Statistica Sinica **29** (2019), 2163-2186
doi:https://doi.org/10.5705/ss.202017.0354

AN ADAPTIVE TEST ON HIGH-DIMENSIONAL PARAMETERS IN GENERALIZED LINEAR MODELS

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

[1] *University of Minnesota and* [2] *University of Michigan*

Abstract: Significance testing for high-dimensional generalized linear models (GLMs) has become increasingly important in various applications. However, existing methods are mainly based on a sum of the squares of the elements of the score vector and are only powerful under certain alternative hypotheses. In practice, the density of the true association pattern under an alternative hypothesis dictates whether existing tests are powerful. We propose an adaptive test on a high-dimensional parameter of a GLM (in the presence of a low-dimensional nuisance parameter) that maintains 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 data set to detect possible associations between Alzheimer's disease and gene pathways that have a large number of single nucleotide polymorphisms (SNPs). We implemented the proposed method in the R package *GLMaSPU*, which 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)) are increasingly being used in high-dimensional settings owing to the increase in the amount of to the surge of high-dimensional data in many fields, such as business and 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 a p-value $< 5 \times 10^{-8}$)

and, thus, is often underpowered. If we fail to identify any or a sufficient number of associated SNPs using a univariate test, we may wish to test a genetic marker set with possibly a large number of SNPs to both gain statistical power and enhance the 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 are widely used (McCullagh and Nelder (1989)). However, the power of such tests tends to diminish quite rapidly as p increases (Goeman, Van De Geer and Van Houwelingen (2006)). These tests even break down completely when $p > n$ because 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 (e.g., Goeman, Van De Geer and Van Houwelingen (2006); Goeman, Van Houwelingen and Finos (2011); Zhong and Chen (2011); Lan, Wang and Tsai (2014); Guo and Chen (2016)). In particular, Zhong and Chen (2011) proposed a modified F-test for high-dimensional linear regression models, allowing $p \to \infty$ as $n \to \infty$; Lan, Wang and Tsai (2014) extended the test to GLMs with a general random design matrix. In addition, Goeman, Van De Geer and Van Houwelingen (2006) proposed a test statistic for high-dimensional linear models and Goeman, Van Houwelingen and Finos (2011) derived its asymptotic distribution for a fixed p in GLMs. Guo and Chen (2016) further modified Goeman's test statistic (Goeman, Van Houwelingen and Finos (2011)) to a simpler form, allowing 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, Shojaie and Witten (2014)). However, the goal of such methods differs from ours of testing on a high-dimensional parameter. Thus, we do not discuss these methods further here.

Existing methods are based mainly on the sum of the squares of the elements of the score vector for the parameters of interest. Such methods 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, whereas tests based on the supremum of the score vector are more powerful. Importantly, as shown in the simulation section, there are intermediate situations in which neither type of tests is powerful. In practice, it is often unclear which type of test should be applied

because the underlying truth is unknown.

In this study, we develop an adaptive test that yields high statistical power under various high-dimensional scenarios, ranging from highly dense to highly sparse signal situations. The main idea is that, because 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 is powerful for a given situation. The proposed adaptive test then selects the test with the most significant result, including 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 to 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 method.

The rest of the paper is organized as follows. In Section 2, we review several existing tests. In Section 3, we describe the proposed the new adaptive test and study its asymptotic properties in contexts with and without nuisance parameters. The results of our simulation studies and real-data analyses are presented in Section 4. All technical 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. Existing Tests

Suppose we have n identical and independently distributed (i.i.d.) samples $\{(Y_i, Z_i, X_i) : i = 1, 2, \ldots, n\}$, 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 let $X_i = (X_{i1}, \ldots, X_{ip})$ be the p-dimensional variables of interest. Without loss of generality, we assume that $E(\mathbb{X}) = 0$, 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 the p-vector β and the 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 tests diminishes quite rapidly as the dimension p increases (Goeman, Van De Geer and Van Houwelingen (2006)). More importantly, in a high-dimensional situation with $p > n$, these tests break down completely, because the MLEs for the parameters no longer exist uniquely. Goeman, Van De Geer and Van Houwelingen (2006) derived the following test statistic for testing hypothesis (2.2), based on the score vector

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

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

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

where μ_0 is the expectation of Y under the null hypothesis. Goeman, Van De Geer and Van Houwelingen (2006) calculated the p-value of this test statistic using permutations or moment matching. Goeman, Van Houwelingen and Finos (2011) modified T_{Goe} with the following statistic:

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

where $\hat{\mu}_0$ and $\mathbb{D}$ are the MLE 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, Van Houwelingen and Finos (2011) derived its asymptotic null distribution for fixed p. Because the denominator of T_{GT} increases the variance, and thus adversely affects the power, Guo and Chen (2016) proposed the following test statistic:

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

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

Remark 1. To the best of our knowledge, most high-dimensional tests are based on a sum of squares of the elements of a 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. The SKAT is a similar test (Wu et al. (2011)).

3. Proposed Method

For the purpose of presentation, we first consider the case without nuisance parameters, followed by 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 notational simplicity, we write $S_{ij} = (Y_i - \mu_{0i})X_{ij}$ for $1 \le i \le n$ and $1 \le j \le p$. As 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 use U to construct 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 $w_j = U_j^{\gamma-1}$ is 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 $L(\gamma, \mu_0) \propto L(\gamma, \mu_0)^{1/\gamma} \to \max_{1\le j\le p} |(1/n)\sum_{i=1}^{n}(Y_i - \mu_{0i})X_{ij}|$; thus, we define $L(\infty, \mu_0)$ as

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

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 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 further details.

The class of SPU tests covers several tests used in GWASs as special cases. For example, for large n and small p, $L(2, \mu_0)$ is like the SKAT with a linear kernel (Wu et al. (2011)), and $L(1, \mu_0)$ is a burden test in genetic rare-variant association analyses (Morgenthaler and Thilly (2007)). As shown in simulations, if most variables of $\mathbb{X}$ are associated with the response Y, such as with similar effect sizes and the same association direction, then a burden test $L(1, \mu_0)$ yields

high statistical power. In contrast, with only moderately dense signals or with different association directions, $L(\gamma, \mu_0)$ with an even integer $\gamma \geq 2$ is more powerful. In particular, the supremum-based test statistic, $L(\infty, \mu_0)$, yields high statistical power if few variables are strongly associated with Y (i.e., a sparse nonzero components of β). In summary, the power of $L(\gamma, \mu_0)$ depends on the unknown true association pattern (i.e., value of β), such as the signal sparsity or magnitudes. To choose the most powerful test automatically, we propose the following adaptive test that combines the multiple tests:

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

where $P_{\text{SPU}(\gamma,\mu_0)}$ is the p-value of the $L(\gamma, \mu_0)$ test. For simplicity, we use $L(\gamma, \mu_0)$, $\text{SPU}(\gamma, \mu_0)$ and $\text{SPU}(\gamma)$ interchangeably. 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 a test statistics and no longer a genuine p-value. Thus, 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 choose a Γ set to maintain the 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, Liu and Xia (2014); from their Theorem 2, the power of the supremum-based test converges to one 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 Guo and Chen (2016). As default values, we recommend using $\gamma = 1, 2, \infty$ 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 γ_u such that $L(\gamma_u)$ gives similar results to that of $L(\infty)$; our simulation studies show that $\gamma_u = 6$ or 8 often suffices and that 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 analyses of rare variants with large n and small p. For simplicity, we use the same name "aSPU" for our proposed test here. Because 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 one; 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, Liu and Xia (2014) used the same assumptions (C1, C4, and C5) to derive 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, and is widely used in spatial statistics and time series. For high-dimensional two-sample mean testing, a similar mixing condition is used by Xu et al. (2016) and Chen, Li and Zhong (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 applies to 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 to estimate high-dimensional covariance matrices (Bickel and Levina (2008)). In addition, the correlations between the 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). Thus, the α-mixing assumption fits the application well, and, thus, is employed in this study.

We write $L(\gamma,\mu_0)=\sum_{j=1}^{p}L^{(j)}(\gamma,\mu_0)$, with $L^{(j)}(\gamma,\mu_0)=((1/n)\sum_{i=1}^{n}S_{ij})^{\gamma}$, then denote $\mu(\gamma)=\sum_{j=1}^{p}\mu^{(j)}(\gamma)$ as $\mu^{(j)}(\gamma)=E\left(L^{(j)}(\gamma,\mu_0)\right)$, and $\sigma^2(\gamma)=\operatorname{var}(L(\gamma,\mu_0))$.

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

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

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

Proposition 1 follows directly from Lemma 1 in the online Supplementary Material.

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

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

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 between $L(\gamma,\mu_0)$.

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,*

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

(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. This yields 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 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 of* $L(\infty, \mu_0)$*. 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 to calculate 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 a 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. For example, we may be interested only in genetic effects, while adjusting for demographic variables. Hence, the coefficients of the 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. Because μ_0 is unknown, we use $\hat{\mu}_0$ and the test statistic $L(\gamma, \hat{\mu}_0)$. 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 $E(X_{ij}|\mathbb{Z}) \neq 0$ holds for $j \in P_0$ only with the size of P_0 (p_0), satisfying $p_0 = O(p^{\eta})$ for a small positive η. We further assume a 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 $\epsilon_{0i} = Y_i - \mu_{0i}$, $1 \le i \le n$. 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 $\sup\{|Pr(A \cap B|\mathcal{F}) - Pr(A|\mathcal{F})Pr(B|\mathcal{F})| : 1 \le t < p, A \in \chi_1^t, B \in \chi_{t+s}^{\infty}\}$, where $\mathcal{F}$ is a sub-σ-algebra of W. We assume $W = \{W^{(j)} = (X_{ij}, i = 1, \ldots, n) : j \ge 1\}$ is conditionally α-mixing, given $\mathbb{Z}$ such that $\alpha_{W|\sigma(\mathbb{Z})}(s) \le M\delta^s$, where $\delta \in (0, 1)$ and M is some constant.

Remark 6. Assumption C6 states that the dimension of the nuisance parameters, q, is fixed as $n \to \infty$, which is appropriate in many applications, including the GWASs of interest here. However, this assumption may not be appropriate in some applications. For example, when 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 zero, partially making it reasonable to assume $E[X_{ij}|\mathbb{Z}] \ne 0$ for $j \in P_0$ only, 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, Song et al. (2010) and assumption 3.3 in Guo and Chen (2016). Assumption C8 is an updated version of C4 and somewhat restrictive, but is needed to prove Theorem 2. Note that, instead of considering only a sum-of-squares-type statistic (with $\gamma = 2$) similar to the HDGLM (Guo and Chen (2016)), 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 shown in the simulations, the asymptotic distribution still performs well for more general high-dimensional situations. We leave this interesting problem to future work. Conditional α-mixing is introduced by Rao (2009) and assumption C9 is an updated version of C2 to adjust for the case of the nuisance parameters.

Although the estimated parameter $\hat{\alpha}$ complicates derivations, we have the following theorem, which is 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 i.i.d., which makes σ_{kj} well defined (unconditionally); we also derive the unconditional version of the asymptotic null distribution.

Because $\mu(\gamma)$, $\sigma(\gamma)$, and R can be approximated using Propositions 1–3, respectively, the p-values for individual $L(\gamma, \hat{\mu}_0)$ can be calculated using 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 that 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 that 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 using 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 statistics from the data, and calculate the p-values for t_O and t_E as $p_O = Pr[\max_{\text{odd } \gamma\in\Gamma'} |\{L(\gamma, \hat{\mu}_0) - \mu(\gamma)\}/\sigma(\gamma)| > t_O]$ and $p_E = Pr[\max_{\text{even } \gamma\in\Gamma'}\{L(\gamma, \hat{\mu}_0) - \mu(\gamma)\}/\sigma(\gamma) > t_E]$, respectively. We can use function *pmvnorm()* in the R package *mvtnorm* to calculate the multivariate normal tail probabilities for 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. Here, 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, Ren and Zhou (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 the covariance matrix $\mathbf{\Sigma}$. Specifically, we first calculate the sample covariance matrix $\mathbb{S} = (s_{ij})$, where $s_{ij} = (1/(n-1))\sum_{k=1}^{n}(Y_k - \hat{\mu}_{0k})^2 X_{ki}X_{kj}$. Then, we further calculate the bandable covariance matrix with bandwidth k_n as $\hat{\mathbf{\Sigma}}_{k_n} = (s_{ij}I(|i-j| \le k_n))$. From a theoretical perspective, the optimal bandwidth k_n and the minimax risk rates of $\hat{\mathbf{\Sigma}}_{k_n}$ have been studied in Bickel and Levina (2008). Because a theoretically optimal k_n is determined by some unknown hyperparameters, 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 $\hat{\sigma}^2(\gamma) = \{1+o(1)\}\sigma^2(\gamma)$, respectively, for properly chosen $k_n = o(n^{1/2})$.

With a relatively small sample size, five-fold cross-validation may select a smaller than optimal bandwidth, yielding an underestimated $\hat{\sigma}^2(\gamma)$ and a smaller p-value. As an alternative, we propose using 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, we 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. The accuracy of the usual resampling methods is bounded by the number of resampling B. Thus, a large B is needed to calculate a very small p-value. In contrast, we can use a relatively small B to calculate $\hat{\mu}(\gamma)$, $\hat{\sigma}^2(\gamma)$, and R and then an asymptotic p-value. Although they estimate the mean and covariance matrix differently, the above two methods still use the asymptotics to calculate the p-values and, hence, are referred to as asymptotics-based methods in this paper. In contrast, we can also simply use the parametric bootstrap to calculate the p-values (without using the asymptotic results directly), which is 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 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) because they use similar test statistics. However, in some situations, $L(2, \hat{\mu}_0)$ is not as powerful as other $L(\gamma, \hat{\mu}_0)$ tests. 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. Owing 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, namely the HDGLM (Guo and Chen (2016)) and the GT (Goeman, Van Houwelingen and Finos (2011)), owing to their popularity and the availability of their computer code.

We set the sample size as $n = 200$ and the dimension of β as $p = 2{,}000$, 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 $X_i \sim N(0, \boldsymbol{\Xi})$ for $i = 1, 2, \ldots, n$. We show the results with unit variances and a blocked first-order autoregressive correlation matrix $\boldsymbol{\Xi} = (\boldsymbol{\Xi}_{ij})$, with $\boldsymbol{\Xi}_{ij} = 0.4^{|i-j|}$ 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 the alternative hypothesis H_1, respectively. Here, we focus on the results for a binary outcome because in our real-data application, the response is binary and more challenging than that of a continuous outcome. Under H_1, $\lfloor ps \rfloor$ elements of β were set to be nonzero, where $s \in [0, 1]$ controls the degree of signal sparsity. We varied s to mimic varying sparsity levels: from highly sparse signals at $s = 0.001$, to less sparse, to moderately dense at $s = 0.1$, and finally to dense and highly dense signals at $s = 0.7$. The indices of nonzero 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 method (Pan et al. (2014))

Table 1. Empirical type I error rates and power (%) of various tests in simulations with $n = 200$ and $p = 2{,}000$. The sparsity parameter is $s = 0.1$, leading to 200 nonzero elements in β with a constant value c. The results outside and inside parentheses are calculated using the 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)

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 $\mathbf{\Sigma}$, and then calculated the means and covariances of the SPU test statistics based on Propositions 1–3. For each setup, we simulated 1,000 data sets and averaged the test results. The nominal significance level was set to $\alpha = 0.05$. For the aSPU test, the candidate set of γ was, by default, set to $\Gamma = \{1, 2, \ldots, 6, \infty\}$.

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 using the asymptotics- and parametric bootstrap-based methods, respectively; the results based on the two methods were very similars, 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 $p = 4{,}000$. The results for these simulation settings were similar to those shown in Table 1, thus, are relegated to the Supplementary Material, Tables S1–S5.

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 nonsparsity 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 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

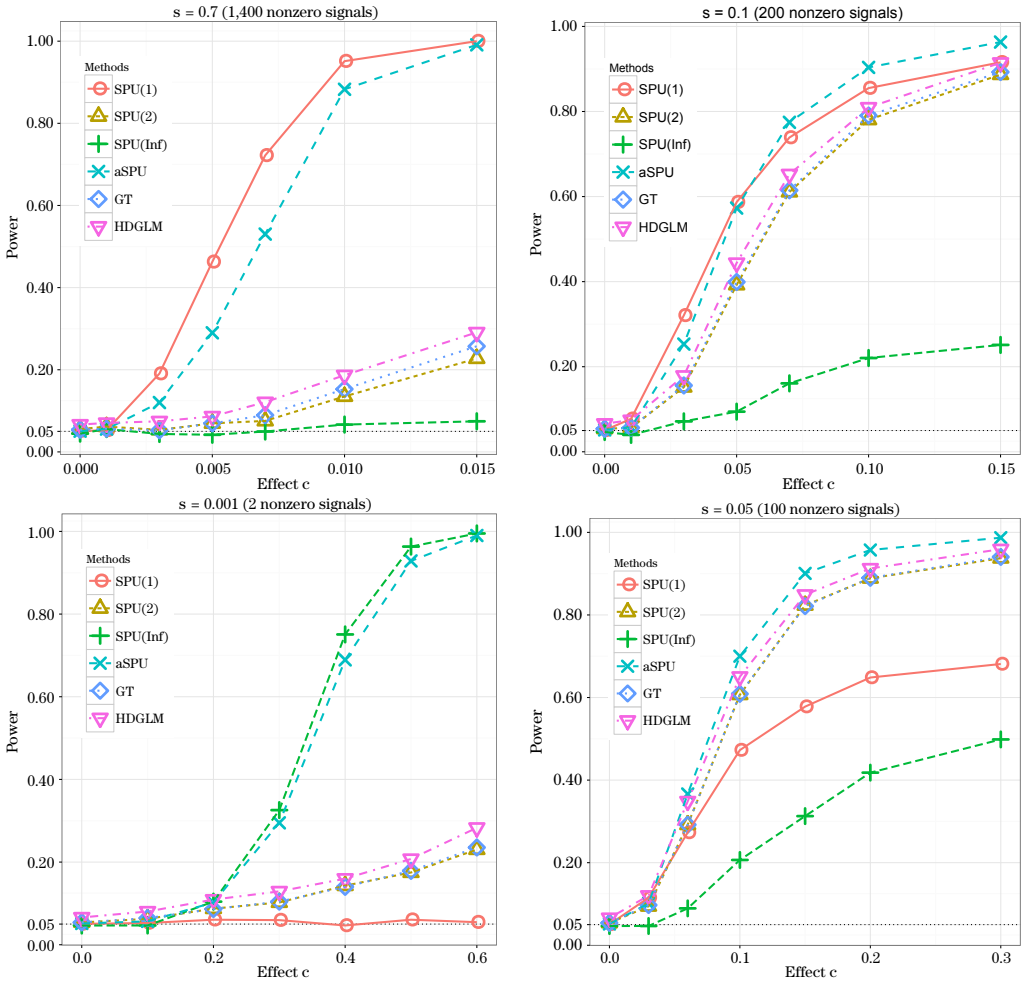

Figure 1. Empirical power of the SPU(1), SPU(2), SPU(∞), aSPU, GT (Goeman, Van Houwelingen and Finos (2011)), and HDGLM (Guo and Chen (2016)). The signal sparsity parameter s varies from 0.001 to 0.7. We set $n = 200$ and $p = 2{,}000$.

followed by the SPU(1) and SPU(2) tests. At $s = 0.7$, the aSPU test remained the best, and the SPU(1) test was more powerful than the sum-of-squares-type and supremum-type tests. Under all of the situations considered, the aSPU consistently maintained high power, being either the best or close to the best.

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.

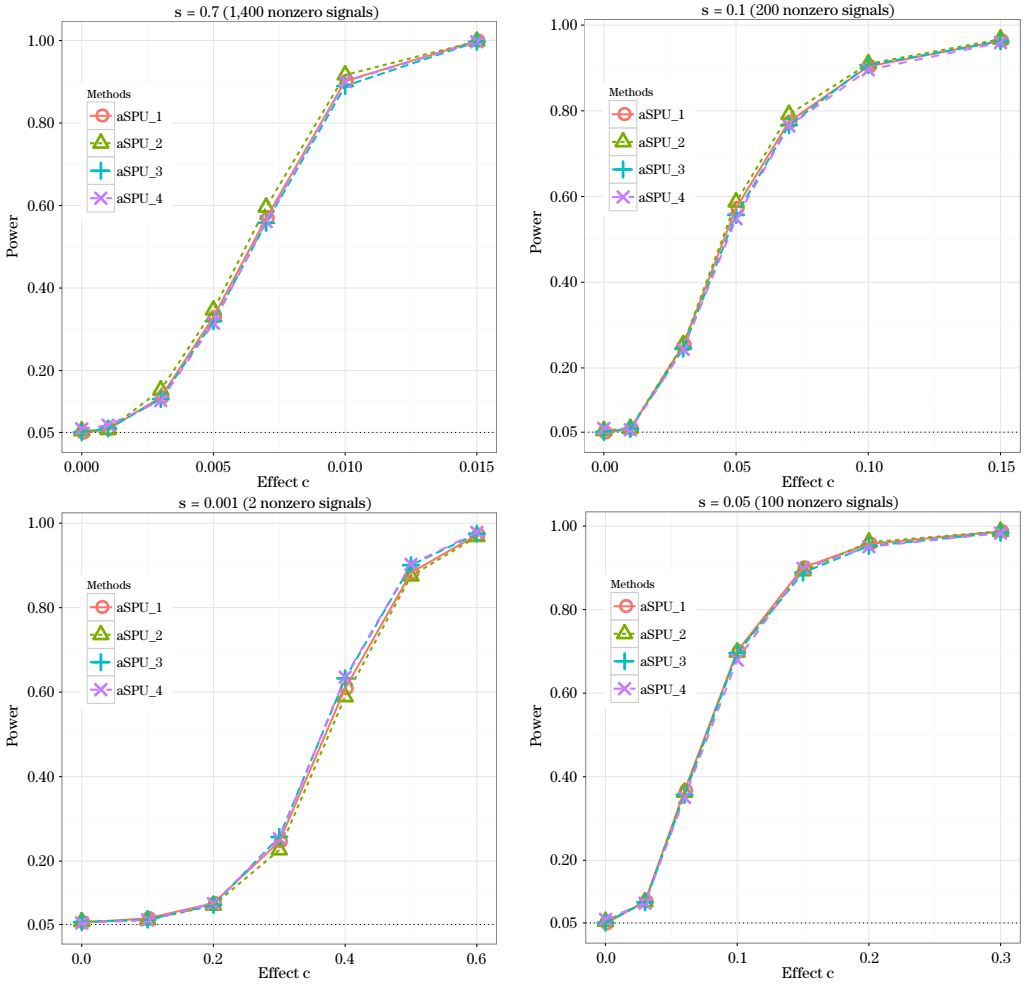

Figure 2. Empirical power of the aSPU with different Γ set. aSPU_1, aSPU_2, aSPU_3, aSPU_4 represent an aSPU with $\Gamma_1 = \{1, 2, \ldots, 4, \infty\}$, $\Gamma_2 = \{1, 2, \ldots, 6, \infty\}$, $\Gamma_3 = \{1, 2, \ldots, 8, \infty\}$, and $\Gamma_4 = \{1, 2, \ldots, 10, \infty\}$, respectively. The signal sparsity parameter s varies from 0.001 to 0.7. We set $n = 200$ and $p = 2{,}000$.

To further study the impact of the covariance structures, we considered the following two structures used in Cai, Liu and Xia (2014). The first is a block diagonal structure: $\boldsymbol{\Xi} = (\sigma^*_{i,j})$, with $\sigma^*_{i,i} = 1$, $\sigma^*_{i,j} = 0.8$ for $2(k-1)+1 \le i \ne j \le 2k$ and $k = 1, \ldots, [p/2]$, and $\sigma^*_{i,j} = 0$ otherwise. The second is a nonsparse structure: let $\boldsymbol{\Xi}^+ = (\sigma^+_{i,j})$. with $\sigma^+_{i,i} = 1$ and $\sigma^+_{i,j} = |i-j|^{-5}/2$, for $i \ne j$, 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, $\boldsymbol{\Xi} = \mathbb{D}^{1/2}\boldsymbol{\Xi}^+\mathbb{D}^{1/2}$.

For these two covariance structures, the results of the asymptotic approximation and power comparison were similar to those in Table 1 and Figure 1. Therefore, these are relegated to the Supplementary Material, Tables S6–S13 and Figures S1–S2. Finally, we also considered a continuous outcome Y; again, the results were similar (Supplementary Material, Table S14).

In summary, because of its adaptiveness, the aSPU test either achieved the highest power or was close to the best 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 millions of people worldwide. 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. The Principal Investigator of this initiative is Michael W. Weiner, VA Medical Center and University of California. The major goal of the 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, aged between 55 and 90, to participate in the research. For latest information, see `www.adni-info.org`.

One objective of the ADNI is to reveal a person's genetic susceptibility to AD. Owing to the relatively small sample size and the usually small genetic effect sizes, applying a univariate test to the ADNI data failed to identify any SNP that passed the genome-wide significance level at 5×10^{-8} (Kim, Zhang and Pan (2016)). Furthermore, an even much larger meta-analysis of 74,046 individuals 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 owing to effect aggregation and a reduced burden of multiple testing, and to reveal the underlying genetic architecture.

We ran quality control steps first. Specifically, 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 $< 10^{-5}$. To test the polygenic effects (on chromosome level), we pruned the SNPs using a criterion of linkage disequilibrium $r^2 > 0.1$, with a sliding window of size 200 SNPs and a moving step of 20. For the pathway-level analysis, we pruned the SNPs using a criterion

of linkage disequilibrium $r^2 > 0.8$, with a sliding window of size 50 SNPs and a moving step of 5. We imputed the missing SNPs using the Michigan Imputation Server (Das et al. (2016)) with the 1,000 Genomes Project European ancestry samples as the reference panel. We used ADNI1 data of 756 individuals. We assigned individuals with AD or MCI as cases ($Y = 0$) and healthy individuals as controls ($Y = 1$). For the covariates, we included gender, years of education, handedness, age, and intracranial volume measured at the baseline. To better demonstrate the possible power differences between the 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 representative results for the 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 values, indicating good approximations by the asymptotics. The aSPU test gave significant p-values (< 0.0023) for five chromosomes. In contrast, the HDGLM (Guo and Chen (2016)) yielded significant p-values for only two chromosomes. As expected, the p-values of the HDGLM were close to those of SPU(2), because the two test statistics are similar. Perhaps owing to the 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 owing to the moderately sparse and weak signals, SPU(3) gave the most significant p-value. Another example is that of chromosome 14, where SPU(3) yielded a significant result, while the HDGLM gave a nonsignificant one. A meta-analysis of 74, 046 individuals identified two 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). Owing to its adaptiveness, the aSPU test often yielded more significant results than those of HDGLM across the chromosomes.

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 (The Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium (2015)), we restricted our analysis to pathways of at most 200 genes and at least 10 genes, and excluded those pathways with fewer than 1,000 SNPs, leading to 141 pathways for the analysis. We set a significance cutoff of

Table 2. The p-values of various tests for the 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)

$0.05/141 \simeq 3 \times 10^{-4}$ for each pathway after the Bonferroni adjustment. Figure 3 compares the p-values of the asymptotics- and parametric bootstrap-based methods, showing that those 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 aSPU, GT, or the HDGLM. The three tests identified 10, 0, and 1 significant pathways, respectively. The KEGG Alzheimer's disease pathway (hsa05010) can be treated as a true positive because the common variant in the APOE gene (one gene in the KEGG AD pathway) alone explains 6% of total AD phenotypic variance (Ridge et al. (2013)). For the HSA05010 pathway, the aSPU test gave a signficant p-value $< 3 \times 10^{-4}$, whereas neither the GT (p-value $= 0.0038$) nor the HDGLM (p-value $= 0.0014$) yielded significant values. Sporadic amyotrophic lateral sclerosis (ALS) is an age-associated disease and there is some evidence that ALS and AD are triggered by common factors (Wang et al. (2014)). Furthermore, acute myeloid leukemia has been discovered to be associated with AD by other studies (Satoh (2012)), supporting for other two identified pathways (HSA05014 and HSA05221). Perhaps owing to very strong, but sparse signals in these three pathways, aSPU identified these three pathways, while the GT and the HDGLM failed to do so.

In summary, the two real-data applications here demonstrate that our proposed aSPU test is competitive and potentially useful in practice, owing to its adaptiveness.

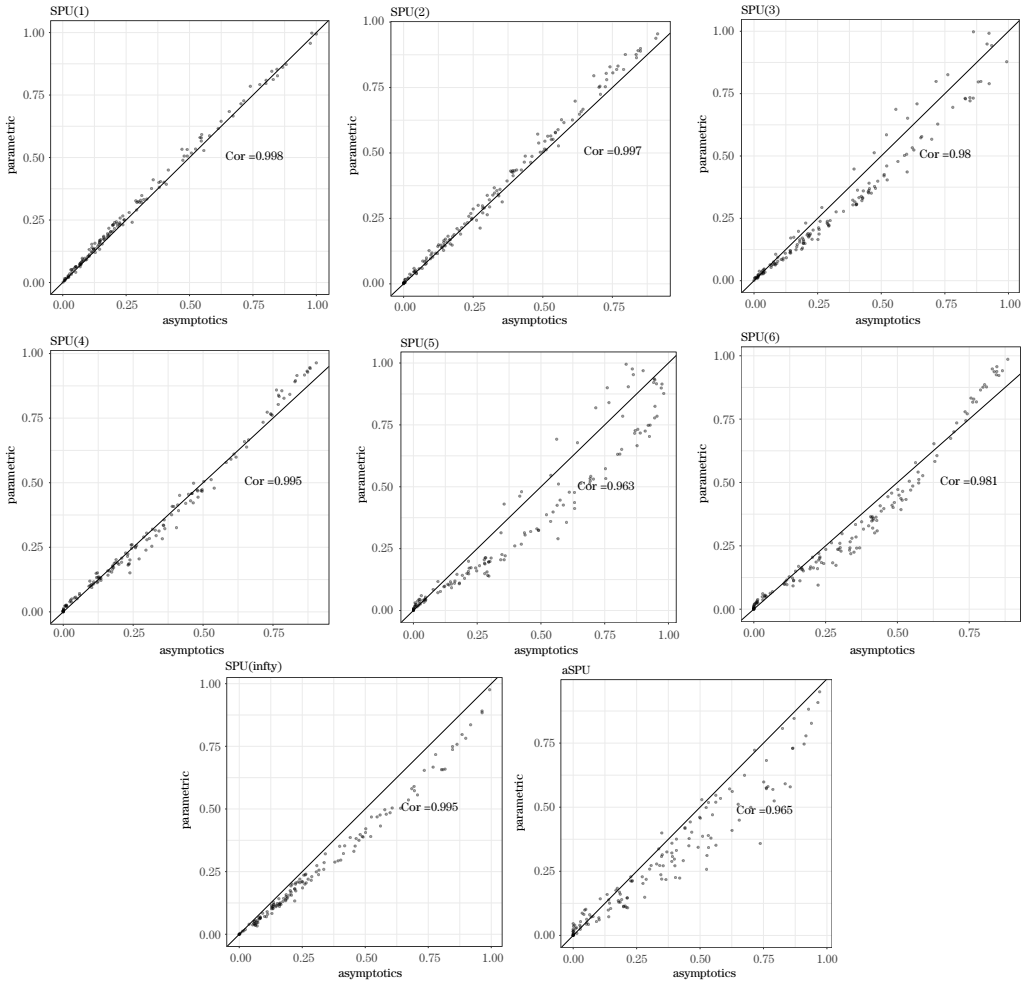

Figure 3. Comparison between the asymptotics- and the parametric bootstrap-based p-values of SPU(γ) and aSPU.

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 first glance, the proofs of Theorems 1 and 2 are similar to those of (Xu et al. (2016)). However, the problem is more challenging here owing to the presence of nuisance parameters.

As shown in both the simulations and in the real-data analyses, the proposed aSPU test is powerful across a wide range of scenarios. In comparison, two

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

				p values		
KEGG ID	Pathway Name	# Genes	# SNPs	aSPU	GT	HDGLM
hsa05010	Alzheimer's disease	151	7,251	0.0E+00	3.8E-03	1.4E-03
hsa05014	Amyotrophic lateral sclerosis	52	2,503	0.0E+00	2.3E-03	3.2E-04
hsa05221	Acute myeloid leukemia	55	2,024	0.0E+00	2.6E-03	7.6E-04
hsa04520	Adherens junction	72	6,179	9.0E-09	4.4E-01	4.7E-01
hsa00071	Fatty acid degradation	40	1,110	5.3E-08	1.6E-02	8.0E-03
hsa00830	Retinol metabolism	61	1,256	2.1E-07	4.1E-03	7.9E-04
hsa00350	Tyrosine metabolism	38	1,194	4.0E-07	7.7E-03	2.4E-03
hsa00982	Drug metabolism	70	1,472	2.2E-05	3.6E-02	2.6E-02
hsa00534	Heparin	26	1,630	6.4E-05	6.2E-04	1.1E-05
hsa00980	Metabolism of xenobiotics	68	1,576	1.6E-04	9.5E-02	9.1E-02

other existing tests, the HDGLM (Guo and Chen (2016)) and GT (Goeman, Van Houwelingen and Finos (2011)), based on the sum of the squares of the elements of a score vector, performed similarly to SPU(2). The latter tests 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 a pathway or a 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 Materials

The online Supplementary Material includes proofs of the theoretical results and additional simulation results.

Acknowledgment

The authors thank the reviewers and editors for their 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. The data collection and sharing for this project were funded by the Alzheimer's Dis-

ease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904).

References

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

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

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

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

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

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

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

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

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

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

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

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

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

Jin, J. and Ke, T. (2014). Rare and weak effects in large-scale inference: methods and phase diagrams. *Statistica Sinica* **26**, 1–34.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. and Hirakawa, M. (2009). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Research* **38**, D355–D360.

Kim, J., Zhang, Y. and Pan, W. (2016). Powerful and adaptive testing for multi-trait and

multi-snp associations with GWAS and sequencing data. *Genetics* **203**, 715–731.

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

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

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

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Vol. 37. CRC press.

Morgenthaler, S. and Thilly, W. G. (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**, 28–56.

The Network and Pathway Analysis Subgroup of the Psychiatric Genomics Consortium (2015). Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nature Neuroscience* **18**, 199–209.

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

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

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

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

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

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

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

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

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

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

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

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

Division of Biostatistics, University of Minnesota, Minneapolis, MN 55455, USA.
E-mail: wuxx0845@umn.edu
Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA.
E-mail: gongjun@umich.edu
Division of Biostatistics, University of Minnesota, Minneapolis, MN 55455, USA.
E-mail: weip@biostat.umn.edu

(Received January 2017; accepted February 2018)