma!}{d!2^d} n^{-d} \sum_{j=1}^{p} \sigma_{jj}^d + o(pn^{-d}), & \text{if } \gamma = 2d, \\ o(pn^{-(d+1)}), & \text{if } \gamma = 2d+1, \end{cases}$$

where $\sigma_{jj} = E[(S_{ij})^2]$.

Proposition 1 follows directly from the Lemma 1 in the online supplementary.

PROPOSITION 2. *Under assumptions C1–C3 and* H_0, $\sigma^2(1) = \frac{1}{n} \sum_{1 \le i,j \le p} \sigma_{ij} + o(pn^{-1})$ *and for* $\gamma \ge 2$,

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

where $\sigma_{ij} = E[S_{ki} S_{kj}]$.

Note that the order of $\sigma^2(\gamma)$ is $pn^{-\gamma}$. Then we derive the following result to approximate the correlations among the $L(\gamma, \mu_0)$.

PROPOSITION 3. *Under assumptions C1–C3 and* $H_0 : \beta = \beta_0$, *for any finite and positive integers* $s, t \in \Gamma$, *we have*

(i) if $s+t$ is even,

$$Cov\{L(t,\mu_0),L(s,\mu_0)\}$$
$$= \mu(t+s) - \sum_{i=1}^{p} \mu^{(i)}(t)\mu^{(i)}(s) + \frac{1}{n^c} \sum_{i\neq j} \sum_{\substack{2c_1+c_3=t \\ 2c_2+c_3=s \\ c_3>0}} \frac{t!s!}{c_3!c_1!c_2!2^{c_1+c_2}} \sigma_{ii}^{c_1}\sigma_{jj}^{c_2}\sigma_{ij}^{c_3} + o(pn^{-(t+s)/2}).$$

(ii) if $s+t$ is odd, $Cov\{L(t,\mu_0),L(s,\mu_0)\} = o(pn^{-(t+s)/2})$.

Let Γ be a candidate set of γ with $\infty \in \Gamma$. We further define $R = (\rho_{st})$, where $\rho_{ss} = 1$ for $s \in \Gamma \setminus \{\infty\}$ and $\rho_{st} = \mathrm{Cov}\{L(s,\mu_0),L(t,\mu_0)\}/\{\sigma(s)\sigma(t)\}$ for $s \neq t \in \Gamma \setminus \{\infty\}$. In particular, $\rho_{st} = o(1)$ when $s+t$ is odd. Then we introduce Theorem 1, which describes the asymptotic distribution of $L(\gamma,\mu_0)$.

THEOREM 1. *Under assumptions C1–C5 and the null hypothesis H_0, we have:*
(i) For set $\Gamma' = \Gamma \setminus \{\infty\}$, the vector of the standardized test statistics $[\{L(\gamma,\mu_0) - \mu(\gamma)\}/\sigma(\gamma)]^{\intercal}_{\gamma\in\Gamma'}$ converges weakly to a normal distribution $N(0,R)$ as n, $p \to \infty$.
(ii) When $\gamma = \infty$, let $a_p = 2\log p - \log\log p$, for any $x \in \mathbb{R}$, $Pr\{L(\infty,\mu_0) - a_p \leq x\} \to \exp\{-\pi^{-1/2}\exp(-x/2)\}$.
(iii) $[\{L(\gamma,\mu_0) - \mu(\gamma)\}/\sigma(\gamma)]^{\intercal}_{\gamma\in\Gamma'}$ is asymptotically independent with $L(\infty,\mu_0)$. That is, the joint distribution of $[\{L(\gamma,\mu_0) - \mu(\gamma)\}/\sigma(\gamma)]^{\intercal}_{\gamma\in\Gamma'}$ and $L(\infty) - a_p$ converges weakly to the product of the limiting distributions given in (i) and (ii).

Remark 5. Testing without nuisance parameters can be treated as a special case of

testing with nuisance parameters. The methods described in the following subsection can be used for calculating the p-values for testing without nuisance parameters by replacing $\hat{\mu}_0$ with μ_0.

3.2 Testing With Nuisance Parameters

In this subsection, we consider testing on a high-dimensional regression coefficient vector in the presence of a low-dimensional nuisance parameter, which is a common task in practice. For example, in a study of complex disease, we usually have both SNP data and other demographic variables, which may confound the association between the SNPs and the outcome of interest. One may be interested only in genetic effects while adjusting for demographic variables, hence the coefficients for demographic variables are treated as low-dimensional nuisance parameters, which have to be estimated. Here, we are interested in testing hypothesis (2.2) under GLM (2.1).

Let $\mu_0(\alpha) = \mu_0 = g^{-1}(\mathbb{Z}\alpha + \mathbb{X}\beta_0)$ and $\hat{\mu}_0 = g^{-1}(\mathbb{Z}\hat{\alpha} + \mathbb{X}\beta_0)$, where the MLE $\hat{\alpha}$ is obtained under the null hypothesis. Since μ_0 is unknown, we use $\hat{\mu}_0$ and the test statistic $L(\gamma, \hat{\mu}_0)$ accordingly. To derive its asymptotic distribution, the following additional assumptions are needed.

C6. The dimension of nuisance parameters α, q, is fixed, and each covariate in $\mathbb{Z}$ is bounded almost surely. We assume $E(X_{ij}|\mathbb{Z}) \neq 0$ only holds for $j \in P_0$ with the

size of P_0, p_0, satisfying $p_0 = O(p^\eta)$ for a small positive η. We further assume the consistent and asymptotic normal MLE $\hat{\alpha}$ under the null hypothesis (Fahrmeir and Kaufmann, 1985).

C7. There exist some positive constants K_1 and K_2 such that $K_1 < E[\epsilon_{0i}^2|\mathbb{Z} = z] < K_2$ almost everywhere for z in the support of the probability density of Z, where $\epsilon_{0i} = Y_i - \mu_{0i}$, $1 \leq i \leq n$. We further assume $E[\epsilon_{0i}|\mathbb{X}, \mathbb{Z}] = 0$.

C8. We assume $p/n^2 = o(1)$.

C9. The conditionally α-mixing coefficient $\alpha_{W|\mathcal{F}}(s)$ is defined as $\sup\{|Pr(A \cap B|\mathcal{F}) - Pr(A|\mathcal{F})Pr(B|\mathcal{F})| : 1 \leq t < p, A \in \chi_1^t, B \in \chi_{t+s}^\infty\}$, where $\mathcal{F}$ is a sub-σ-algebra of W. We assume $W = \{W^{(j)} = (X_{ij}, i = 1, \ldots, n) : j \geq 1\}$ is conditionally α-mixing given $\mathbb{Z}$ such that $\alpha_{W|\sigma(\mathbb{Z})}(s) \leq M\delta^s$, where $\delta \in (0, 1)$ and M is some constant.

Remark 6. Assumption C6 states that the dimension of nuisance parameters, q, is fixed as $n \to \infty$, which is appropriate in many applications, including GWASs of interest here. However, this assumption may not be appropriate in some applications. For example, in testing gene-environmental interactions, the main effects are treated as nuisance parameters, which may be high-dimensional (Lin et al., 2013). Note that, we assume that each X_j is already centered and has sample mean 0, partially making it reasonable to assume $E[X_{ij}|\mathbb{Z}] \neq 0$ only for $j \in P_0$ with the size of P_0 in a small order of p (i.e. $p_0 = O(p^\eta)$). This assumption is technically needed to prove Theorem 2. For finite γ, we can relax the assumption to $p_0 = O(p^{1/2-\delta})$, where

δ is a small constant. If we are concerned about the validity of this assumption, we can regress each X_j on $\mathbb{Z}$ and use its residuals as the new X_j to approximately satisfy $E[X_{ij}|\mathbb{Z}] = 0$ for any $j = \{1, 2, \ldots, p\}$. Assumption C7 is common in GLMs, for instance, as assumption G in Fan et al. (2010) and assumption 3.3 in Guo and Chen (2016). Assumption C8 is an updated version of C4 and somewhat restrictive, which however is technically needed to prove Theorem 2. Note that, instead of considering only the sum-of-squares-type statistic (with $\gamma = 2$) similar to the HDGLM (Guo and Chen, 2016), here we derive the asymptotic distributions for any finite γ and $\gamma = \infty$, for which a stronger assumption is therefore used. However, this assumption may be relaxed: as to be shown in simulations, the asymptotic distribution still performed well for more general high dimensional situations, and we leave this interesting problem to future work. Conditionally α-mixing is introduced by Rao (2009) and assumption C9 is an updated version of C2 to adjust the case of nuisance parameters.

Although the estimated parameter $\hat{\alpha}$ does complicate the derivations, we still have the following theorem similar to Theorem 1.

THEOREM 2. *Under assumptions C1–C9 and the null hypothesis H_0, we have:*
(i) For set $\Gamma' = \Gamma \setminus \{\infty\}$, $[\{L(\gamma, \hat{\mu}_0) - \mu(\gamma)\}/\sigma(\gamma)]^{\intercal}_{\gamma \in \Gamma'}$ converges weakly to the normal distribution $N(0, R)$ specified in Theorem 1 as $n, p \to \infty$.

(ii) When $\gamma = \infty$*, let* $a_p = 2\log p - \log\log p$*, for any* $x \in \mathbb{R}$*,* $Pr\{L(\infty, \hat{\mu}_0) - a_p \leq x\} \to \exp\{-\pi^{-1/2}\exp(-x/2)\}$.

(iii) $[\{L(\gamma, \hat{\mu}_0) - \mu(\gamma)\}/\sigma(\gamma)]^{\intercal}_{\gamma\in\Gamma'}$ *is asymptotically independent with* $L(\infty, \hat{\mu}_0)$.

Remark 7. In a GLM, conditional on $\mathbb{Z}$ and $\mathbb{X}$, we usually have $\mathrm{Cov}[S_{ik}, S_{ij}|\mathbb{Z}, \mathbb{X}] \neq \mathrm{Cov}[S_{i'k}, S_{i'j}|\mathbb{Z}, \mathbb{X}]$ for $i \neq i'$. In our derivations, we treat $\mathbb{Z}$ and $\mathbb{X}$ as random and assume the data are independently and identically distributed, which makes σ_{kj} well defined (unconditionally); and we derive the unconditional version of the asymptotic null distribution.

Since $\mu(\gamma)$, $\sigma(\gamma)$, and R can be approximated according to Propositions 1–3, respectively, the p-values for individual $L(\gamma, \hat{\mu}_0)$ can be calculated via either a normal or an extreme value distribution. We illustrate how to calculate the p-value for aSPU. Define $L_O = [\{L(\gamma, \hat{\mu}_0) - \mu(\gamma)\}/\sigma(\gamma) : \text{odd } \gamma \in \Gamma']$ and $L_E = [\{L(\gamma, \hat{\mu}_0) - \mu(\gamma)\}/\sigma(\gamma) : \text{even } \gamma \in \Gamma']$. By Proposition 3, $\mathrm{Cov}(L(t), L(s))$ is a small order term if $t + s$ is odd, implying L_O and L_E are asymptotically uncorrelated. By Theorem 2, L_O and L_E converge jointly and weakly to a multivariate normal distribution as n, $p \to \infty$, implying L_O and L_E are asymptotically independent. Further, by Theorem 2, $L(\infty, \hat{\mu}_0)$ is asymptotically independent of both L_O and L_E. Then we can calculate the p-value for aSPU via the following procedure.

Step 1 Define $t_O = \max_{\text{odd } \gamma\in\Gamma'} |\{L(\gamma, \hat{\mu}_0) - \mu(\gamma)\}/\sigma(\gamma)|$ and $t_E = \max_{\text{even } \gamma\in\Gamma'} \{L(\gamma, \hat{\mu}_0) -$

$\mu(\gamma)\}/\sigma(\gamma)$ as the observed test statistic from the data and calculate the p-value for t_O and t_E as $p_O = Pr[\max_{\text{odd } \gamma\in\Gamma'} |\{L(\gamma, \hat{\mu}_0) - \mu(\gamma)\}/\sigma(\gamma)| > t_O]$ and $p_E = Pr[\max_{\text{even } \gamma\in\Gamma'}\{L(\gamma, \hat{\mu}_0) - \mu(\gamma)\}/\sigma(\gamma) > t_E]$. Use function *pmvnorm()* in R package *mvtnorm* to calculate the multivariate normal tail probabilities p_O and p_E.

Step 2 Calculate the p-value p_∞ of $L(\infty, \hat{\mu}_0)$ based on its asymptotic extreme value distribution.

Step 3 By the asymptotic independence, the asymptotic p-value for the aSPU test is $p_{\text{aSPU}} = 1 - (1 - p_{\min})^3$, where we have $p_{\min} = \min\{p_O, p_E, p_\infty\}$.

The above discussion assumes that the covariance matrix $\mathbf{\Sigma}$ is known. In practice, $\mathbf{\Sigma}$ has to be estimated. We may apply an existing method, such as the banding and thresholding technique, to estimate a high-dimensional sparse covariance matrix (Bickel and Levina, 2008; Cai and Liu, 2011); see Cai et al. (2016) for an excellent review. Under the α mixing assumption C2, σ_{ij} is close to zero when $|i - j|$ is large and thus we may apply the banding approach of Bickel and Levina (2008) to estimate covariance matrix $\mathbf{\Sigma}$. Specifically, we first calculate the sample covariance matrix $\mathbb{S} = (s_{ij})$, where $s_{ij} = \frac{1}{n-1}\sum_{k=1}^{n}(Y_k - \hat{\mu}_{0k})^2 X_{ki}X_{kj}$, then we further calculate the bandable covariance matrix with bandwidth k_n as $\hat{\mathbf{\Sigma}}_{k_n} = (s_{ij}I(|i - j| \le k_n))$. Theoretically optimal bandwidth k_n and minimax risk rates of $\hat{\mathbf{\Sigma}}_{k_n}$ have been studied

in Bickel and Levina (2008). Since a theoretically optimal k_n is determined by some unknown hyper-parameters, we use five-fold cross-validation to select an optimal bandwidth k_n (Bickel and Levina, 2008; Cai and Liu, 2011). Following Xu et al. (2016), under the assumptions in Theorem 2, we can show that $\hat{\mu}(\gamma)$ and $\hat{\sigma}^2(\gamma)$ estimated based on the bandable covariance matrix $\hat{\mathbf{\Sigma}}_{k_n}$ satisfy $\hat{\mu}(\gamma) = \{1+o(1)\}\mu(\gamma)$ and $\hat{\sigma}^2(\gamma) = \{1 + o(1)\}\sigma^2(\gamma)$ for properly chosen $k_n = o(n^{1/2})$.

With a relatively small sample size, five-fold cross-validation may select a smaller bandwidth than the optimal one and thus yield an underestimated $\hat{\sigma}^2(\gamma)$ and a smaller p-value. Alternatively, we propose to use the parametric bootstrap to estimate $\hat{\mu}(\gamma)$, $\hat{\sigma}^2(\gamma)$, and R. We first fit a null model under H_0 to obtain $\hat{\mu}_{0i} = \hat{E}(Y_i|Z_i, H_0)$, then simulate a new set of responses $Y_i^{(b)}$ from the corresponding model for $b = 1, 2, \ldots, B$. For example, for a binary outcome of interest, generate $Y_i^{(b)} \sim \text{Bin}(1, \hat{\mu}_{0i})$. We refit the model with $\{Y_i^{(b)} : i = 1, 2, \ldots, n\}$ and calculate the corresponding test statistic $L(\gamma, \hat{\mu}_0)^{(b)}$. Then $\hat{\mu}(\gamma) = \sum_{b=1}^{B} L(\gamma, \hat{\mu}_0)^{(b)}/B$, $\hat{\sigma}^2(\gamma) = \sum_{b=1}^{B}(L(\gamma, \hat{\mu}_0)^{(b)} - \hat{\mu}(\gamma))^2/(B-1)$ and $\hat{R} = \text{cor}(L(\Gamma, \hat{\mu}_0))$, where cor is the sample correlation. Unlike that the accuracy of a usual resampling method is bounded by the number of resampling B and thus a large B is needed for calculating a very small p-value, we can use a relatively small B to calculate $\hat{\mu}(\gamma)$, $\hat{\sigma}^2(\gamma)$, R and then an asymptotic p-value. Although estimating the mean and covariance matrix differently, the above two methods are still based on the asymptotics to calculate the

p-values, hence are called asymptotics-based methods in the following. In contrast, we can also simply use the parametric bootstrap to calculate the p-values (without direct use of the asymptotic results), which will be more time-consuming (requiring a large B for a highly significant p-value) but may perform better for finite samples; in the sequel, by default, the parametric bootstrap refers to this way of calculating the p-values.

Remark 8. The optimal value of γ for the test $L(\gamma, \hat{\mu}_0)$ to achieve the highest power depends on the true alternative. As to be shown in the numerical results, when the signal β is highly dense with the same sign, $L(1, \hat{\mu}_0)$ is more powerful than the competing tests. $L(2, \hat{\mu}_0)$ performs similarly to the tests of Guo and Chen (2016) since they have similar test statistics. There are some other situations, under which $L(2, \hat{\mu}_0)$ is not as powerful as other $L(\gamma, \hat{\mu}_0)$ tests, and therefore in these cases, the proposed test is more powerful than the competing tests. When the signal is strong and highly sparse, $L(\infty, \hat{\mu}_0)$ is more powerful. Due to the nature of its adaptiveness, the power of the aSPU test is often either the highest or close to the highest.

4. Numerical Results

4.1 Simulations

We conducted extensive simulations to compare the performance of the proposed adaptive test with two existing methods, the HDGLM (Guo and Chen, 2016) and the GT (Goeman et al., 2011), due to their popularity and the availability of their computer code.

We set the sample size $n = 200$ and the dimension of β $p = 2000$, though other values were also considered. We generated a data matrix $\mathbb{X}_{n\times p}$ from a multivariate normal distribution; that is, we had independent $X_i \sim N(0, \mathbf{\Xi})$ for $i = 1, 2, \ldots, n$. We show the results with unit variances and a blocked first-order autoregressive correlation matrix $\mathbf{\Xi} = (\mathbf{\Xi}_{ij})$ with $\mathbf{\Xi}_{ij} = 0.4^{|i-j|}$ if $|i - j| \leq 3$ and 0 otherwise. Other simulation results with other covariance structures are presented in the supplementary material.

We further generated a data matrix with two covariates $\mathbb{Z}$ from a normal distribution $N(0, 0.5)$. The outcome Y was generated from a logistic regression model as in GLM (2.1) with a logit link function, $\alpha = (1, 1)^\intercal$, and $\beta = 0$ or $\neq 0$, corresponding to the null hypothesis H_0 or an alternative hypothesis H_1 respectively. Here, we mainly focused on the results for a binary outcome since in our real data application the response is binary and it is generally more challenging than that for a continuous

outcome. Under H_1, $\lfloor ps \rfloor$ elements of β were set to be non-zero, where $s \in [0, 1]$ controlled the degree of signal sparsity. We varied s to mimic varying sparsity levels, covering from highly sparse signals at $s = 0.001$ to less sparse and then to moderate dense at $s = 0.1$, finally to dense and highly dense signals at $s = 0.7$, respectively. The indices of non-zero elements in β were assumed to be uniformly distributed in $\{1, 2, \ldots, p\}$, and their values were constant at c. We varied s, c, n and p to evaluate the performance of the new method under various situations. We used the parametric bootstrap (Pan et al., 2014) to obtain a 'bronze-standard' (slightly inferior to a 'gold standard', where the true p-value is known) analysis, to which we compared the asymptotic results based on Theorem 2. In all simulations, we treated $\mathbf{\Sigma}$ as unknown and thus estimated $\mathbf{\Sigma}$, then calculated the means and covariances of the SPU test statistics according to Propositions 1–3. For each set-up, we simulated 1,000 data sets and averaged the testing results of these 1,000 data sets. The nominal significance level was set to $\alpha = 0.05$. For the aSPU test, the candidate set of γ was by default set to be $\Gamma = \{1, 2, \ldots, 6, \infty\}$.

Table 1 shows the type I error rates and power for $s = 0.1$. The results outside and inside parentheses in Table 1 were calculated from asymptotics- and parametric bootstrap-based methods, respectively; the results based on the two methods were very close to each other, confirming the results in Theorem 2. We further studied the performance of the asymptotics-based method under different sparsity levels ($s =$

$0.001, 0.05, 0.7$) and dimension $p = 4000$. The results for those simulation settings were similar to Table 1 and were relegated to the supplementary Tables S1–S5.

Table 1: Empirical type I error rates and power (%) of various tests in simulations with $n = 200$ and $p = 2000$. The sparsity parameter was $s = 0.1$, leading to 200 non-zero elements in β with a constant value c. The results outside and inside parentheses were calculated from asymptotics- and parametric bootstrap-based methods, respectively.

c	0	0.03	0.05	0.07	0.1	0.15
SPU(1)	5 (5)	33 (32)	59 (59)	73 (74)	84 (86)	92 (92)
SPU(2)	6 (5)	18 (15)	44 (39)	65 (61)	81 (78)	91 (89)
SPU(3)	4 (5)	28 (30)	58 (59)	76 (76)	89 (90)	96 (96)
SPU(4)	4 (6)	11 (14)	33 (36)	55 (58)	74 (75)	87 (87)
SPU(5)	4 (5)	15 (18)	37 (41)	59 (62)	78 (81)	88 (89)
SPU(6)	3 (6)	7 (11)	18 (24)	36 (43)	53 (59)	70 (72)
SPU(∞)	5 (5)	7 (7)	8 (9)	13 (16)	19 (22)	21 (25)
aSPU	5 (5)	22 (25)	53 (57)	75 (77)	90 (90)	96 (96)

Figure 1 shows the empirical power for different methods under high-dimensional scenarios. When the signals were extremely sparse at $s = 0.001$, as expected, the supremum-type test SPU(∞) and aSPU performed much better than the competing tests, the GT and the HDGLM, in terms of power. When the signal non-sparsity increased from 0.001 to 0.05, the aSPU test performed similarly to the sum-of-squares-type tests, such as the GT and the HDGLM, and it was much more powerful than the supremum-type test SPU(∞). As the signals became more dense at $s = 0.1$, the aSPU test was the most powerful, closely followed by the SPU(1) and SPU(2) tests. At $s = 0.7$, the aSPU test remained to be the winner, and the SPU(1) test was more powerful than the sum-of-squares-type and supremum-type tests. Under all the

situations considered, the aSPU consistently maintained high power, being either the winner or close to the winner.

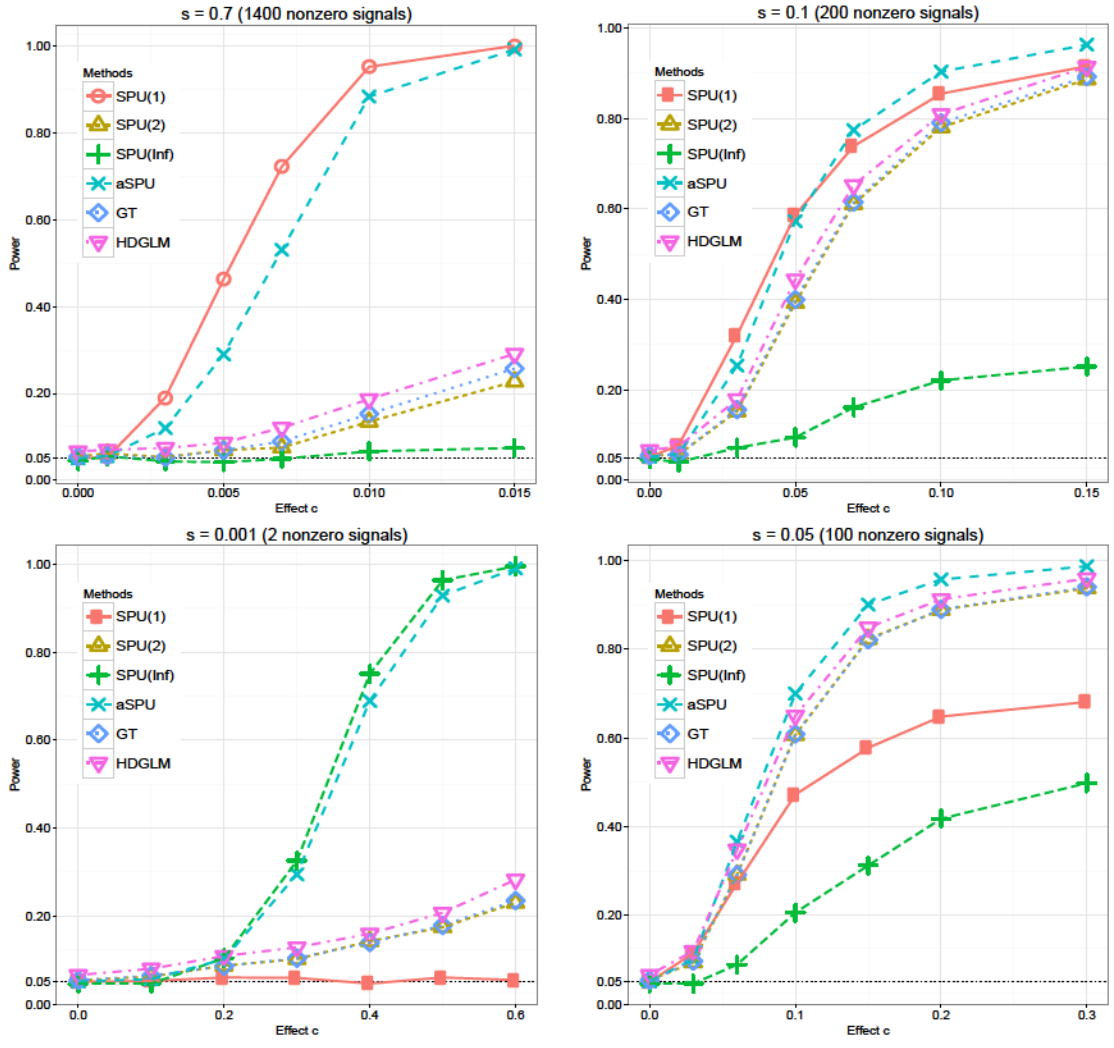

Figure 1: Empirical powers of SPU(1), SPU(2), SPU(∞), aSPU, GT (Goeman et al., 2011), and HDGLM (Guo and Chen, 2016). The signal sparsity parameter s varies from 0.001 to 0.7. We set $n = 200$ and $p = 2000$.

Next, we analyzed the sensitivity of the aSPU test to the choice of Γ. Figure 2

shows the results for aSPU with $\Gamma_1 = \{1, 2, \ldots, 4, \infty\}$, $\Gamma_2 = \{1, 2, \ldots, 6, \infty\}$, $\Gamma_3 = \{1, 2, \ldots, 8, \infty\}$, and $\Gamma_4 = \{1, 2, \ldots, 10, \infty\}$ under different scenarios. As shown in Figure 2, the aSPU test was relatively robust to the choice of Γ.

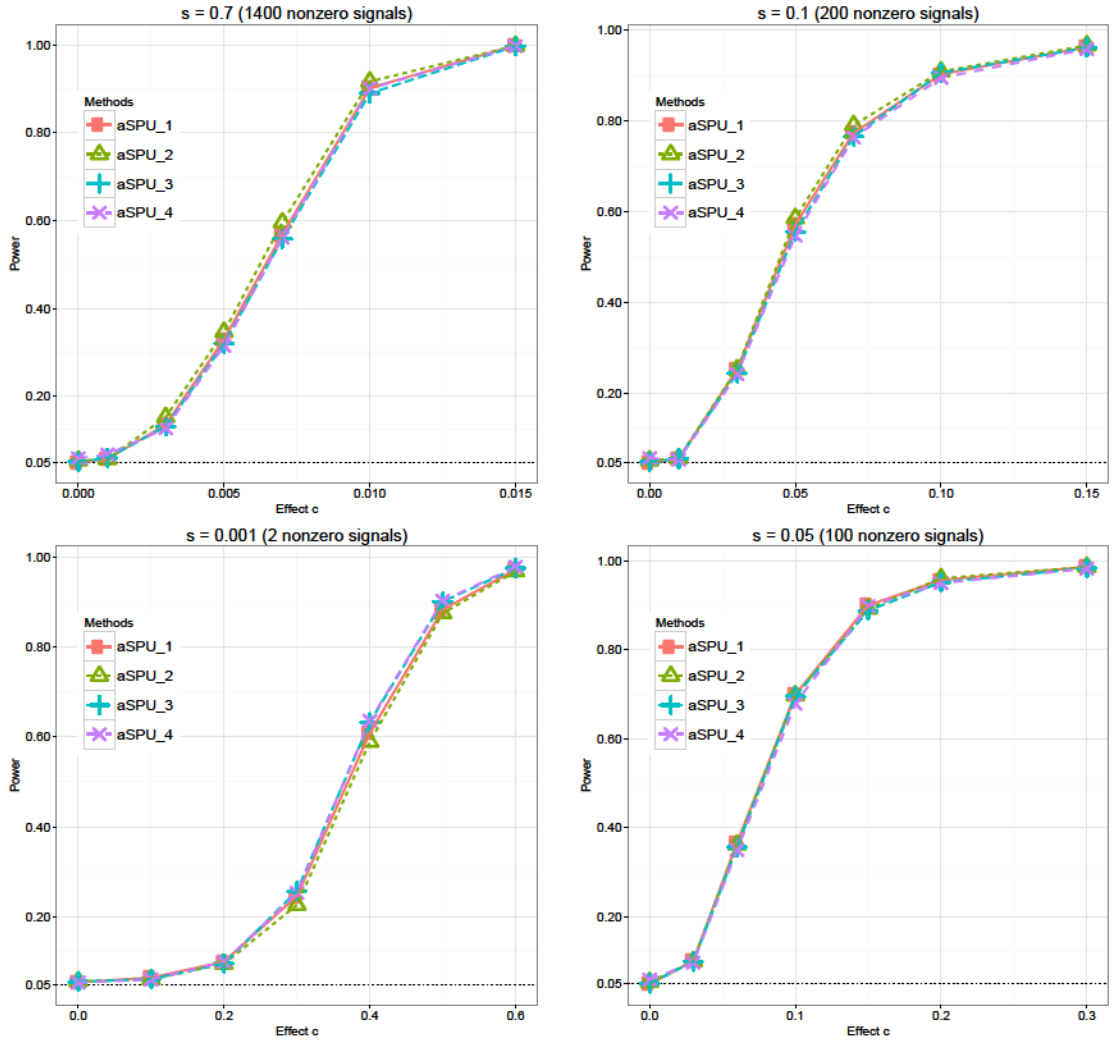

Figure 2: Empirical powers of aSPU with different Γ set. aSPU_1, aSPU_2, aSPU_3, aSPU_4 represent aSPU with $\Gamma_1 = \{1, 2, \ldots, 4, \infty\}$, $\Gamma_2 = \{1, 2, \ldots, 6, \infty\}$, $\Gamma_3 = \{1, 2, \ldots, 8, \infty\}$, and $\Gamma_4 = \{1, 2, \ldots, 10, \infty\}$, respectively. The signal sparsity parameter s varies from 0.001 to 0.7. We set $n = 200$ and $p = 2000$.

To further study the impact of covariance structures, we considered the following two other covariance structures as used in Cai et al. (2014). The first was a block diagonal structure: $\mathbf{\Xi} = (\sigma^*_{i,j})$ with $\sigma^*_{i,i} = 1$, $\sigma^*_{i,j} = 0.8$ for $2(k-1)+1 \leq i \neq j \leq 2k$ and $k = 1, \ldots, [p/2]$, and $\sigma^*_{i,j} = 0$ otherwise. The second was a non-sparse structure: let $\mathbf{\Xi}^+ = (\sigma^+_{i,j})$ with $\sigma^+_{i,i} = 1$ and $\sigma^+_{i,j} = |i-j|^{-5}/2$ for $i \neq j$ and let $\mathbb{D} = (d_{i,j})$ be a diagonal matrix with diagonal elements $d_{i,i}$ following a uniform distribution between 1 and 3 for $i = 1, \ldots, p$. Then $\mathbf{\Xi} = \mathbb{D}^{1/2}\mathbf{\Xi}^+\mathbb{D}^{1/2}$. For these two covariance structures, the results of the asymptotic approximation and power comparison were similar to those in Table 1 and Figure 1, thus were relegated to the supplementary Tables S6–S13 and Figures S1–S2. Finally, we also considered a continuous outcome Y; again, the results were similar (supplementary Table S14).

In summary, due to the nature of its adaptiveness, the aSPU test either achieved the highest power or was close to the winner under various scenarios, validating its consistently good performance across a wide range of scenarios.

4.2 Real Data Analysis

Alzheimer's disease (AD) is the most common form of dementia, affecting many millions around the world. The Alzheimer's Disease Neuroimaging Initiative (ADNI) is a longitudinal multisite observational study of healthy elders, mild cognitive impairment (MCI), and AD (Jack et al., 2008). It is jointly funded by the National

Institutes of Health (NIH) and industry via the Foundation for the NIH and the Principal Investigator of this initiative is Michael W. Weiner, VA Medical Center and University of California. The major goal of ADNI is to test whether serial MRI, positron emission tomography (PET), and other biological markers can be combined to measure the progression of MCI and early AD. ADNI has recruited more than $1,500$ subjects, ages range from 55 to 90, to participate in the research. For latest information, see `www.adni-info.org`.

One objective of ADNI is to elucidate genetic susceptibility to AD. Due to a relatively small sample size and usually small genetic effect sizes, applying a univariate test to the ADNI data failed to identify any SNP passing the genome-wide significance level at 5×10^{-8} (Kim et al., 2016), and even a much larger meta-analysis of 74,046 individuals only identified very few genome-wide significant SNPs (Lambert et al., 2013). Hence, it is natural to consider possible associations at the pathway or even chromosome level, which may be more powerful through effect aggregation and a reduced burden of multiple testing, and shed light on the underlying genetic architecture.

We ran quality control steps first. To be specific, we filtered out all SNPs with a minor allele frequency < 0.05, those with a genotyping rate < 0.95, and those with a Hardy-Weinberg equilibrium test p-value $< 10^{-5}$. For testing polygenic effects (on chromosome level), we pruned SNPs with a criterion of linkage disequilibrium

$r^2 > 0.1$ using a sliding window of size 200 SNPs and a moving step of 20. For pathway-level analysis, we pruned SNPs with a criterion of linkage disequilibrium $r^2 > 0.8$ using a sliding window of size 50 SNPs and a moving step of 5. We imputed the missing SNPs via a Michigan Imputation Server (Das et al., 2016) with the 1000 Genomes Project European ancestry samples as the reference panel. For covariates, we included gender, years of education, handedness, age, and intracranial volume measured at baseline. To better demonstrate the possible power differences among the different tests, we applied the tests at either the chromosome or pathway level.

First, we conducted polygenic testing at the chromosome level. The family-wise nominal significance level was set at 0.05, yielding a $0.05/22 \simeq 0.0023$ significance cutoff for each chromosome after the Bonferroni adjustment. Table 2 shows some representative results for both asymptotics and parametric bootstrap-based p-values for each test. Most asymptotic p-values of the proposed SPU and aSPU tests were close to their parametric bootstrap-based ones, indicating good approximations by asymptotics. The aSPU test gave significant p-values (< 0.0023) for 5 chromosomes. In contrast, The HDGLM (Guo and Chen, 2016) yielded significant p-values for only two chromosomes. As expected, the p-values of HDGLM were close to that of SPU(2) since the two test statistics are similar. Perhaps due to dense and weak signals on these chromosomes, the supremum type test SPU(∞) was not significant in any chromosome while the burden test SPU(1) was often more significant. However, in

some situations, SPU(γ) with a larger γ might perform better. For example, for chromosome 5, perhaps due to moderately sparse and weak signals, SPU(3) gave the most significant p-value. Another example was for chromosome 14, SPU(3) yielded a significant result, while HDGLM gave a non-significant one. A meta-analysis of 74,046 individuals identified 2 SNPs at the genome-wide significance level on chromosome 14 (Lambert et al., 2013), validating that chromosome 14 was not a false positive finding by SPU(3). Due to its adaptiveness, the aSPU test often yielded more significant results than the HDGLM across the chromosomes.

Table 2: The p-values of various tests for ADNI data. The results outside and inside parentheses were calculated from the asymptotics- and parametric bootstrap-based methods, respectively.

Test	Chromosome (number of SNPs)			
	5 (3445)	13 (2071)	14 (1878)	21 (840)
SPU(1)	0.01 (0.01)	2×10^{-4} (6×10^{-4})	0.002 (0.002)	1×10^{-4} (2×10^{-4})
SPU(2)	0.03 (0.04)	0.11 (0.10)	0.25 (0.22)	0.15 (0.14)
SPU(3)	0.004 (0.003)	7×10^{-5} (7×10^{-4})	5×10^{-4} (2×10^{-3})	5×10^{-4} (2×10^{-3})
SPU(4)	0.11 (0.09)	0.14 (0.13)	0.30 (0.28)	0.33 (0.02)
SPU(5)	0.01 (0.02)	5×10^{-4} (3×10^{-3})	0.001 (0.005)	6×10^{-3} (0.01)
SPU(6)	0.32 (0.29)	0.22 (0.20)	0.28 (0.25)	0.38 (0.32)
SPU(∞)	0.95 (0.87)	0.66 (0.57)	0.07 (0.12)	0.27 (0.23)
aSPU	0.02 (0.03)	3×10^{-4} (9×10^{-4})	0.003 (0.006)	7×10^{-4} (5×10^{-4})
HDGLM	0.04 (0.04)	0.14 (0.12)	0.29 (0.25)	0.20 (0.17)

Next we conducted a pathway-based analysis. We retrieved a total of 214 pathways from the KEGG database (Kanehisa et al., 2009). As in practice (Network et al., 2015), we restricted our analysis to pathways of at most 200 genes and at least 10 genes, and excluded the pathways with less than 1000 SNPs, leading to 141 path-

ways for the following analysis. We set a $0.05/141 \simeq 3 \times 10^{-4}$ significance cutoff for each pathway after the Bonferroni adjustment. Figure 3 compares the p-values of the asymptotics- and parametric bootstrap-based methods, showing that the p-values of the former method were close to those of the latter, validating the good performance of the asymptotic results in Theorem 2 for real data analyses. The Pearson correlations of the p-values between the two methods ranged from 0.965 to 0.998. Table 3 shows 10 KEGG pathways with p-values less than 3×10^{-4} by either aSPU or GT or HDGLM. The three tests identified 10, 0, 1 significant pathways, respectively. The KEGG Alzheimer's disease pathway (hsa05010) can be treated as a true positive since the common variant in the APOE gene (one gene in the KEGG Alzheimer's disease pathway) alone explains 6% of total AD phenotypic variance (Ridge et al., 2013). For HSA05010 pathway, only the aSPU test gave a signficant p-value $< 3 \times 10^{-4}$, however, not by either GT (p-value= 0.0038) or HDGLM (p-value= 0.0014). Sporadic amyotrophic lateral sclerosis (ALS) is an age-associated disease and there are some evidence showing that ALS and AD are triggered by some common factors (Wang et al., 2014), while acute myeloid leukemia has been discovered to be associated with AD by other studies (Satoh, 2012), lending some support for other two identified pathways (HSA05014 and HSA05221). Perhaps due to very strong but sparse signals in these three pathways, aSPU could identified these three pathways while GT and HDGLM failed.

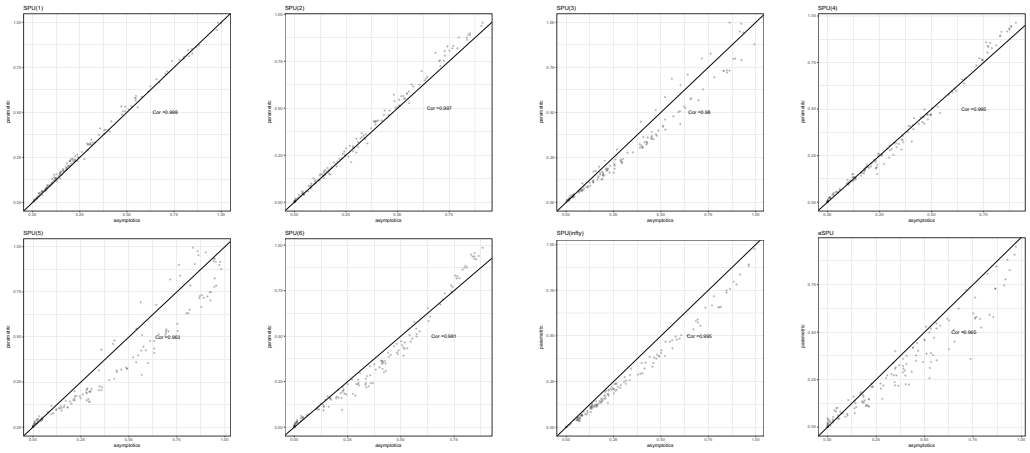

Figure 3: Comparison between the asymptotics- and the parametric bootstrap-based p-values of SPU(γ) and aSPU.

Table 3: Results of the ADNI data analysis: the significant KEGG pathways with p-values $< 3 \times 10^{-4}$ by any of aSPU, GT and HDGLM.

				p values		
KEGG ID	Pathway Name	# Genes	# SNPs	aSPU	GT	HDGLM
hsa05010	Alzheimer's disease	151	7251	0.0E+00	3.8E-03	1.4E-03
hsa05014	Amyotrophic lateral sclerosis	52	2503	0.0E+00	2.3E-03	3.2E-04
hsa05221	Acute myeloid leukemia	55	2024	0.0E+00	2.6E-03	7.6E-04
hsa04520	Adherens junction	72	6179	9.0E-09	4.4E-01	4.7E-01
hsa00071	Fatty acid degradation	40	1110	5.3E-08	1.6E-02	8.0E-03
hsa00830	Retinol metabolism	61	1256	2.1E-07	4.1E-03	7.9E-04
hsa00350	Tyrosine metabolism	38	1194	4.0E-07	7.7E-03	2.4E-03
hsa00982	Drug metabolism	70	1472	2.2E-05	3.6E-02	2.6E-02
hsa00534	Heparin	26	1630	6.4E-05	6.2E-04	1.1E-05
hsa00980	Metabolism of xenobiotics	68	1576	1.6E-04	9.5E-02	9.1E-02

In summary, the two real data applications here demonstrate that our proposed aSPU test was competitive and can be potentially useful in practice due to its adaptiveness.

5. Discussion

We have proposed a highly adaptive association test on a high-dimensional parameter in a GLM in the presence of a low-dimensional nuisance parameter. Its asymptotic null distribution is established, facilitating its asymptotic p-value calculations. At the first glance, the technical details of proving Theorems 1 and 2 are similar to those in a previous paper (Xu et al., 2016), however, the problem is more challenging here due to the presence of nuisance parameters.

As shown in both simulations and real data analyses, the proposed aSPU test is powerful across a wide range of scenarios considered. In comparison, two other existing tests, HDGLM (Guo and Chen, 2016) and GT (Goeman et al., 2011), based on the sum of squares of the score vector, performed similarly to SPU(2), all of which were powerful only in situations with moderately dense signals, but less powerful than some other SPU tests when the signals were either highly dense or highly sparse. In contrast, by combining multiple SPU tests, the aSPU test maintained high power across various scenarios. In addition to polygenic testing, we also applied the proposed aSPU test to pathway or gene set analysis, demonstrating its potential

usefulness in practice. An R package *GLMaSPU* implementing the proposed test is publicly available on GitHub and CRAN; to facilitate its use, we have also created an online website at `http://wuchong.org/GLMaSPU.html`.

Supplementary Material

The online supplementary material includes proofs of the theoretical results and additional simulation results.

Acknowledgement

The authors thank the reviewers and editors for helpful comments. This research was supported by NIH grants R01GM113250, R01HL105397 and R01HL116720, by NSF grants DMS-1712717 and SES-1659328, by NSA grant H98230-17-1-0308, and by the Minnesota Supercomputing Institute. CW was supported by a University of Minnesota Doctoral Dissertation Fellowship.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904).

References

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

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

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

Cai, T. T., Z. Ren, H. H. Zhou, et al. (2016). Estimating structured high-dimensional covariance and precision matrices: Optimal rates and adaptive estimation. *Electronic Journal of Statistics 10*(1), 1–59.

Chen, S. X., J. Li, and P.-S. Zhong (2014). Two-sample tests for high dimensional means with thresholding and data transformation. *arXiv preprint arXiv:1410.2848*.

Das, S., L. Forer, S. Schönherr, C. Sidore, A. E. Locke, A. Kwong, S. I. Vrieze, E. Y. Chew, S. Levy, M. McGue, et al. (2016). Next-generation genotype imputation service and methods. *Nature Genetics 48*(10), 1284–1287.

Donoho, D. and J. Jin (2015). Higher criticism for large-scale inference, especially for rare and weak effects. *Statistical Science 30*(1), 1–25.

Fahrmeir, L. and H. Kaufmann (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics 13*(1), 342–368.

Fan, J., R. Song, et al. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *The Annals of Statistics 38*(6), 3567–3604.

Goeman, J. J., S. A. Van De Geer, and H. C. Van Houwelingen (2006). Testing against a high dimensional alternative. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(3), 477–493.

Goeman, J. J., H. C. Van Houwelingen, and L. Finos (2011). Testing against a high-dimensional alternative in the generalized linear model: asymptotic type 1 error control. *Biometrika 98*(2), 381–390.

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

Jack, C. R., M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L Whitwell, C. Ward, et al. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging 27*(4), 685–691.

Jin, J. and T. Ke (2014). Rare and weak effects in large-scale inference: methods and phase diagrams. *arXiv preprint arXiv:1410.4578*.

Kanehisa, M., S. Goto, M. Furumichi, M. Tanabe, and M. Hirakawa (2009). KEGG

for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research 38*(Database issue), D355–D360.

Kim, J., Y. Zhang, and W. Pan (2016). Powerful and adaptive testing for multi-trait and multi-snp associations with GWAS and sequencing data. *Genetics 203*(2), 715–731.

Lambert, J.-C., C. A. Ibrahim-Verbaas, D. Harold, A. C. Naj, R. Sims, C. Bellenguez, G. Jun, A. L. DeStefano, J. C. Bis, G. W. Beecham, et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics 45*(12), 1452–1458.

Lan, W., H. Wang, and C.-L. Tsai (2014). Testing covariates in high-dimensional regression. *Annals of the Institute of Statistical Mathematics 66*(2), 279–301.

Lin, X., S. Lee, D. C. Christiani, and X. Lin (2013). Test for interactions between a genetic marker set and environment in generalized linear models. *Biostatistics 14*(4), 667–681.

McCullagh, P. and J. A. Nelder (1989). *Generalized linear models*, Volume 37. CRC press.

Morgenthaler, S. and W. G. Thilly (2007). A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test

(CAST). *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 615*(1), 28–56.

Network, T., P. A. S. of the Psychiatric Genomics Consortium, et al. (2015). Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nature Neuroscience 18*(2), 199–209.

Pan, W. (2009). Asymptotic tests of association with multiple snps in linkage disequilibrium. *Genetic Epidemiology 33*(6), 497–507.

Pan, W., J. Kim, Y. Zhang, X. Shen, and P. Wei (2014). A powerful and adaptive association test for rare variants. *Genetics 197*(4), 1081–1095.

Rao, B. P. (2009). Conditional independence, conditional mixing and conditional association. *Annals of the Institute of Statistical Mathematics 61*(2), 441–460.

Ridge, P. G., S. Mukherjee, P. K. Crane, J. S. Kauwe, et al. (2013). Alzheimer's disease: analyzing the missing heritability. *PloS One 8*(11), e79771.

Satoh, J.-i. (2012). Molecular network of microRNA targets in Alzheimer's disease brains. *Experimental Neurology 235*(2), 436–446.

Van de Geer, S., P. Bühlmann, Y. Ritov, and R. Dezeure (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics 42*(3), 1166–1202.

Voorman, A., A. Shojaie, and D. Witten (2014). Inference in high dimensions with the penalized score test. *arXiv preprint arXiv:1401.2678*.

Wang, X., J. Blanchard, I. Grundke-Iqbal, J. Wegiel, H.-X. Deng, T. Siddique, and K. Iqbal (2014). Alzheimer disease and amyotrophic lateral sclerosis: an etiopathogenic connection. *Acta Neuropathologica 127*(2), 243–256.

Wu, M. C., S. Lee, T. Cai, Y. Li, M. Boehnke, and X. Lin (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics 89*(1), 82–93.

Xu, G., L. Lin, P. Wei, and W. Pan (2016). An adaptive two-sample test for high-dimensional means. *Biometrika 103*(3), 609–624.

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

Zhong, P.-S. and S. X. Chen (2011). Tests for high-dimensional regression coefficients with factorial designs. *Journal of the American Statistical Association 106*(493), 260–274.