Tests for High Dimensional Generalized Linear Models

Chen, Song Xi and Guo, Bin

Peking Univeristy and Iowa State University, Peking Univeristy

2014

Tests for High Dimensional Generalized Linear Models

Song Xi Chen and Bin Guo

Peking University and Iowa State University, and Peking University

April 17, 2014

Abstract

We consider testing regression coefficients in high dimensional generalized linear models. By modifying a test statistic proposed by Goeman et al. (2011) for large but fixed dimensional settings, we propose a new test which is applicable for diverging dimension and is robust for a wide range of link functions. The power properties of the tests are evaluated under the setting of the local and fixed alternatives. A test in the presence of nuisance parameters is also proposed. The proposed tests can provide p-values for testing significance of multiple gene-sets, whose usefulness is demonstrated in a case study on an acute lymphoblastic leukemia dataset.

Key words: Generalized Linear Model; Gene-Sets; High Dimensional Covariate; Nuisance Parameter; U-statistics.

1 Introduction

The generalized linear models (McCullagh and Nelder, 1989) are widely used in many fields of statistical applications. The surge of high dimensional data collection and analysis in bioinformatics and related studies have led to the use of generalized linear models in high dimensional settings. The high dimensionality can arise at least in two forms. One is in the various multiple response variables but with low or fixed dimensional covariates where the responses represent readings for large number of genes and the covariates represent certain design and demographic variables. Another is to have low dimensional response (for instance indicator for a disease) but high dimensional covariates representing genes expressions levels. Research on the first form of high dimensionality includes Auer and Doerge (2010) and Lund et al. (2012) in the context of next generation sequencing data. The current paper will be focused on the latter case where the high dimensionality is associated with the covariates. Statistical inference for the generalized

linear models under the high dimensional settings has been the focus of latest research. van de Geer (2008) considered variable selection via a LASSO approach. Fan and Song (2010) and Chang et al. (2013) proposed approaches via the sure independence screening of Fan and Lv (2008).

The focus of the paper is on testing for significance of the regression coefficients in high dimensional generalized linear models, which is of important interest to practitioners, for instance in the context of discovering significant gene-sets. The inferential context of gene-set testing encounters both high dimensionality and multiplicity, as genes in different gene-sets can overlap. These two features call for methods which can produce the p-value for the significance of each gene-set, which is an aim of the current paper in the context of the generalized linear models. For fixed dimensional data, the likelihood ratio test and the Wald test have been popular choices as elaborated in McCullagh and Nelder (1989). However, the high dimensionality renders the inapplicability of these tests. There are a set of published works on testing for the coefficients of high dimensional linear regression for the large p (dimension), small n (sample size) paradigm, which include the tests proposed in Goeman et al. (2006) for an empirical Bayesian formulation, Zhong and Chen (2011) that accommodates the factorial designs, and in Lan et al. (2014) that allows testing on subsets of the regression coefficients. There are also works on the post-variable selection inference associated LASSO and other variable selection methods for the linear models under the sparsity assumption, see Berk et al. (2013), Lee et al. (2014), Taylor et al. (2014), van de Geer et al. (2013) and Zhang and Zhang (2014).

In this paper, we consider testing for high dimensional regression coefficient for the generalized linear models without assuming the non-zero coefficients are sparse. In an important development, Goeman et al. (2011) proposed a test for the coefficients of high dimensional generalized linear models in the presence of nuisance parameters. The test has provided a much needed tool for performing multivariate tests when the conventional likelihood ratio and the Wald tests are not applicable. While allowing for the dimension p larger than the sample size n, the test of Goeman et al. (2011) was formulated for fixed p.

We propose tests for the entire regression coefficients and part of the regression coefficients in the presence of nuisance parameters for high dimensional generalized linear models with diverging p. We modify the test statistic of Goeman et al. (2011) by removing the denominator in their ratio statistic as well as some terms in the numerator. The modification is designed to make the tests operational with accurate size and reasonable power when p diverges along with the sample size. Our analysis shows that this modification is critical for models whose inverse link functions have unbounded derivative like the log link in the Poisson or Negative Binomial regression. For models whose links are bounded or with bounded derivative, like the logit and identity links, the test of Goeman et al. (2011) is found to be valid with robust power under diverging dimensions, and is asymptotically equivalent to the proposed tests. The proposed tests are studied by both theoretical analysis and numerical simulations. And they are applied to find significant gene-sets in an empirical study on an acute lymphoblastic leukemia dataset. It is shown in the case study that the p-values produced from the proposed tests when used in conjunction with a proper control on the False Discovery Rate (Benjamini and Hochberg, 1995) can lead to selecting

significant gene-sets in the context of high dimensionality and multiplicity.

The paper is organized as follows. In Section 2, we review the inferential setting for the generalized linear models. Section 3 considers Goeman et al. (2011)'s test for diverging p, which motivates our proposal for the global test in Section 4 and the test with nuisance parameters in Section 5. Results from simulation studies are reported in Section 6. Section 7 presents the case study on the acute lymphoblastic leukemia dataset. All technical details are relegated to the Appendix.

2 Models and existing test

Let Y be a response variable to a p-dimensional covariate X. The generalized linear models (McCullagh and Nelder, 1989) provide a rich collection of specifications for the conditional mean of Y given X. Although they are intimately connected to the exponential family of distributions, a more general view can be attained via the semiparametric quasi-likelihood of Wedderburn (1974).

Conditioning on the covariate X, there exists a monotone function $g(\cdot)$ and a non-negative function $V(\cdot)$ such that

$$E(Y|X) = \mu(\beta) = g(X^{\mathrm{T}}\beta) \quad \text{and} \quad \mathrm{Var}(Y|X) = V\{\mu(\beta); \phi\}, \tag{2.1}$$

where β is a p-dimensional regression coefficient, $g^{-1}(\cdot)$ is called the link function and ϕ is a dispersion parameter.

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be the independent copies of (X, Y). The maximum quasi-likelihood estimator $\hat{\beta}_n$ of β can be obtained by solving the quasi-likelihood score equation:

$$\ell_n(\beta) = \sum_{i=1}^{n} \frac{\{Y_i - g(X_i^{\mathrm{T}}\beta)\} g'(X_i^{\mathrm{T}}\beta) X_i}{V\{\mu_i(\beta); \hat{\phi}\}} = 0 \tag{2.2}$$

where $\mu_i(\beta) = g(X_i^{\mathrm{T}}\beta)$ and $\hat{\phi}$ is an estimator of ϕ, which can be obtained via the method of moment (for instance that given in Chen and Cui, 2003). When the variance function is multiplicative with respect to ϕ, say $V\{\mu(\beta); \phi\} = \phi V\{\mu(\beta)\}$, there is no need to carry out the initial estimation of ϕ for the inference on β. The consistency and asymptotic normality of $\hat{\beta}_n$ are well established for fixed dimensional covariate (McCullagh and Nelder, 1989).

Let $\beta = (\beta^{(1)\mathrm{T}}, \beta^{(2)\mathrm{T}})^{\mathrm{T}}$ be a partition of the coefficient vector and $X_i = (X_i^{(1)\mathrm{T}}, X_i^{(2)\mathrm{T}})^{\mathrm{T}}$ be the corresponding partition of the covariates, where $\beta^{(1)}$ and $X_i^{(1)}$ are p_1-dimensional, $\beta^{(2)}$ and $X_i^{(2)}$ are p_2-dimensional, and $p_1 + p_2 = p$. Suppose one is interested in testing a hypothesis

$$H_0 : \beta^{(2)} = \beta_0^{(2)} \quad \text{versus} \quad H_1 : \beta^{(2)} \neq \beta_0^{(2)}$$

on the effect of the covariate $X_i^{(2)}$ while treating $\beta^{(1)}$ as the nuisance parameter.

When the dimensions p_1 and p_2 are fixed, modified Wald and the score tests based on the asymptotic Chi-square approximations (Fahrmeir and Tutz, 1994) can be performed to test the

above hypothesis. However, the high dimensionality often requires that $p_2 > n$, see Pan (2009). When $p_2 > n$, the conventional Wald or the likelihood ratio tests are no longer applicable since the invertibility of the information matrix is not attainable and the maximum likelihood estimators for the parameters may not be obtained.

Goeman et al. (2011) considered the following test formulation in the case of $p_2 > n$ with $g^{-1}(\cdot)$ being the canonical link. To make the discussion more generally applicable, non-canonical links are considered via $\psi(X_i, \beta_0, \phi) = g'(X_i^{\mathrm{T}}\beta_0)/V\{\mu_i(\beta_0); \phi\}$ where g' denotes the derivative of g. The canonical link means $\psi(X_i, \beta_0, \phi)$ is a constant. Using the general $\psi(\cdot)$ function does not alter the formulation of Goeman et al. (2011)'s test.

Let $\hat{\beta}_0^{(1)}$ and $\hat{\phi}_0$ be the estimators of the nuisance parameters $\beta^{(1)}$ and ϕ under H_0, $\hat{\beta}_0 = (\hat{\beta}_0^{(1)\mathrm{T}}, \beta_0^{(2)\mathrm{T}})^{\mathrm{T}}$, $\hat{\mu}_{0i} = \mu_i(\hat{\beta}_0)$, $\widehat{\mu}_0 = (\hat{\mu}_{01}, \ldots, \hat{\mu}_{0n})^{\mathrm{T}}$ and $\widehat{\Psi}_0 = \{\psi(X_1, \hat{\beta}_0, \hat{\phi}_0), \ldots, \psi(X_n, \hat{\beta}_0, \hat{\phi}_0)\}^{\mathrm{T}}$. Moreover, let $\mathbb{X}^{(2)} = (X_1^{(2)}, \ldots, X_n^{(2)})^{\mathrm{T}}$, $\mathbb{Y} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$ and $\mathbb{D}$ be the $n \times n$ diagonal matrix that collects the diagonal elements of $\mathbb{X}^{(2)}\mathbb{X}^{(2)\mathrm{T}}$. The test statistic of Goeman et al. (2011) is

$$\widehat{S}_n = \frac{\{(\mathbb{Y} - \widehat{\mu}_0) \circ \widehat{\Psi}_0\}^{\mathrm{T}}\mathbb{X}^{(2)}\mathbb{X}^{(2)\mathrm{T}}\{(\mathbb{Y} - \widehat{\mu}_0) \circ \widehat{\Psi}_0\}}{\{(\mathbb{Y} - \widehat{\mu}_0) \circ \widehat{\Psi}_0\}^{\mathrm{T}}\mathbb{D}\{(\mathbb{Y} - \widehat{\mu}_0) \circ \widehat{\Psi}_0\}} \tag{2.3}$$

where $A \circ B = (a_{ij}b_{ij})$ for matrices $A = (a_{ij})$ and $B = (b_{ij})$. Note that, under the null hypothesis, the score function of $\beta^{(2)}$ is

$$\ell_2(\hat{\beta}_0^{(1)}, \beta_0^{(2)}) = \mathbb{X}^{(2)\mathrm{T}}\{(\mathbb{Y} - \widehat{\mu}_0) \circ \widehat{\Psi}_0\}.$$

Hence, the numerator of $\widehat{S}_n$ is a quadratic form of the score function, which will be small (large) when the null hypothesis is true (not true). The denominator is a plug-in estimator to the mean of the numerator for standardization.

3 Goeman et al. (2011)'s test when $p \to \infty$

The proposal of Goeman et al. (2011) was formulated for fixed dimension p while allowing $p > n$. We analyze in this section its properties under the regime of diverging p as $n \to \infty$. It will be shown that the test of Goeman et al. (2011) remains powerful for diverging p when either g or g' is bounded. At the same time, we reveal a loss of power for the test for some link functions. The analysis will provide useful insight on how to construct better tests for the case of $p \to \infty$.

To make the discussion focused while still being relevant, we concentrate on testing the global hypothesis in the absence of nuisance regression parameters, namely

$$H_0 : \beta = \beta_0 \quad \text{versus} \quad H_1 \quad \beta \neq \beta_0.$$

To simplify our analysis, we assume $E(X) = 0$ without loss of generality as otherwise X can be re-centered by its mean. Throughout the paper, we denote $\Sigma_{\mathrm{X}} = \mathrm{cov}(X)$, $\epsilon = Y - g(X^{\mathrm{T}}\beta)$, $\epsilon_0 = Y - g(X^{\mathrm{T}}\beta_0)$. We use $\|\cdot\|$ to denote the Euclidean norm, and for two sequences $\{a_n\}$ and $\{b_n\}$, $a_n \asymp b_n$ means $a_n = O(b_n)$ and $b_n = O(a_n)$.

The following assumptions are needed in our analysis.

Assumption 3.1. There exists a m-variate random vector $Z_i = (z_{i1}, \ldots, z_{im})^{\mathrm{T}}$ for some $m \geq p$ so that $X_i = \Gamma Z_i$, where Γ is a $p \times m$ constant matrix such that $\Gamma\Gamma^{\mathrm{T}} = \Sigma_{\mathrm{X}}$ and $E(Z_i) = 0$, $\mathrm{var}(Z_i) = \mathbb{I}_m$, where $\mathbb{I}_m$ is the $m \times m$ identity matrix. Each z_{ij} has a finite 8th moment and $E(z_{ij}^4) = 3 + \Delta$ for a constant $\Delta > -3$, and for any integers $\ell_\nu \geq 0$ and distinct $j_1, \ldots, j_q$ with $\sum_{\nu=1}^{q} \ell_\nu \leq 8$,

$$E(z_{ij_1}^{\ell_1} z_{ij_2}^{\ell_2} \cdots z_{ij_q}^{\ell_q}) = E(z_{ij_1}^{\ell_1}) E(z_{ij_2}^{\ell_2}) \cdots E(z_{ij_q}^{\ell_q}).$$

Assumption 3.2. As $n \to \infty$, $p \to \infty$, $\mathrm{tr}(\Sigma_{\mathrm{X}}^2) \to \infty$ and $\mathrm{tr}(\Sigma_{\mathrm{X}}^4) = o\{\mathrm{tr}^2(\Sigma_{\mathrm{X}}^2)\}$.

Assumption 3.3. Let f_x be the probability density of X and $D(f_x)$ be its support. There exist positive constants K_1 and K_2 such that $E(\epsilon^2 | X = x) > K_1$ and $E(\epsilon^8 | X = x) < K_2$ for any $x \in D(f_x)$.

Assumption 3.4. g is once continuous differentiable, $V(\cdot) > 0$, and there exist positive constants c_1 and c_2 such that $c_1 \leq \psi^2(x, \beta_0, \phi) \leq c_2$ for any $x \in D(f_x)$.

Assumption 3.1 is used in Bai and Saranadasa (1996) and Zhong and Chen (2011) to facilitate the analysis in ultra high dimensional tests for the means and linear regression. The model contains the Gaussian and some other important multivariate distributions as special cases; see Chen et al. (2009). Assumption 3.2 is a weaker substitute to conditions which are explicit on the relative rates between p and n, for instance, $\log(p) \asymp n^{1/3}$, say. It is noted that when all the eigenvalues of Σ_{X} are bounded, $\mathrm{tr}(\Sigma_{\mathrm{X}}^4) = o\{\mathrm{tr}^2(\Sigma_{\mathrm{X}}^2)\}$ is true for any diverging p, and the condition allows diverging eigenvalues. Assumption 3.3 and 3.4 are standard in the analysis of generalized linear models, for instance, the assumption G in Fan and Song (2010). Assumption 3.4 is satisfied if Y is from the exponential family with canonical links.

To reduce the amount of notation, we assume the dispersion parameter ϕ can be ignored in the inference for β. We will consider ϕ in Section 5 when treating nuisance parameters. To facilitate the analysis, we define three matrices:

$$\Delta_{\beta,\beta_0} = E[\{g(X^{\mathrm{T}}\beta) - g(X^{\mathrm{T}}\beta_0)\}\psi(X, \beta_0)X],$$

$$\Sigma_\beta(\beta_0) = E[V\{g(X^{\mathrm{T}}\beta)\}\psi^2(X, \beta_0)XX^{\mathrm{T}}] \quad \text{and}$$

$$\Xi_{\beta,\beta_0} = E[\{g(X^{\mathrm{T}}\beta) - g(X^{\mathrm{T}}\beta_0)\}^2\psi^2(X, \beta_0)XX^{\mathrm{T}}].$$

The test statistic $\widehat{S}_n$ of Goeman et al. (2011) can be expressed as

$$\widehat{S}_n = 1 + U_n/A_n \quad \text{where}$$

$$U_n = \frac{1}{n}\sum_{i \neq j}^{n}\{(Y_i - \mu_{0i})(Y_j - \mu_{0j})\psi(X_i, \beta_0)\psi(X_j, \beta_0)X_i^{\mathrm{T}}X_j\} \quad \text{and}$$

$$A_n = \frac{1}{n}\sum_{i=1}^{n}\{(Y_i - \mu_{0i})^2\psi^2(X_i, \beta_0)X_i^{\mathrm{T}}X_i\}.$$

In additional to the remark made at end of the last section regarding the meaning of the statistic $\widehat{S}_n$, more insight can be made via the means of U_n and A_n. Derivations show that the means are, respectively,

$$\mu_{\mathrm{U}_n} = (n-1)\Delta_{\beta,\beta_0}^{\mathrm{T}}\Delta_{\beta,\beta_0} \quad \text{and} \quad \mu_{\mathrm{A}_n} = \mathrm{tr}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0}\}. \tag{3.1}$$

We note that for the generalized linear models, the difference between β and β_0 is measured by $g(X^{\mathrm{T}}\beta) - g(X^{\mathrm{T}}\beta_0)$, which is reflected by Δ_{β,β_0} and Ξ_{β,β_0} defined above. Hence, U_n measures the difference $g(X^{\mathrm{T}}\beta) - g(X^{\mathrm{T}}\beta_0)$, and A_n is a certain measure of the noise.

Lemma A.1 in the Appendix shows that the variances of U_n and A_n are respectively

$$\sigma_{\mathrm{A}_n}^2 = n^{-1}\left[E\{\epsilon_0^4\psi^4(X,\beta_0)(X^{\mathrm{T}}X)^2\} - E^2\{\epsilon_0^2\psi^2(X,\beta_0)(X^{\mathrm{T}}X)\}\right] \quad \text{and}$$

$$\sigma_{\mathrm{U}_n}^2 = 4(n-2)(1-n^{-1})\xi_1 + 2(1-n^{-1})\xi_2 \tag{3.2}$$

where $\xi_1 = \Delta_{\beta,\beta_0}^{\mathrm{T}}\{\Sigma_\beta(\beta_0)+\Xi_{\beta,\beta_0}\}\Delta_{\beta,\beta_0} - (\Delta_{\beta,\beta_0}^{\mathrm{T}}\Delta_{\beta,\beta_0})^2$ and $\xi_2 = \mathrm{tr}\{\Sigma_\beta(\beta_0)+\Xi_{\beta,\beta_0}\}^2 - (\Delta_{\beta,\beta_0}^{\mathrm{T}}\Delta_{\beta,\beta_0})^2$.

From the central limit theorem, even $p \to \infty$,

$$\sigma_{\mathrm{A}_n}^{-1}(A_n - \mu_{\mathrm{A}_n}) \xrightarrow{d} N(0,1) \quad \text{as} \quad n \to \infty.$$

By Taylor expansion,

$$\widehat{S}_n = 1 + \mu_{\mathrm{A}_n}^{-1}\mu_{\mathrm{U}_n} - \mu_{\mathrm{A}_n}^{-2}\mu_{\mathrm{U}_n}(A_n - \mu_{\mathrm{A}_n}) + \mu_{\mathrm{A}_n}^{-1}(U_n - \mu_{\mathrm{U}_n}) + \mu_{\mathrm{A}_n}^{-3}\mu_{\mathrm{U}_n}(A_n - \mu_{\mathrm{A}_n})^2 + \cdots. \tag{3.3}$$

To identify the leading order term of (3.3), we consider two families of alternative hypothesis which produce different leading order terms. One is the so-called "local" alternatives:

$$\mathscr{L}_\beta = \Bigg\{\beta_0 \in R^p \;\Bigg|\; \Delta_{\beta,\beta_0}^{\mathrm{T}}\Sigma_{\mathrm{X}}\Delta_{\beta,\beta_0} = o\{n^{-1}\mathrm{tr}(\Sigma_{\mathrm{X}}^2)\} \text{ and either } \{g(X^{\mathrm{T}}\beta) - g(X^{\mathrm{T}}\beta_0)\}^2 = O(1) \text{ a.s.}$$
$$\text{or } (\beta-\beta_0)^{\mathrm{T}}\Sigma_{\mathrm{X}}(\beta-\beta_0) = O(1) \quad \text{and} \quad |g'(t)| \le C_0 \text{ for any } t \in (-\infty,\infty)\Bigg\}; \tag{3.4}$$

for a positive constant C_0. The other is the so-called "fixed" alternatives:

$$\mathscr{L}_\beta^{\mathrm{F}} = \left\{\beta_0 \in R^p \;\middle|\; \Delta_{\beta,\beta_0}^{\mathrm{T}}\Xi_{\beta,\beta_0}\Delta_{\beta,\beta_0} = o\{n^{-1}\mathrm{tr}(\Xi_{\beta,\beta_0}^2)\} \text{ and } \mathrm{tr}(\Sigma_{\mathrm{X}}^2) = o\{\mathrm{tr}(\Xi_{\beta,\beta_0}^2)\}\right\}. \tag{3.5}$$

Clearly, the H_0 is embedded in the "local" alternatives $\mathscr{L}_\beta$. While $\mathscr{L}_\beta$ encompasses β where the difference $\|\Delta_{\beta,\beta_0}\|$ is relatively small, it also includes β_0 not necessarily close to the true β when either g is uniformly bounded as in the logistic and the probit models, or g' is bounded as in the linear regression. We use the term "local" simply because H_0 is part of $\mathscr{L}_\beta$. It is noticed that $\mathscr{L}_\beta^{\mathrm{F}}$ is applicable to models with unbounded g' function such as Poisson or Negative Binomial regression.

If $\beta_0 \in \mathscr{L}_\beta$, the proof of Theorem 1 shows that

$$\sigma_{\mathrm{A}_n}^2 = O(n^{-1}\mu_{\mathrm{A}_n}^2) \quad \text{and} \quad \sigma_{\mathrm{U}_n}^2 = 2\mathrm{tr}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0}\}^2\{1+o(1)\} \tag{3.6}$$

which imply that

$$\widehat{S}_n = 1 + \mu_{\mathrm{A}_n}^{-1}\mu_{\mathrm{U}_n} + \mu_{\mathrm{A}_n}^{-1}(U_n - \mu_{\mathrm{U}_n}) + o_p(\mu_{\mathrm{A}_n}^{-1}\sigma_{\mathrm{U}_n}). \tag{3.7}$$

The above analysis shows that under the "local" alternatives, the test statistic $\widehat{S}_n$ is dominated by a linear function of U_n. It can be shown that this is the same as the fixed dimensional case, the setting of Goeman et al (2011)'s proposal. This is due to the fact that the quadratic term and beyond in the Taylor expansion (3.3) can be controlled in the case of diverging p if either g or its derivative is bounded under $\mathscr{L}_\beta$.

Let $\sigma^2_{\widehat{S}_n} = 2\mathrm{tr}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0}\}^2\mathrm{tr}^{-2}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0}\}$, which is the leading order variance of $\widehat{S}_n$ under $\mathscr{L}_\beta$.

Theorem 1. *Suppose Assumptions 3.1-3.4 hold, then under the "local" alternatives $\mathscr{L}_\beta$,*

$$\sigma^{-1}_{\widehat{S}_n}(\widehat{S}_n - 1 - \mu_{\mathrm{A}_n}^{-1}\mu_{\mathrm{U}_n}) \xrightarrow{d} N(0,1) \quad \textit{as } n \to \infty \textit{ and } p \to \infty.$$

Under the null hypothesis, $\mu_{\mathrm{A}_n}^{-1}\mu_{\mathrm{U}_n} = 0$ and $\sigma^2_{\widehat{S}_n} = 2\mathrm{tr}\{\Sigma^2_{\beta_0}(\beta_0)\}\mathrm{tr}^{-2}\{\Sigma_{\beta_0}(\beta_0)\}$. To formulate a test procedure based on the asymptotic normality, we need to estimate $\sigma_{\widehat{S}_n}$ and hence $\mathrm{tr}\{\Sigma^2_{\beta_0}(\beta_0)\}$ and $\mathrm{tr}^2\{\Sigma_{\beta_0}(\beta_0)\}$. Let

$$\mathrm{tr}\{\widehat{\Sigma^2_{\beta_0}(\beta_0)}\} = \frac{1}{n(n-1)}\sum_{i\neq j}^{n}\big[\{Y_i - g(X_i^{\mathrm{T}}\beta_0)\}^2\{Y_j - g(X_j^{\mathrm{T}}\beta_0)\}^2\psi^2(X_i,\beta_0)\psi^2(X_j,\beta_0)(X_i^{\mathrm{T}}X_j)^2\big] \text{ and}$$

$$\mathrm{tr}^2\{\widehat{\Sigma_{\beta_0}(\beta_0)}\} = \frac{1}{n(n-1)}\sum_{i\neq j}^{n}\big[\{Y_i - g(X_i^{\mathrm{T}}\beta_0)\}^2\{Y_j - g(X_j^{\mathrm{T}}\beta_0)\}^2\psi^2(X_i,\beta_0)\psi^2(X_j,\beta_0)(X_i^{\mathrm{T}}X_i)(X_j^{\mathrm{T}}X_j)\big]. \tag{3.8}$$

Lemma A.3 in the Appendix shows that both are ratioly consistent such that under H_0

$$\frac{\mathrm{tr}\{\widehat{\Sigma^2_{\beta_0}(\beta_0)}\}}{\mathrm{tr}\{\Sigma^2_{\beta_0}(\beta_0)\}} \xrightarrow{p} 1 \quad \text{and} \quad \frac{\mathrm{tr}^2\{\widehat{\Sigma_{\beta_0}(\beta_0)}\}}{\mathrm{tr}^2\{\Sigma_{\beta_0}(\beta_0)\}} \xrightarrow{p} 1 \quad \text{as } n \to \infty.$$

Theorem 1 and the Slutsky Lemma lead to an asymptotic α-level test that rejects H_0 if

$$\widehat{S}_n > 1 + z_\alpha\big[2\mathrm{tr}\{\widehat{\Sigma^2_{\beta_0}(\beta_0)}\}/\mathrm{tr}^2\{\widehat{\Sigma_{\beta_0}(\beta_0)}\}\big]^{1/2}, \tag{3.9}$$

where z_α is the upper α-quantile of $N(0,1)$.

Goeman et al. (2011) approximated the null distribution of $\widehat{S}_n$ by a ratio of quadratic forms based on normally distributed variables, which involves a numerical inversion of the characteristic function. A R package "globaltest" is available at www.bioconductor.org to implement the algorithm. The critical value obtained via the procedure of Goeman et al. (2011) is asymptotically equivalent to the right hand side of (3.9) under H_0 in the case of $p \to \infty$, which is confirmed by our simulation study. We will use the explicit critical value in (3.9) in the following power analysis.

Define the power of the test in (3.9) under the "local" alternatives $\mathscr{L}_\beta$ as

$$\Omega(\beta,\beta_0) = \mathrm{P}\left(\widehat{S}_n > 1 + z_\alpha\big[2\mathrm{tr}\{\widehat{\Sigma^2_{\beta_0}(\beta_0)}\}/\mathrm{tr}^2\{\widehat{\Sigma_{\beta_0}(\beta_0)}\}\big]^{1/2} \mid \beta_0 \in \mathscr{L}_\beta\right).$$

Corollary 1. *Under Assumptions 3.1-3.4 and the "local" alternatives $\mathscr{L}_\beta$,*

$$\Omega(\beta,\beta_0)=\Phi\left(-z_\alpha+\frac{n\|\Delta_{\beta,\beta_0}\|^2}{\left[2\mathrm{tr}\{\Sigma_\beta(\beta_0)+\Xi_{\beta,\beta_0}\}^2\right]^{1/2}}\right)\{1+o(1)\}\quad \textit{as } n\to\infty \textit{ and } p\to\infty.$$

The corollary shows that the power of the test in (3.9) is determined by

$$\mathrm{SNR}(\beta,\beta_0)=\frac{n\|\Delta_{\beta,\beta_0}\|^2}{\left[2\mathrm{tr}\{\Sigma_\beta(\beta_0)+\Xi_{\beta,\beta_0}\}^2\right]^{1/2}}.$$

We note that $\|\Delta_{\beta,\beta_0}\|^2=\|E[\{g(X^{\mathrm{T}}\beta)-g(X^{\mathrm{T}}\beta_0)\}\psi(X,\beta_0)X]\|^2$ measures the difference between H_0 and H_1, and can be viewed as the signal of the test problem. At the same time, $\left[2\mathrm{tr}\{\Sigma_\beta(\beta_0)+\Xi_{\beta,\beta_0}\}^2\right]^{1/2}$ can be regarded as the noise due to its close connection to the standard deviation of $\widehat{S}_n$. Hence $\mathrm{SNR}(\beta,\beta_0)$ is the signal-to-noise ratio of the test.

Let $\lambda_1\le\cdots\le\lambda_p$ be the eigenvalues of Σ_{X} and λ_{m_0} be the smallest non-zero one for a $m_0\in\{1,\ldots,p\}$. Since $\Delta_{\beta,\beta_0}^{\mathrm{T}}\Sigma_{\mathrm{X}}\Delta_{\beta,\beta_0}\le\lambda_p\|\Delta_{\beta,\beta_0}\|^2$ and $\mathrm{tr}(\Sigma_{\mathrm{X}}^2)\ge\lambda_{m_0}^2(p-m_0)$, a sufficient condition that ensures the first component of $\mathscr{L}_\beta$ is

$$\|\Delta_{\beta,\beta_0}\|^2=o\{\lambda_p^{-1}\lambda_{m_0}^2n^{-1}(p-m_0)\}. \tag{3.10}$$

Now let $\tilde\lambda_1\le\tilde\lambda_2\le\cdots\le\tilde\lambda_p$ be the eigenvalues of $\Sigma_\beta(\beta_0)+\Xi_{\beta,\beta_0}$. Assumption 3.3 and $\mathscr{L}_\beta$ imply that each $\tilde\lambda_i$ is bounded below and above by constant multiple of λ_i. Using the same argument leading to (3.10), we can show that $\mathrm{SNR}(\beta,\beta_0)$ is bounded within

$$\left(n\|\Delta_{\beta,\beta_0}\|^2\{2\tilde\lambda_p^2(p-m_0)\}^{-1/2},\quad n\|\Delta_{\beta,\beta_0}\|^2\{2\tilde\lambda_{m_0}^2(p-m_0)\}^{-1/2}\right).$$

Thus, if $\|\Delta_{\beta,\beta_0}\|$ is a larger order than $n^{-1/2}\lambda_p^{1/2}(p-m_0)^{1/4}$, $\mathrm{SNR}(\beta,\beta_0)\to+\infty$ and hence the power converges to 1. If $\|\Delta_{\beta,\beta_0}\|$ is a smaller order (weaker) than $n^{-1/2}\lambda_{m_0}^{1/2}(p-m_0)^{1/4}$, the test does not have power beyond the significant level α. Non-trivial power $\Omega(\beta,\beta_0)$ is attained if $\|\Delta_{\beta,\beta_0}\|\asymp n^{-1/2}\lambda_p^{1/2}(p-m_0)^{1/4}$.

Let us evaluate the power of the test under the "fixed" alternatives $\mathscr{L}_\beta^{\mathrm{F}}$, which is denoted as

$$\Omega^{\mathrm{F}}(\beta,\beta_0)=\mathrm{P}\left(\widehat{S}_n>1+z_\alpha\left[2\mathrm{tr}\{\widehat{\Sigma_{\beta_0}^2(\beta_0)}\}/\mathrm{tr}^2\{\widehat{\Sigma_{\beta_0}(\beta_0)}\}\right]^{1/2}\;\middle|\;\beta_0\in\mathscr{L}_\beta^{\mathrm{F}}\right).$$

Unlike the "local" alternatives case, the leading order terms under the "fixed" alternatives involve an additional term $\mu_{\mathrm{A}_n}^{-2}\mu_{\mathrm{U}_n}(A_n-\mu_{\mathrm{A}_n})$, as shown in the proof of Theorem 2. This term is a smaller order term in the fixed dimensional case as considered in Goeman et al. (2011). It is also ignorable in the high dimensional case when either g or g' are bounded, as have been shown earlier. However, it may not be the case under the "fixed" alternatives. Having $\mu_{\mathrm{A}_n}^{-2}\mu_{\mathrm{U}_n}(A_n-\mu_{\mathrm{A}_n})$ does not generate more signal for the test, but can increase the variance and hence causes a reduction in the power.

To make this point explicit, we consider a specific case of the "fixed" alternatives where

$$\|\Delta_{\beta,\beta_0}\|^2\asymp n^{\delta-1}\mathrm{tr}^{1/2}(\Xi_{\beta,\beta_0}^2)\quad\text{and}\quad\frac{E[\{g(X^{\mathrm{T}}\beta)-g(X^{\mathrm{T}}\beta_0)\}^4\psi^4(X,\beta_0)(X^{\mathrm{T}}X)^2]}{E^2[\{g(X^{\mathrm{T}}\beta)-g(X^{\mathrm{T}}\beta_0)\}^2\psi^2(X,\beta_0)(X^{\mathrm{T}}X)]}\asymp n^{1-2\delta} \tag{3.11}$$

for a $\delta\in(0,1/2)$. We need one more assumption analogous to Assumption 3.2 in the following analysis.

Assumption 3.5. As $n \to \infty$, $p \to \infty$, $\mathrm{tr}(\Xi^2_{\beta,\beta_0}) \to \infty$ and $\mathrm{tr}(\Xi^4_{\beta,\beta_0}) = o\{\mathrm{tr}^2(\Xi^2_{\beta,\beta_0})\}$.

Theorem 2. *Under Assumptions 3.1-3.5, (3.11) and the "fixed" alternatives $\mathscr{L}^{\mathrm{F}}_\beta$,*

$$\Omega^{\mathrm{F}}(\beta,\beta_0) = \Phi\left(\frac{1}{(1+\tau^2)^{1/2}}\left[-z_\alpha + \frac{n\|\Delta_{\beta,\beta_0}\|^2}{\{2\mathrm{tr}(\Xi^2_{\beta,\beta_0})\}^{1/2}}\right]\right)\{1+o(1)\} \tag{3.12}$$

as $n \to \infty$, $p \to \infty$, where $\tau^2 = (\mu^2_{\mathrm{U}_n}\sigma^2_{\mathrm{A}_n})/(\mu^2_{\mathrm{A}_n}\sigma^2_{\mathrm{U}_n}) \in (0,\infty)$ is a constant.

The reason for obtaining the power expression in (3.12) is that under the conditions of Theorem 2, $\sigma^2_{\mathrm{A}_n} = O(n^{-2\delta}\mu^2_{\mathrm{A}_n})$, $\sigma^2_{\mathrm{U}_n} = 2\mathrm{tr}(\Xi^2_{\beta,\beta_0})\{1+o(1)\}$ and

$$\widehat{S}_n = 1 + \mu^{-1}_{\mathrm{A}_n}\mu_{\mathrm{U}_n} - \mu^{-2}_{\mathrm{A}_n}\mu_{\mathrm{U}_n}(A_n - \mu_{\mathrm{A}_n}) + \mu^{-1}_{\mathrm{A}_n}(U_n - \mu_{\mathrm{U}_n}) + o_p(\mu^{-1}_{\mathrm{A}_n}\sigma_{\mathrm{U}_n}). \tag{3.13}$$

Note that, both $\mu^{-2}_{\mathrm{A}_n}\mu_{\mathrm{U}_n}(A_n - \mu_{\mathrm{A}_n})$ and $\mu^{-1}_{\mathrm{A}_n}(U_n - \mu_{\mathrm{U}_n})$ are the joint leading order terms of $\widehat{S}_n$. The role of condition (3.11) is to make the quadratic terms and beyond in the Taylor expansion (3.3) of $\widehat{S}_n$ are of smaller orders of the two linear terms in (3.13). A consequence of the $A_n - \mu_{A_n}$ term in the leading order term leads to τ^2 appeared in the power function, which causes a power reduction as reflected by the first fraction inside Φ in (3.12).

If the second part of (3.11) is more relaxed so that it is of a larger order of $n^{1-2\delta}$ but a smaller order of $n^{1-\delta}$, the power expression (3.12) still holds but with $\tau^2 \to \infty$. This means a dramatic deterioration in the power. If the order of the second term in (3.11) is higher than $n^{1-\delta}$, the quadratic terms and beyond in the expansion (3.3) will be of larger orders than the linear terms in (3.13), making the power analysis much harder to accomplish and the power performance unpredictable since the main signal bearing term $U_n - \mu_{\mathrm{U}_n}$ is no longer important.

4 A new proposal

An important insight acquired in the previous section is that the A_n term in the statistic

$$\widehat{S}_n = 1 + U_n/A_n$$

does not contribute to the signal of the test but can increase the variance (noise) and hence can adversely affect the power. Although A_n has a negligible effect on the power under the "local" alternatives $\mathscr{L}_\beta$, its role on the power becomes more pronounced under the "fixed" alternatives $\mathscr{L}^{\mathrm{F}}_\beta$. Dividing by A_n is a standard formulation that dates back to the Fisher's F-test for regression coefficients. However, under the high dimensionality, doing so may not be necessary since its contribution to the variance (noise) can be significant as shown in Theorem 2.

The above consideration leads us to propose a statistic by excluding A_n from $\widehat{S}_n$. Specifically, we consider using

$$U_n = \frac{1}{n}\sum_{i\neq j}^{n}\{(Y_i - \mu_{0i})(Y_j - \mu_{0j})\psi(X_i,\beta_0)\psi(X_j,\beta_0)X_i^{\mathrm{T}}X_j\}$$

as the test statistic. Comparing with the involved expansion (3.3) of $\widehat{S}_n$, U_n has a much simpler form. Despite being simpler, it captures the signal of the test since $E(U_n) = (n-1)\|\Delta_{\beta,\beta_0}\|^2$ as shown in (3.1). We will demonstrate in this section that a test based on U_n can achieve better power for diverging p than Goeman et al. (2011)'s test under $\mathscr{L}_\beta^{\mathrm{F}}$ while maintaining the same asymptotic power under $\mathscr{L}_\beta$.

We still consider testing the global hypothesis $H_0 : \beta = \beta_0$ in this section. A test for the presence of the nuisance parameters will be unveiled in the next section. Recall from (3.6) that under $\mathscr{L}_\beta$, $\sigma^2_{\mathrm{U}_n} = 2\mathrm{tr}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0}\}^2\{1 + o(1)\}$.

Theorem 3. *Under Assumptions 3.1-3.4 and the "local" alternatives $\mathscr{L}_\beta$,*

$$\frac{U_n - n\|\Delta_{\beta,\beta_0}\|^2}{\left[2\mathrm{tr}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0}\}^2\right]^{1/2}} \xrightarrow{d} N(0,1) \quad \textit{as } n \to \infty \textit{ and } p \to \infty.$$

Theorem 3 implies that under the null hypothesis,

$$\frac{U_n}{\left[2\mathrm{tr}\{\Sigma^2_{\beta_0}(\beta_0)\}\right]^{1/2}} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty \text{ and } p \to \infty.$$

Using $\mathrm{tr}\{\widehat{\Sigma^2_{\beta_0}(\beta_0)}\}$ given in (3.8) to estimate $\mathrm{tr}\{\Sigma^2_{\beta_0}(\beta_0)\}$, the proposed α-level test rejects H_0 if

$$U_n > z_\alpha\left[2\mathrm{tr}\{\widehat{\Sigma^2_{\beta_0}(\beta_0)}\}\right]^{1/2}. \tag{4.1}$$

Let $\widetilde{\Omega}(\beta,\beta_0) = \mathrm{P}\left(U_n > z_\alpha\left[2\mathrm{tr}\{\widehat{\Sigma^2_{\beta_0}(\beta_0)}\}\right]^{1/2} \mid \beta_0 \in \mathscr{L}_\beta\right)$ be the power of the above test under the "local" alternatives $\mathscr{L}_\beta$.

Corollary 2. *Under Assumptions 3.1-3.4 and the "local" alternatives $\mathscr{L}_\beta$,*

$$\widetilde{\Omega}(\beta,\beta_0) = \Phi\left(-z_\alpha + \frac{n\|\Delta_{\beta,\beta_0}\|^2}{\left[2\mathrm{tr}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0}\}^2\right]^{1/2}}\right)\{1 + o(1)\} \quad \textit{as } n \to \infty \textit{ and } p \to \infty.$$

We note here that the power of the proposed test is asymptotically equivalent to $\Omega(\beta,\beta_0)$ of Goeman et al. (2011) given in Corollary 1. This is expected since in the case of the "local" alternatives $\mathscr{L}_\beta$,

$$1 + \mu^{-1}_{\mathrm{A}_n}\mu_{\mathrm{U}_n} + \mu^{-1}_{\mathrm{A}_n}(U_n - \mu_{\mathrm{U}_n}) \tag{4.2}$$

is the leading order term of $\widehat{S}_n$. Hence, the two tests are asymptotically equivalent.

From the proof of Theorem 4, the asymptotic variance of U_n under the "fixed" alternatives $\mathscr{L}_\beta^{\mathrm{F}}$ is

$$\sigma^2_{\mathrm{U}_n} = 2\mathrm{tr}(\Xi^2_{\beta,\beta_0})\{1 + o(1)\}.$$

Let $\widetilde{\Omega}^{\mathrm{F}}(\beta,\beta_0) = \mathrm{P}\left(U_n > z_\alpha\left[2\mathrm{tr}\{\widehat{\Sigma^2_{\beta_0}(\beta_0)}\}\right]^{1/2} \mid \beta_0 \in \mathscr{L}_\beta^{\mathrm{F}}\right)$ be the power under $\mathscr{L}_\beta^{\mathrm{F}}$.

Theorem 4. *Under Assumptions 3.1-3.5 and the "fixed" alternatives $\mathscr{L}_\beta^{\mathrm{F}}$,*

$$\widetilde{\Omega}^{\mathrm{F}}(\beta,\beta_0) = \Phi\left(-z_\alpha + \frac{n\|\Delta_{\beta,\beta_0}\|^2}{\{2\mathrm{tr}(\Xi^2_{\beta,\beta_0})\}^{1/2}}\right)\{1 + o(1)\} \quad \textit{as } n \to \infty \textit{ and } p \to \infty.$$

The conditions in Theorem 4 are much simpler than those in Theorem 2, as condition (3.11) is not needed. To compare the two power functions under the "fixed" alternatives while assuming the conditions of Theorem 2, (3.11) implies that

$$\frac{n\|\Delta_{\beta,\beta_0}\|^2}{\{2\mathrm{tr}(\Xi^2_{\beta,\beta_0})\}^{1/2}} \asymp n^{\delta} \to \infty.$$

A power gain is evident as $\widetilde{\Omega}^{\mathrm{F}}(\beta,\beta_0) > \Omega^{\mathrm{F}}(\beta,\beta_0)$ asymptotically, since the power function (3.12) has an extra τ^2 in the denominator.

5 Test with nuisance parameter

We consider testing for parts of the regression coefficient vector β. This is motivated by practical needs to consider the significance for a subset of covariates $X^{(2)}$, in the presence of other covariates $X^{(1)}$. For instance, one may have both gene expression levels and demographic variables collected in a study on the cause of a disease. The researcher may be interested only in the genetic effect. In this case, the demographic coefficients together with the dispersion parameter may be viewed as nuisance parameters.

Without loss of generality, we partition $\beta = (\beta^{(1)\mathrm{T}}, \beta^{(2)\mathrm{T}})^{\mathrm{T}}$ and denote the nuisance parameters $\theta = (\beta^{(1)\mathrm{T}}, \phi)^{\mathrm{T}}$, where ϕ is the nuisance dispersion parameter. Suppose the dimension of θ is p_1 and that of $\beta^{(2)}$ is p_2. It is of interest to test

$$H_{01} : \beta^{(2)} = \beta_0^{(2)} \quad \text{versus} \quad H_{11} : \beta^{(2)} \neq \beta_0^{(2)}.$$

A test statistic along the line of the global test statistic U_n will be proposed. To this end, the nuisance parameters $\beta^{(1)}$ and ϕ have to be estimated first under H_{01}. The quasi-likelihood score of $\beta^{(1)}$ is

$$\ell_1(\beta^{(1)}, \beta^{(2)}, \phi) = \mathbb{X}^{(1)\mathrm{T}}\{(\mathbb{Y} - \mu) \circ \Psi\},$$

where $\mathbb{X}^{(1)}$ is similarly defined as $\mathbb{X}^{(2)}$ in Section 2, $\Psi = \{\psi(X_1, \beta, \phi), \ldots, \psi(X_n, \beta, \phi)\}^{\mathrm{T}}$ and $\mu = \{\mu_1(\beta), \ldots, \mu_n(\beta)\}^{\mathrm{T}}$. The maximum quasi-likelihood estimator of $\beta^{(1)}$ under H_{01} solves

$$\ell_1(\beta^{(1)}, \beta_0^{(2)}, \hat{\phi}_0) = 0,$$

which is denoted as $\hat{\beta}_0^{(1)}$, by plugging-in $\hat{\phi}_0$, either a maximum likelihood estimator or a moment estimator of ϕ as elaborated in McCullagh and Nelder (1989) and Chen and Cui (2003). Let $\hat{\beta}_0 = (\hat{\beta}_0^{(1)\mathrm{T}}, \beta_0^{(2)\mathrm{T}})^{\mathrm{T}}$, $\hat{\theta}_0 = (\hat{\beta}_0^{(1)\mathrm{T}}, \hat{\phi}_0)^{\mathrm{T}}$ and $\hat{\mu}_{0i} = \mu_i(\hat{\beta}_0)$.

We consider a statistic,

$$\widetilde{U}_n = \frac{1}{n}\sum_{i\neq j}^{n}\{(Y_i - \hat{\mu}_{0i})(Y_j - \hat{\mu}_{0j})\psi(X_i, \hat{\beta}_0, \hat{\phi}_0)\psi(X_j, \hat{\beta}_0, \hat{\phi}_0)X_i^{(2)\mathrm{T}}X_j^{(2)}\}. \tag{5.1}$$

Let $\Sigma_{\mathrm{X}^{(\mathrm{i})}} = E(X^{(\mathrm{i})}X^{(\mathrm{i})\mathrm{T}})$ for $i = 1$ and 2. The following assumptions are needed.

Assumption 5.1. As $n \to \infty$, $p_2 \to \infty$, $\mathrm{tr}(\Sigma^2_{\mathrm{X}^{(2)}}) \to \infty$ and $\mathrm{tr}(\Sigma^4_{\mathrm{X}^{(2)}}) = O\{n^{-1}\mathrm{tr}^2(\Sigma^2_{\mathrm{X}^{(2)}})\}$.

Assumption 5.2. As $n \to \infty$, $p_1 n^{-1/4} \to 0$ and there exists a $\theta^* = (\beta^{*(1)\mathrm{T}}, \phi^*)^{\mathrm{T}} \in R^{p_1}$ such that $\|\hat{\theta}_0 - \theta^*\| = O_p(p_1 n^{-1/2})$, and in particular under H_{01}, $\theta^* = \theta$, where $\theta = (\beta^{(1)\mathrm{T}}, \phi)^{\mathrm{T}}$ is the true value of nuisance parameter.

Assumption 5.3. There exists a positive constant λ_0 such that $0 < \lambda_0 \leq \lambda_{\min}(\Sigma_{\mathrm{X}^{(1)}}) \leq \lambda_{\max}(\Sigma_{\mathrm{X}^{(1)}}) \leq \lambda_0^{-1} < \infty$, where $\lambda_{\min}(\Sigma_{\mathrm{X}^{(1)}})$ and $\lambda_{\max}(\Sigma_{\mathrm{X}^{(1)}})$ represent the smallest and largest eigenvalues of the matrix $\Sigma_{\mathrm{X}^{(1)}}$ respectively.

Assumption 5.4. There exist positive constants c_1 and c_2 such that for $\beta_0^* = (\beta^{*(1)\mathrm{T}}, \beta_0^{(2)\mathrm{T}})^{\mathrm{T}}$ where $\beta^{*(1)}$ is defined in Assumption 5.2, $c_1 \leq \psi^2(x, \beta_0^*, \phi^*) \leq c_2$ and $[\partial\psi\{g(t)\}/\partial g(t)]^2 \mid_{t=x^{\mathrm{T}}\beta_0^*} \leq c_2$ for any $x \in D(f_x)$ and a neighborhood of $x^{\mathrm{T}}\beta_0^*$.

These assumptions are variations of Assumptions 3.2-3.4 in Section 2. Specifically, Assumption 5.1 is equivalent to Assumption 3.2 in the presence of the nuisance parameter. The requirement of the growing rate of p_1 being slower than $n^{1/4}$ is to allow accurate estimation of the nuisance parameter under the high dimensionality. Assumption 5.2 maintains that the initial estimator $\hat{\theta}_0$ is consistent to a θ^* which may deviate from the true parameter θ, when the discrepancy between $\beta_0^{(2)}$ and $\beta^{(2)}$ is large. The θ^* is the one that minimizes the Kullback-Leibler divergence between the misspecified model under H_{01} and the model under H_{11}; see van de Vaart (2000) for details. Assumption 5.3 is easier to be satisfied due to $\Sigma_{\mathrm{X}^{(1)}}$'s dimension is much more manageable than the case considered in the previous section. Assumption 5.4 is an updated version of Assumption 3.4 to suit the case of nuisance parameters.

To analyze the power, we introduce two matrices

$$\Delta^{(2)}_{\beta,\beta_0^*} = E[\{g(X^{\mathrm{T}}\beta) - g(X^{\mathrm{T}}\beta_0^*)\}\psi(X, \beta_0^*, \phi^*)X^{(2)}] \quad \text{and}$$
$$\Sigma^{(2)}_{\beta}(\beta_0^*) = E[V\{g(X^{\mathrm{T}}\beta)\}\psi^2(X, \beta_0^*, \phi^*)X^{(2)}X^{(2)\mathrm{T}}],$$

which are counterparts of Δ_{β,β_0} and $\Sigma_\beta(\beta_0)$ used in the study of the global test. There is no need to define a counterpart of Ξ_{β,β_0} since the second part of the "local" alternatives $\mathscr{L}_{\beta^{(2)}}$ defined below makes it unnecessary.

The involvement of the estimated nuisance parameter $\hat{\theta}_0$ does complicate the power analysis. To expedite the study, our analysis is confined under the following family of the "local" alternatives

$$\mathscr{L}_{\beta^{(2)}} = \left\{ \beta_0^{(2)} \in R^{p_2} \;\middle|\; \Delta^{(2)\mathrm{T}}_{\beta,\beta_0^*}\Sigma_{\mathrm{X}^{(2)}}\Delta^{(2)}_{\beta,\beta_0^*} = o\{n^{-1}\mathrm{tr}(\Sigma^2_{\mathrm{X}^{(2)}})\} \text{ and } E\{g(X^{\mathrm{T}}\beta) - g(X^{\mathrm{T}}\beta_0^*)\}^4 = o(n^{-3/2}) \right\}.$$

We note here that the second component of $\mathscr{L}_{\beta^{(2)}}$ is stronger than that in $\mathscr{L}_\beta$ in (3.4), which simplifies the analysis in the presence of the nuisance parameter.

Theorem 5. *Under Assumptions 3.1, 3.3, 5.1-5.4, and the "local" alternatives $\mathscr{L}_{\beta^{(2)}}$,*

$$\frac{\widetilde{U}_n - n\|\Delta_{\beta,\beta_0^*}\|^2}{\left[2\mathrm{tr}\{\Sigma^{(2)}_\beta(\beta_0^*)\}^2\right]^{1/2}} \xrightarrow{d} N(0,1) \quad \textit{as } n \to \infty \textit{ and } p_2 \to \infty.$$

To formulate a test procedure, we use

$$\widehat{R}_n = \frac{1}{n(n-1)} \sum_{i \neq j}^{n} (Y_i - \hat{\mu}_{0i})^2 (Y_j - \hat{\mu}_{0j})^2 \psi^2(X_i, \hat{\beta}_0, \hat{\phi}_0) \psi^2(X_j, \hat{\beta}_0, \hat{\phi}_0) (X_i^{(2)\mathrm{T}} X_j^{(2)})^2$$

to estimate $\mathrm{tr}\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2$ under H_{01}. Lemma A.6 in the Appendix shows that the estimator is ratioly consistent under H_{01}, that is

$$\frac{\widehat{R}_n}{\mathrm{tr}\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2} \xrightarrow{p} 1 \quad \text{as } n \to \infty.$$

Hence, an asymptotic α-level test rejects H_{01} if $\widetilde{U}_n > z_\alpha (2\widehat{R}_n)^{1/2}$ and the proofs of Theorem 5 and Lemma A.6 show that the test procedure is invariant to the scale transformation of Y. Define the power of the test under the "local" alternatives $\mathscr{L}_{\beta^{(2)}}$

$$\widetilde{\Omega}^{(2)}(\beta, \beta_0^*) = \mathrm{P}\left(\widetilde{U}_n > z_\alpha (2\widehat{R}_n)^{1/2} \mid \beta_0^{(2)} \in \mathscr{L}_{\beta^{(2)}}\right).$$

Corollary 3. *Under Assumptions 3.1, 3.3, 5.1-5.4 and the "local" alternatives* $\mathscr{L}_{\beta^{(2)}}$,

$$\widetilde{\Omega}^{(2)}(\beta, \beta_0^*) = \Phi\left(-z_\alpha + \frac{n\|\Delta_{\beta,\beta_0^*}^{(2)}\|^2}{\left[2\mathrm{tr}\{\Sigma_\beta^{(2)}(\beta^*)\}^2\right]^{1/2}}\right)\{1 + o(1)\} \quad \textit{as } n \to \infty \textit{ and } p_2 \to \infty.$$

The power $\widetilde{\Omega}^{(2)}(\beta, \beta_0^*)$ has a similar form as $\widetilde{\Omega}(\beta, \beta_0)$ in Corollary 2. This is expected due to the close connection between the two tests and their test statistics respectively. We note that the denominator inside Φ only involves $\Sigma_\beta^{(2)}(\beta^*)$ due to the second part of $\mathscr{L}_{\beta^{(2)}}$.

We did not study the power under the "fixed" alternatives similar to the one in Section 3, as we would expect the power performance would be largely similar to the one depicted in Section 4. We also did not study the power property of the Goeman et al. (2011)'s test with nuisance parameter as the analysis would be quite involved due to the division of A_n term and the estimated nuisance parameter. However, we would expect similar power properties revealed in the previous section would prevail, namely the power of Goeman et al. (2011)'s test would be hampered when g' is unbounded. This is indeed confirmed by the simulation results reported in the next section.

6 Simulation studies

We report in this section results from simulation studies which were designed to evaluate the performances of the proposed high dimensional test procedures for the generalized linear models. Both the global test and the test in the presence of nuisance parameter were considered for both the proposed and Goeman et al. (2011)'s tests. The R package "globaltest" is used to carry out a version of the Goeman et al. (2011)'s test for diverging p. We also carried out the test given in

(3.9) based on the asymptotic normality. Both had close size, confirming the fact that the two forms of the critical value lead to equivalent tests.

Throughout this section, the covariates $X_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$ were generated according to a moving average model

$$X_{ij} = \rho_1 Z_{ij} + \rho_2 Z_{i(j+1)} + \cdots + \rho_T Z_{i(j+T-1)}, \quad j = 1, \ldots, p;$$

for some $T < p$, where $Z_i = (Z_{i1}, \ldots, Z_{i(p+T-1)})^{\mathrm{T}}$ were from a $(p+T-1)$ dimensional standard normal distribution $N(0, \mathbb{I}_{p+T-1})$. The coefficients $\{\rho_l\}_{l=1}^T$ were generated independently from the $U(0,1)$ distribution, and were treated as fixed once generated. Here, T was used to prescribe different levels of dependence among the components of the high dimensional vector X_i. We had experimented $T = 5, 10$ and 20, and only reported the results for $T = 5$ since those for $T = 10$ and 20 were largely similar.

Four generalized linear models were considered in the simulation study: the logistic, linear, Poisson and Negative Binomial regression models. In the logistic regression model, the conditional mean of the response Y was given by

$$E(Y_i|X_i) = g(X_i^{\mathrm{T}}\beta) = \frac{\exp(X_i^{\mathrm{T}}\beta)}{1 + \exp(X_i^{\mathrm{T}}\beta)},$$

and conditioning on X_i, $Y_i \sim \text{Bernoulli}\{1, g(X_i^{\mathrm{T}}\beta)\}$. In the linear regression,

$$E(Y_i|X_i) = g(X_i^{\mathrm{T}}\beta) = X_i^{\mathrm{T}}\beta,$$

and conditioning on X_i, $Y_i \sim N(X_i^{\mathrm{T}}\beta, 1)$. We note here that the test is invariant with respect to the nuisance dispersion parameter σ^2. Hence, setting $\sigma = 1$ was not crucial for the test performance. In the Poisson regression,

$$E(Y_i|X_i) = g(X_i^{\mathrm{T}}\beta) = \exp(X_i^{\mathrm{T}}\beta),$$

and conditioning on X_i, $Y_i \sim \text{Poisson}\{g(X_i^{\mathrm{T}}\beta)\}$. The setup for the Negative Binomial model was

$$Y|\lambda \sim \text{Poisson}(\lambda) \quad \text{and} \quad \lambda \sim \text{Gamma}\{\exp(X^{\mathrm{T}}\beta), 1\}.$$

The conditional distribution of Y given X is the negative binomial distribution $NB\{\exp(X^{\mathrm{T}}\beta), 1/2\}$, which prescribes an over-dispersion to the Poisson model, and makes it a popular alternative to the Poisson regression in practice.

To create regimes of high dimensionality, we chose a relationship $p = \exp(n^{0.4})$ and specifically considered $(n, p) = (80, 320)$ and $(200, 4127)$ in the simulations. Seven nominal type I errors ranging from 0.05 to 0.2 were considered, and the corresponding empirical sizes and powers were evaluated from 2000 replications.

We first considered testing the global hypothesis

$$H_0 : \beta = 0_{p\times 1} \quad \text{versus} \quad H_1 : \beta \neq 0_{p\times 1}. \tag{6.1}$$

In designing the alternative hypothesis, for the linear model we made $\|\beta\|^2 = 0.2$ and chose the first five coefficients in β to be non-zero of equal magnitude and the rest of the coefficients to be zero. For the other three models, $\|\beta\|^2 = 2$. Hence, the non-zero coefficients were quite sparse. In order to have a reasonable range for the response variable, as in Goeman et al. (2011), we restricted $E(Y_i|X_i)$ between $\exp(-4)/\{1+\exp(-4)\} = 0.02$ and $\exp(4)/\{1+\exp(4)\} = 0.98$ for the logistic model, between -1000 and 1000 for the linear model, and between $\exp(0) = 1$ and $\exp(4) = 55$ for the Poisson and Negative Binomial models respectively.

The empirical power profiles (curves of empirical power versus empirical size) of the global tests were plotted in Figure 1. It is observed that the proposed global test and Goeman et al. (2011)'s test had largely similar power profiles for the logistic and linear models as displayed by Panels (a) - (d) of the figure. This is consistent with our findings in Corollaries 1 and 2, which indicate that both tests have the same asymptotic powers under the "local" alternatives $\mathscr{L}_\beta$.

Panels (a) and (b) of Figure 1 displayed that the proposed test had a slightly higher power than Goeman et al. (2011)'s test in the case of the logistic model. This can be understood as the impact of A_n term on the variance of $\widehat{S}_n$ despite its being the second order under the "local" alternatives.

Panels (c)-(d) of Figure 1 displayed that both Goeman et al. (2011)'s test and the proposed test had almost identical size and power performance in the linear regression model. This confirms the provision of our theory regarding bounded g' as shown in Corollary 1.

Panels (e)-(h) showed a much larger discrepancy in the power profiles between the two tests for the Poisson and Negative Binomial models with the proposed test being significantly more powerful. It is noted that both models have unbounded g', which imply that the testing was operated in the regime of the "fixed" alternatives $\mathscr{L}_\beta^{\mathrm{F}}$. The simulated power profiles confirmed the findings in Theorem 2 in that an unbounded g' function can adversely impact the power of Goeman et al. (2011)'s test, whereas the proposed test withstands such situations due to its test statistic formulation.

We then conducted simulation for testing

$$H_0 : \beta^{(2)} = 0_{p_2\times 1} \quad \text{versus} \quad H_1 : \beta^{(2)} \neq 0_{p_2\times 1} \tag{6.2}$$

in the presence of nuisance parameter $\beta^{(1)}$ for the same four generalized linear models considered above. The nuisance parameter $\beta^{(1)}$ was $p_1 = 10$ dimensional, generated randomly from $U(0,1)$ as in the design of the global hypothesis. We still chose $(n, p_2) = (80, 320)$ and $(200, 4127)$ by assigning $p_2 = \exp(n^{0.4})$. To evaluate the power of the test, the first five elements of $\beta^{(2)}$ were set to be non-zero of equal magnitude with $\|\beta^{(2)}\|^2 = 0.5$ for the linear model and $\|\beta^{(2)}\|^2 = 2$ for the other three generalized linear models, while the rest of $\beta^{(2)}$ were zeros.

The power profiles of the proposed and Goeman et al. (2011)'s tests were displayed in Figure 2. It is observed from Panel (a) of Figure 2 that, for the logistic model with $n = 80$ and $p_2 = 320$, the test of Goeman et al. (2011) had very severe size distortion, which may be due to the estimation of the nuisance parameter. We observed that when the sample size was increased to $n = 200$, Panel (b) of Figure 2 shows the size distortion is no longer that severe as compared with the case of $n = 80$. However, our test statistic is more robust. Indeed, the size distortion

presence for the test of Goeman et al. (2011) was largely absent for the proposed test. Figure 2 shows that the proposed test had quite reasonable power with good control of the type I error. For the Poisson and Negative Binomial models (Panels (e)-(h)), we observed that the proposed test had much more advantageous power profiles than those of Goeman et al. (2011)'s test. The latter was similar to the global tests demonstrated in Figure 1.

7 Case study

We analyze a dataset that contains microarray readings for 128 persons who suffer the acute lymphoblastic leukemia (ALL). The dataset also has information on patients' age, gender and response to multidrug resistance. Among the 128 individuals, 75 of them were patients of the B-cell type leukemia which were classified further to two types: the BCR/ABL fusion (35 *patients*) and cytogenetically normal NEG (40 *patients*). The dataset has been analyzed by Chiaretti et al. (2004), Dudoit et al. (2008), Chen and Qin (2010) and Li and Chen (2012) and others motivated from different aspects of the inference.

Biological studies have shown that each gene tends to work with other genes to perform certain biological missions. Biologists have defined gene-sets under the Gene Ontology system which provides structured vocabularies producing names of Gene Ontology terms. The gene-sets under the Gene Ontology system have been classified to three broad functional categories: Biological Processes, Cellular Components and Molecular Functions. There have been a set of research works focusing on identifying differentially expressed sets of genes in the analysis of gene expression data; see Efron and Tibshirani (2007), Rahmatallah et al. (2012). After preliminary gene-filtering with the algorithm proposed in Gentleman et al. (2005), there were 2250 unique Gene Ontology terms in Biological Processes, 328 in Cellular Component and 402 in Molecular Function categories respectively, which involved 3265 genes in total.

Our aim here is to identify gene-sets within each functional category, which are significant in determining the two types of B-cell ALL: BCR/ABL fusion or cytogenetically normal NEG. We formulate it as a binary regression with the response Y_i being 1 if the ith patient had the BCR/ABL type ALL and 0 if had the NEG type. The covariate of the ith patient corresponding to a gene-set, label by g in the subscript, is $X_{ig} = (X_{ig}^{(1)\mathrm{T}}, X_{ig}^{(2)\mathrm{T}})^{\mathrm{T}}$, where $X_{ig}^{(1)}$ contains the gender, age and the patient's response to multidrug resistance (1 if negative and 0 positive), and $X_{ig}^{(2)}$ is the vector of gene expression levels of the gth Gene Ontology term.

We considered the logistic and probit models for the gene-set data due to the binary nature of the response variable. The two models are, respectively,

$$E(Y_i|X_{ig}^{(1)}, X_{ig}^{(2)}) = \frac{\exp(X_{ig}^{(1)\mathrm{T}}\beta_g^{(1)} + X_{ig}^{(2)\mathrm{T}}\beta_g^{(2)})}{1 + \exp(X_{ig}^{(1)\mathrm{T}}\beta_g^{(1)} + X_{ig}^{(2)\mathrm{T}}\beta_g^{(2)})} \quad \text{and}$$

$$E(Y_i|X_{ig}^{(1)}, X_{ig}^{(2)}) = \Phi(X_{ig}^{(1)\mathrm{T}}\beta_g^{(1)} + X_{ig}^{(2)\mathrm{T}}\beta_g^{(2)}).$$

For the leukemia data, it is of fundamental interest in discovering significant Gene Ontology terms while considering the effects of the three covariates in $X^{(1)}$, namely by treating $\beta_g^{(1)}$ as the

nuisance parameter and testing the following hypothesis:

$$H_0: \ \beta_g^{(2)} = 0 \quad \text{versus} \quad H_1: \ \beta_g^{(2)} \neq 0.$$

By controlling the false discovery rate (Benjamini and Hochberg, 1995) at 0.01, 1084 gene-sets in Biological Processes, 154 in Cellular Components and 153 in Molecular Function were found significant under the logistic model, and 981 in Biological Processes, 140 in Cellular Components and 132 in Molecular Function were significant under the probit model. Table 1 reports the two by two rejection/non-rejection classification between the tests under the two models. It shows that the testing results were largely agreeable between the two models. This was especially the case for the gene-set categories of Biological Processes and Cellular Components, with more than 90% of the gene-sets rejected under the logistic model being also rejected under the probit model, and the non-rejected gene-sets matched perfectly. The discrepancy in the test conclusions got larger for gene-sets in the Molecular Function category. But still, the percentages of agreement between the two models exceeded 72% in the rejection and 92% in the non-rejection. These showed again the testings under the two models attained similar results.

We also carried out the global test for the significance of the entire regression coefficient vector β_g by performing test on

$$H_0: \ \beta_g = 0 \quad \text{versus} \quad H_1: \ \beta_g \neq 0$$

where $\beta_g = (\beta_g^{(1)\mathrm{T}}, \beta_g^{(2)\mathrm{T}})^{\mathrm{T}}$ with the first three coefficients corresponding to the three non-genetic covariates: the gender, age and multidrug resistance. We note that the value of the standardized global test statistics under the logistic and the probit models were identical. This is because under the H_0, $g(X_i^{\mathrm{T}}\beta_g) = g(0) = 0.5$ and $\psi(X_i, 0)$ are constant for both models, which means that $\psi(X_i, 0)$ are canceled out in the standardized test statistics. Hence, the test procedures were identical for testing the global hypothesis regarding each gene-set under both the logistic and probit models.

Figure 3 displays the histograms of p-values and the standardized global test statistics L_n. It is observed that the bulk of the test statistics (right panels) took extremely large values in the scale of the standard normal distribution, implying that most of the p-values would be very small and the significance of many sets of genes. The latter was confirmed by the left panels of Figure 3. The histograms of the standardized test statistics and the p-values of the test for the gene-sets only while treating the first three coefficients as the nuisance parameter are shown in Figures 4 and 5. Comparing Figure 3 with Figures 4 and 5, it is found that the body of the histograms were much less extreme in Figures 4 and 5 than those in Figure 3. This indicates that much of the significance in the global tests were due to the significance of the three nuisance covariates rather than the gene-sets. It also demonstrates that considering the three nuisance parameters was necessary in filtering out the influence of the gene-sets between the two types of B-cell acute lymphoblastic leukemia.

8 Discussion

As the generalized linear models are widely used tools in analyzing genetic data, the proposed tests, being more adaptive to the high dimensionality, are useful additions to the existing test procedures for the significance of regression coefficients. As shown in the case study, testing for the significance of gene-sets requires high dimensional multivariate test procedures which can produce p-values under both high dimensionality and multiplicity (as genes in gene-sets can overlap). The proposed tests and the tests of Goeman et al. (2011) are such tests which can be used for the gene-sets testing in conjunction with the FDR procedure when testing a large number of hypotheses simultaneously.

The test of Goeman et al. (2011) was proposed for fixed dimensional p. We have found it is quite resilient to the diverging p in both theoretical and empirical analysis, as long as either the inverse of the link function or its derivative is bounded. The latter encompasses the logistic and linear regression models. The proposed tests are designed to improve the performance of Goeman et al. (2011)'s test for diverging p. This is especially the case when the first order derivative of the inverse of the link function is unbounded, when the high dimensionality can insert adverse influence on the test of Goeman et al. (2011). The proposed test statistics due to their simpler formulations can avoid some of the high dimensional effects, and hence lead to better performances in terms of more accurate size approximation and more power. Of course, when p is of fixed dimension, the test of Goeman et al. (2011) may be used and the proposed test may not be valid.

There have been works on the post-variable selection inference associated LASSO and other variable selection methods for the linear models as in Berk et al. (2013) , Lee et al. (2014), Taylor et al. (2014), van de Geer et al. (2013) and Zhang and Zhang (2014). These methods are based on the sparsity assumption that the non-zero regression coefficients are sparsely populated. The sparsity assumption is quite hard to be validated from data. Our proposed tests are valid without the sparsity assumption and may be used first when the sparsity level of a testing problem is unknown. More research is needed on how to combine the two strains of inference methods together in the setting of high dimensional generalized linear models.

Appendix

In this section, we provide technical proofs to the main results reported in Section 3-5. To establish the results of the paper, we introduce three lemmas whose proofs are available in Chen and Guo (2014).

We define a few notations:

$$\epsilon_i = Y_i - g(X_i^{\mathrm{T}}\beta), \quad \epsilon_{0i} = Y_i - g(X_i^{\mathrm{T}}\beta_0), \quad V_{0i} = V\{g(X_i^{\mathrm{T}}\beta_0)\}, \quad \psi_{0i} = g'(X_i^{\mathrm{T}}\beta_0)/V\{g(X_i^{\mathrm{T}}\beta_0)\}.$$

Lemma A.1. *The expectations and variances of A_n and U_n are respectively*

$$\mu_{\mathrm{A}_n} = \mathrm{tr}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0}\}, \quad \mu_{\mathrm{U}_n} = (n-1)\Delta_{\beta,\beta_0}^{\mathrm{T}}\Delta_{\beta,\beta_0},$$

$$\sigma_{\mathrm{A}_n}^2 = n^{-1}\big[E\big\{\epsilon_0^4\psi_0^4(X^{\mathrm{T}}X)^2\big\} - E^2\big\{\epsilon_0^2\psi_0^2(X^{\mathrm{T}}X)\big\}\big] \quad \textit{and}$$

$$\sigma_{\mathrm{U}_n}^2 = 4(n-2)(1-n^{-1})\xi_1 + 2(1-n^{-1})\xi_2$$

where $\xi_1 = \Delta_{\beta,\beta_0}^{\mathrm{T}}\{\Sigma_\beta(\beta_0)+\Xi_{\beta,\beta_0}\}\Delta_{\beta,\beta_0} - (\Delta_{\beta,\beta_0}^{\mathrm{T}}\Delta_{\beta,\beta_0})^2$ *and* $\xi_2 = \mathrm{tr}\{\Sigma_\beta(\beta_0)+\Xi_{\beta,\beta_0}\}^2 - (\Delta_{\beta,\beta_0}^{\mathrm{T}}\Delta_{\beta,\beta_0})^2$.

Lemma A.2. *Under Assumptions 3.1-3.4 and the "local" alternatives $\mathscr{L}_\beta$,*

$$\sigma_{\mathrm{U}_n}^2 = 2\mathrm{tr}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0}\}^2\{1 + o(1)\} \quad \textit{as } n \to \infty \textit{ and } p \to \infty.$$

Lemma A.3. *Under Assumptions 3.1-3.4 and the "local" alternatives $\mathscr{L}_\beta$,*

$$\frac{\mathrm{tr}\{\widehat{\Sigma_{\beta_0}^2(\beta_0)}\}}{\mathrm{tr}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0}\}^2} \xrightarrow{p} 1 \quad \textit{and} \quad \frac{\mathrm{tr}^2\{\widehat{\Sigma_{\beta_0}(\beta_0)}\}}{\mathrm{tr}^2\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0}\}} \xrightarrow{p} 1 \quad \textit{as } n \to \infty.$$

In the following, we provide technical proofs for the main results in Section 4 first, since they are used to establish the results in Section 3. The results in Section 5 are given the last.

Proof of Theorem 3

From the mean value theorem, under the Assumptions 3.1, 3.3, 3.4 and the "local" alternatives, we can show that there is a positive constant C such that $\Xi_{\beta,\beta_0} \le C\Sigma_{\mathrm{X}}$. Details are given in Chen and Guo (2014). Define $\sigma_n^2 = 2\mathrm{tr}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0}\}^2$. Notice that

$$U_n - (n-1)\Delta_{\beta,\beta_0}^{\mathrm{T}}\Delta_{\beta,\beta_0} = V_{n1} + V_{n2} \quad \text{where}$$

$$V_{n1} = \frac{1}{n}\sum_{i\neq j}^{n}(\Delta_{\beta,\beta_0}^{\mathrm{T}}\epsilon_{0j}\psi_{0j}X_j + \Delta_{\beta,\beta_0}^{\mathrm{T}}\epsilon_{0i}\psi_{0i}X_i - 2\Delta_{\beta,\beta_0}^{\mathrm{T}}\Delta_{\beta,\beta_0}) \quad \text{and}$$

$$V_{n2} = \frac{1}{n}\sum_{i\neq j}^{n}(\epsilon_{0i}\psi_{0i}X_i - \Delta_{\beta,\beta_0})^{\mathrm{T}}(\epsilon_{0j}\psi_{0j}X_j - \Delta_{\beta,\beta_0}).$$

As $E(V_{n1}) = 0$ and from the Hoeffding decomposition in the proof of Lemma A.1, under the "local" alternatives $\mathscr{L}_\beta$,

$$\mathrm{var}(V_{n1}) = o(\sigma_n^2) \quad \text{and} \quad V_{n1} = o_p(\sigma_n).$$

We use the martingale central limit theorem to show the asymptotic normality of V_{n2}. Let

$$Z_{n,i} = \frac{2}{n\sigma_n}\sum_{j=1}^{i-1}(\epsilon_{0i}\psi_{0i}X_i - \Delta_{\beta,\beta_0})^{\mathrm{T}}(\epsilon_{0j}\psi_{0j}X_j - \Delta_{\beta,\beta_0}) \quad \text{for } i \geq 2, \text{ and}$$

$T_{n,k} = \sum_{i=2}^{k} Z_{n,i}$. Then $T_{n,n} = \sum_{i=2}^{n} Z_{n,i} = V_{n2}/\sigma_n$.

Let $\mathscr{F}_k = \sigma\left\{\left(\begin{smallmatrix} X_1 \\ \epsilon_1 \end{smallmatrix}\right), \ldots, \left(\begin{smallmatrix} X_k \\ \epsilon_k \end{smallmatrix}\right)\right\}$ be the σ-fields generated by $\left(\begin{smallmatrix} X_i \\ \epsilon_i \end{smallmatrix}\right)$ for $i = 1, \ldots, k$. It can be verified that $T_{n,k}$ is a martingale. For $i = 2, \ldots, n$, let $v_{n,i} = E(Z_{n,i}^2|\mathscr{F}_{i-1})$ and $v_n = \sum_{i=2}^{n} v_{n,i}$.

From Hall and Heyde (1980), in order to show the asymptotic normality of V_{n2}, we need to verify the following two conditions:

$$v_n \xrightarrow{p} 1 \quad \text{as } n \to \infty \text{ and } p \to \infty; \tag{A.1}$$

$$\text{for any } \eta > 0, \ \sum_{i=2}^{n} E\{Z_{n,i}^2 I(|Z_{n,i}| > \eta)\} \to 0 \ \text{ as } n \to \infty \text{ and } p \to \infty. \tag{A.2}$$

We first establish (A.1). For $i = 2, \ldots, n$,

$$\begin{aligned} v_{n,i} =& \frac{4}{n^2\sigma_n^2}\Bigg[\sum_{j=1}^{i-1}(\epsilon_{0j}\psi_{0j}X_j - \Delta_{\beta,\beta_0})^{\mathrm{T}}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0} - \Delta_{\beta,\beta_0}\Delta_{\beta,\beta_0}^{\mathrm{T}}\}(\epsilon_{0j}\psi_{0j}X_j - \Delta_{\beta,\beta_0}) \\ &+ \sum_{j_1 \neq j_2}^{i-1}(\epsilon_{0j_1}\psi_{0j_1}X_{j_1} - \Delta_{\beta,\beta_0})^{\mathrm{T}}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0} - \Delta_{\beta,\beta_0}\Delta_{\beta,\beta_0}^{\mathrm{T}}\}(\epsilon_{0j_2}\psi_{0j_2}X_{j_2} - \Delta_{\beta,\beta_0})\Bigg]. \end{aligned}$$

Then

$$v_n = \sum_{i=2}^{n} v_{n,i} = C_1 + C_2 \quad \text{where}$$

$$C_1 = \frac{4}{n^2\sigma_n^2}\sum_{j=1}^{n-1}\left[(n-j)(\epsilon_{0j}\psi_{0j}X_j - \Delta_{\beta,\beta_0})^{\mathrm{T}}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0} - \Delta_{\beta,\beta_0}\Delta_{\beta,\beta_0}^{\mathrm{T}}\}(\epsilon_{0j}\psi_{0j}X_j - \Delta_{\beta,\beta_0})\right] \ \text{ and}$$

$$C_2 = \frac{8}{n^2\sigma_n^2}\sum_{1 \leq j_1 < j_2 \leq n-1}\left[(n-j_2)(\epsilon_{0j_1}\psi_{0j_1}X_{j_1} - \Delta_{\beta,\beta_0})^{\mathrm{T}}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0} - \Delta_{\beta,\beta_0}\Delta_{\beta,\beta_0}^{\mathrm{T}}\}(\epsilon_{0j_2}\psi_{0j_2}X_{j_2} - \Delta_{\beta,\beta_0})\right].$$

Under the "local" alternatives $\mathscr{L}_\beta$, we have

$$E\left[(\epsilon_{0j}\psi_{0j}X_j - \Delta_{\beta,\beta_0})^{\mathrm{T}}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0} - \Delta_{\beta,\beta_0}\Delta_{\beta,\beta_0}^{\mathrm{T}}\}(\epsilon_{0j}\psi_{0j}X_j - \Delta_{\beta,\beta_0})\right] = \mathrm{tr}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0}\}^2\{1 + o(1)\}.$$

Thus $E(C_1) = 1 + o(1)$. Similar to the proof of Lemma A.3 in Chen and Guo (2014),

$$\begin{aligned} \mathrm{var}(C_1) &= \frac{16}{n^4\sigma_n^4}\sum_{j=1}^{n-1} E\left[(n-j)(\epsilon_{0j}\psi_{0j}X_j - \Delta_{\beta,\beta_0})^{\mathrm{T}}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0} - \Delta_{\beta,\beta_0}\Delta_{\beta,\beta_0}^{\mathrm{T}}\}(\epsilon_{0j}\psi_{0j}X_j - \Delta_{\beta,\beta_0})\right]^2 \\ &\leq \frac{16}{n^4\sigma_n^4}\sum_{j=1}^{n-1}(n-j)^2 O[\mathrm{tr}^2\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0}\}^2] = O(n^{-1}). \end{aligned}$$

Therefore $C_1 \xrightarrow{p} 1$. For C_2, we note that $E(C_2) = 0$ and

$$\mathrm{var}(C_2) = \frac{64}{n^4\sigma_n^4} \sum_{1\le j_1<j_2\le n-1} (n-j_2)^2 \mathrm{tr}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0} - \Delta_{\beta,\beta_0}\Delta_{\beta,\beta_0}^{\mathrm{T}}\}^4 = o(1).$$

Thus, $C_2 \xrightarrow{p} 0$. Hence, (A.1) holds.

Next, we verify (A.2). Notice that for any $\eta > 0$,

$$\sum_{i=2}^n E\{Z_{n,i}^2 I(|Z_{n,i}| > \eta)\} \le \frac{1}{\eta^2}\sum_{i=2}^n E(Z_{n,i}^4) \quad \text{and}$$

$$\sum_{i=2}^n E(Z_{n,i}^4) = \frac{16}{n^4\sigma_n^4}\sum_{i=2}^n E\Big\{\sum_{j=1}^{i-1}(\epsilon_{0i}\psi_{0i}X_i - \Delta_{\beta,\beta_0})^{\mathrm{T}}(\epsilon_{0j}\psi_{0j}X_j - \Delta_{\beta,\beta_0})\Big\}^4 = P_1 + P_2$$

where

$$P_1 = \frac{16}{n^4\sigma_n^4}\sum_{i=2}^n\sum_{j=1}^{i-1} E\{(\epsilon_{0i}\psi_{0i}X_i - \Delta_{\beta,\beta_0})^{\mathrm{T}}(\epsilon_{0j}\psi_{0j}X_j - \Delta_{\beta,\beta_0})\}^4 \quad \text{and}$$

$$P_2 = \frac{16}{n^4\sigma_n^4}\sum_{i=2}^n\sum_{j_1\ne j_2}^{i-1} E[\{(\epsilon_{0i}\psi_{0i}X_i-\Delta_{\beta,\beta_0})^{\mathrm{T}}(\epsilon_{0j_1}\psi_{0j_1}X_{j_1}-\Delta_{\beta,\beta_0})(\epsilon_{0i}\psi_{0i}X_i-\Delta_{\beta,\beta_0})^{\mathrm{T}}(\epsilon_{0j_2}\psi_{0j_2}X_{j_2}-\Delta_{\beta,\beta_0})\}^2].$$

By Lemma A.3 in Chen and Guo (2014) and the Cauchy-Schwartz inequality, the orders of P_1 and P_2 are respectively

$$P_1 = O(n^{-2}) \quad \text{and} \quad P_2 = O(n^{-1}).$$

Then, we obtain $\sum_{i=2}^n E(Z_{n,i}^4) = o(1)$ and the desired asymptotic normality of U_n. □

Proof of Theorem 4

We first show that under the "fixed" alternatives $\mathscr{L}_\beta^{\mathrm{F}}$,

$$\frac{U_n - (n-1)\Delta_{\beta,\beta_0}^{\mathrm{T}}\Delta_{\beta,\beta_0}}{\{2\mathrm{tr}(\Xi_{\beta,\beta_0}^2)\}^{1/2}} \xrightarrow{d} N(0,1) \quad \text{as } n\to\infty \text{ and } p\to\infty. \tag{A.3}$$

Similar to the proof of Theorem 3,

$$U_n - (n-1)\Delta_{\beta,\beta_0}^{\mathrm{T}}\Delta_{\beta,\beta_0} = V_{n1} + V_{n2} + V_{n3} + V_{n4} \tag{A.4}$$

where

$$V_{n1} = \frac{1}{n}\sum_{i\ne j}^n \epsilon_i\epsilon_j\psi_{0i}\psi_{0j}X_i^{\mathrm{T}}X_j,$$

$$V_{n2} = \frac{1}{n}\sum_{i\ne j}^n\big[\{(g_i - g_{0i})\epsilon_j\psi_{0i}\psi_{0j}X_i^{\mathrm{T}}X_j\} + \{(g_j - g_{0j})\epsilon_i\psi_{0i}\psi_{0j}X_i^{\mathrm{T}}X_j\}\big],$$

$$V_{n3} = \frac{1}{n}\sum_{i\neq j}^{n}\big[\Delta_{\beta,\beta_0}^{\mathrm{T}}(g_i - g_{0i})\psi_{0i}X_i + \Delta_{\beta,\beta_0}^{\mathrm{T}}(g_j - g_{0j})\psi_{0j}X_j - 2\Delta_{\beta,\beta_0}^{\mathrm{T}}\Delta_{\beta,\beta_0}\big] \quad \text{and}$$

$$V_{n4} = \frac{1}{n}\sum_{i\neq j}^{n}\big[\{(g_i - g_{0i})\psi_{0i}X_i - \Delta_{\beta,\beta_0}\}^{\mathrm{T}}\{(g_j - g_{0j})\psi_{0j}X_j - \Delta_{\beta,\beta_0}\}\big].$$

Notice that V_{ni} are statistics with zero mean for $i = 1, \cdots, 4$. Similar to Lemma A.1, we can show

$$\mathrm{var}(V_{n1}) = 2(1 - n^{-1})\mathrm{tr}\{\Sigma_\beta^2(\beta_0)\} = o\{\mathrm{tr}(\Xi_{\beta,\beta_0}^2)\};$$

$$\mathrm{var}(V_{n2}) = 4(n-2)(1 - n^{-1})\Delta_{\beta,\beta_0}^{\mathrm{T}}\Sigma_\beta(\beta_0)\Delta_{\beta,\beta_0} + 4(1 - n^{-1})\mathrm{tr}\{\Sigma_\beta(\beta_0)\Xi_{\beta,\beta_0}\} = o\{\mathrm{tr}(\Xi_{\beta,\beta_0}^2)\};$$

$$\mathrm{var}(V_{n3}) \leq 4(n-2)(1 - n^{-1})\Delta_{\beta,\beta_0}^{\mathrm{T}}\Xi_{\beta,\beta_0}\Delta_{\beta,\beta_0} + 4(1 - n^{-1})\Delta_{\beta,\beta_0}^{\mathrm{T}}\Xi_{\beta,\beta_0}\Delta_{\beta,\beta_0} = o\{\mathrm{tr}(\Xi_{\beta,\beta_0}^2)\}.$$

Then

$$V_{n1} = o_p\{\mathrm{tr}^{1/2}(\Xi_{\beta,\beta_0}^2)\}, \quad V_{n2} = o_p\{\mathrm{tr}^{1/2}(\Xi_{\beta,\beta_0}^2)\} \quad \text{and} \quad V_{n3} = o_p\{\mathrm{tr}^{1/2}(\Xi_{\beta,\beta_0}^2)\}.$$

Applying the same technique we used in the proof of Theorem 3, we have

$$\frac{V_{n4}}{\{2\mathrm{tr}(\Xi_{\beta,\beta_0}^2)\}^{1/2}} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty \text{ and } p \to \infty.$$

Then from the decomposition (A.4), the asymptotic normality (A.3) holds. The power expression stated in the theorem is readily available from Lemma A.3. □

Proof of Theorem 1

Note that

$$\widehat{S}_n = 1 + \frac{\{(\mathbb{Y} - \mu_0)\circ\Psi_0\}^{\mathrm{T}}(\mathbb{X}\mathbb{X}^{\mathrm{T}} - \mathbb{D})\{(\mathbb{Y} - \mu_0)\circ\Psi_0\}/n}{\{(\mathbb{Y} - \mu_0)\circ\Psi_0\}^{\mathrm{T}}\mathbb{D}\{(\mathbb{Y} - \mu_0)\circ\Psi_0\}/n} = 1 + \frac{U_n}{A_n}.$$

Let $\mu_{\mathrm{U}_n} = E(U_n) = (n-1)\Delta_{\beta,\beta_0}^{\mathrm{T}}\Delta_{\beta,\beta_0}$ and $\mu_{\mathrm{A}_n} = E(A_n) = \mathrm{tr}\{\Sigma_\beta(\beta_0) + \Xi_{\beta,\beta_0}\}$. From the Taylor expansion,

$$\begin{aligned}
\widehat{S}_n =& 1 + \frac{\mu_{\mathrm{U}_n} + (U_n - \mu_{\mathrm{U}_n})}{\mu_{\mathrm{A}_n}(1 + \frac{A_n - \mu_{\mathrm{A}_n}}{\mu_{\mathrm{A}_n}})} \\
=& 1 + \mu_{\mathrm{A}_n}^{-1}\left\{1 - \frac{A_n - \mu_{\mathrm{A}_n}}{\mu_{\mathrm{A}_n}} + (\frac{A_n - \mu_{\mathrm{A}_n}}{\mu_{\mathrm{A}_n}})^2 + \cdots\right\}\{\mu_{\mathrm{U}_n} + (U_n - \mu_{\mathrm{U}_n})\} \\
=& 1 + \mu_{\mathrm{A}_n}^{-1}\mu_{\mathrm{U}_n} - \mu_{\mathrm{A}_n}^{-1}\mu_{\mathrm{U}_n}(\frac{A_n - \mu_{\mathrm{A}_n}}{\mu_{\mathrm{A}_n}}) + \mu_{\mathrm{A}_n}^{-1}(U_n - \mu_{\mathrm{U}_n}) + \mu_{\mathrm{A}_n}^{-1}\mu_{\mathrm{U}_n}(\frac{A_n - \mu_{\mathrm{A}_n}}{\mu_{\mathrm{A}_n}})^2 + \cdots.
\end{aligned} \tag{A.5}$$

Under the "local" alternatives $\mathscr{L}_\beta$, from Lemma A.3 in Chen and Guo (2014),

$$\sigma_{\mathrm{A}_n}^2 \leq n^{-1}E\{\epsilon_0^4\psi_0^4(X^{\mathrm{T}}X)^2\} = O\{n^{-1}\mathrm{tr}^2(\Sigma_{\mathrm{X}})\}, \quad \sigma_{\mathrm{A}_n}^2/\mu_{\mathrm{A}_n}^2 = O(n^{-1}) \quad \text{and}$$

$$\sigma_{\mathrm{U}_n}^{-1}(U_n - \mu_{\mathrm{U}_n}) \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty \text{ and } p \to \infty.$$

Observe that

$$\mathrm{var}\{\mu_{\mathrm{A}_n}^{-2}\mu_{\mathrm{U}_n}(A_n-\mu_{\mathrm{A}_n})\}=\sigma_{\mathrm{A}_n}^2\mu_{\mathrm{U}_n}^2\mu_{\mathrm{A}_n}^{-4}\quad\text{and}\quad\mathrm{var}\{\mu_{\mathrm{A}_n}^{-1}(U_n-\mu_{\mathrm{U}_n})\}=\sigma_{\mathrm{U}_n}^2\mu_{\mathrm{A}_n}^{-2}.$$

From the fact that

$$(\sigma_{\mathrm{A}_n}^2\mu_{\mathrm{U}_n}^2)/(\mu_{\mathrm{A}_n}^2\sigma_{\mathrm{U}_n}^2)=O\big(n(\Delta_{\beta,\beta_0}^{\mathrm{T}}\Delta_{\beta,\beta_0})^2/\mathrm{tr}\{\Sigma_\beta(\beta_0)+\Xi_{\beta,\beta_0}\}^2\big)=o(1),$$

we have $\mu_{\mathrm{A}_n}^{-2}\mu_{\mathrm{U}_n}(A_n-\mu_{\mathrm{A}_n})=o_p(\sigma_{\mathrm{U}_n}\mu_{\mathrm{A}_n}^{-1})$.

Regarding the higher order terms in the expansion (A.5), for $k\geq 1$,

$$(\frac{A_n-\mu_{\mathrm{A}_n}}{\mu_{\mathrm{A}_n}})^k(\frac{U_n-\mu_{\mathrm{U}_n}}{\mu_{\mathrm{A}_n}})=(\frac{A_n-\mu_{\mathrm{A}_n}}{\sigma_{\mathrm{A}_n}})^k(\frac{U_n-\mu_{\mathrm{U}_n}}{\sigma_{\mathrm{U}_n}})\frac{\sigma_{\mathrm{A}_n}^k}{\mu_{\mathrm{A}_n}^k}\frac{\sigma_{\mathrm{U}_n}}{\mu_{\mathrm{A}_n}}=O_p(n^{-\frac{k}{2}}\frac{\sigma_{\mathrm{U}_n}}{\mu_{\mathrm{A}_n}}).$$

Note that under the "local" alternatives $\mathscr{L}_\beta$,

$$\frac{\sigma_{\mathrm{A}_n}^2}{\mu_{\mathrm{A}_n}^2}\frac{\mu_{\mathrm{U}_n}}{\sigma_{\mathrm{U}_n}}=O(n^{-1})\frac{(n-1)\Delta_{\beta,\beta_0}^{\mathrm{T}}\Delta_{\beta,\beta_0}}{[2\mathrm{tr}\{\Sigma_\beta(\beta_0)+\Xi_{\beta,\beta_0}\}^2]^{1/2}}\{1+o(1)\}=o(1).$$

Hence, for $k\geq 2$,

$$\mu_{\mathrm{A}_n}^{-1}\mu_{\mathrm{U}_n}(\frac{A_n-\mu_{\mathrm{A}_n}}{\mu_{\mathrm{A}_n}})^k=(\frac{A_n-\mu_{\mathrm{A}_n}}{\sigma_{\mathrm{A}_n}})^k\frac{\sigma_{\mathrm{A}_n}^{k-2}}{\mu_{\mathrm{A}_n}^{k-2}}\frac{\sigma_{\mathrm{A}_n}^2}{\mu_{\mathrm{A}_n}^2}\frac{\mu_{\mathrm{U}_n}}{\sigma_{\mathrm{U}_n}}\frac{\sigma_{\mathrm{U}_n}}{\mu_{\mathrm{A}_n}}=o_p(n^{-\frac{k-2}{2}}\frac{\sigma_{\mathrm{U}_n}}{\mu_{\mathrm{A}_n}})=o_p(\frac{\sigma_{\mathrm{U}_n}}{\mu_{\mathrm{A}_n}}).$$

Therefore,

$$\widehat{S}_n=1+\mu_{\mathrm{A}_n}^{-1}\mu_{\mathrm{U}_n}+\mu_{\mathrm{A}_n}^{-1}(U_n-\mu_{\mathrm{U}_n})+o_p(\sigma_{\mathrm{U}_n}\mu_{\mathrm{A}_n}^{-1}).\tag{A.6}$$

From the asymptotic normality of U_n and the Slutsky theorem, we have

$$\sigma_{\mathrm{U}_n}^{-1}\mu_{\mathrm{A}_n}(\widehat{S}_n-1-\mu_{\mathrm{A}_n}^{-1}\mu_{\mathrm{U}_n})\xrightarrow{d}N(0,1)\quad\text{as } n\to\infty \text{ and } p\to\infty.$$

□

Proof of Theorem 2

For brevity, we define

$$\begin{aligned}\sigma_{\mathrm{G}}^2=&\mathrm{var}\big\{-\mu_{\mathrm{U}_n}\mu_{\mathrm{A}_n}^{-2}(A_n-\mu_{\mathrm{A}_n})+\mu_{\mathrm{A}_n}^{-1}(U_n-\mu_{\mathrm{U}_n})\big\}\\=&\sigma_{\mathrm{U}_n}^2\mu_{\mathrm{A}_n}^{-2}\{1+\tau(\tau-2\rho_{\mathrm{A}_n,\mathrm{U}_n})\}\end{aligned}$$

where $\tau^2=(\sigma_{\mathrm{A}_n}^2\mu_{\mathrm{U}_n}^2)/(\sigma_{\mathrm{U}_n}^2\mu_{\mathrm{A}_n}^2)$ and $\rho_{\mathrm{A}_n,\mathrm{U}_n}$ is the correlation coefficient between A_n and U_n. It is straightforward to show that, by Cauchy-Schwartz inequality,

$$\mathrm{cov}(A_n,U_n)\leq 2E^{1/2}\big[\epsilon_0^4\psi_0^4(X^{\mathrm{T}}X)^2-E^2\{\epsilon_0^2\psi_0^2(X^{\mathrm{T}}X)\}\big]\big[\Delta_{\beta,\beta_0}^{\mathrm{T}}\{\Xi_{\beta,\beta_0}+\Sigma_\beta(\beta_0)\}\Delta_{\beta,\beta_0}\big]^{1/2}.$$

Notice that under the "fixed" alternatives $\mathscr{L}_\beta^{\mathrm{F}}$,

$$E\big[\epsilon_0^4\psi_0^4(X^{\mathrm{T}}X)^2-E^2\{\epsilon_0^2\psi_0^2(X^{\mathrm{T}}X)\}\big]=n\sigma_{\mathrm{A}_n}^2\quad\text{and}\quad\Delta_{\beta,\beta_0}^{\mathrm{T}}\{\Xi_{\beta,\beta_0}+\Sigma_\beta(\beta_0)\}\Delta_{\beta,\beta_0}=o(n^{-1}\sigma_{\mathrm{U}_n}^2).$$

Therefore

$$\mathrm{cov}(A_n, U_n) = o(\sigma_{\mathrm{A}_n}\sigma_{\mathrm{U}_n}) \quad \text{and} \quad \rho_{\mathrm{A}_n,\mathrm{U}_n} = o(1).$$

Recall that

$$\tau^2 = \frac{\sigma^2_{\mathrm{A}_n}\mu^2_{\mathrm{U}_n}}{\sigma^2_{\mathrm{U}_n}\mu^2_{\mathrm{A}_n}} = \frac{n(\Delta^{\mathrm{T}}_{\beta,\beta_0}\Delta_{\beta,\beta_0})^2}{\mathrm{tr}(\Xi^2_{\beta,\beta_0})}\left[\frac{E\{\epsilon_0^4\psi_0^4(X^{\mathrm{T}}X)^2\}}{E^2\{\epsilon_0^2\psi_0^2(X^{\mathrm{T}}X)\}} - 1\right]\{1+o(1)\}.$$

The condition (3.11) implies

$$\frac{n(\Delta^{\mathrm{T}}_{\beta,\beta_0}\Delta_{\beta,\beta_0})^2}{\mathrm{tr}(\Xi^2_{\beta,\beta_0})}\frac{E\{\epsilon_0^4\psi_0^4(X^{\mathrm{T}}X)^2\}}{E^2\{\epsilon_0^2\psi_0^2(X^{\mathrm{T}}X)\}} \asymp 1.$$

Thus, $\tau^2 \asymp 1$ and $\sigma^2_{\mathrm{G}} = \sigma^2_{\mathrm{U}_n}\mu^{-2}_{\mathrm{A}_n}\{1+\tau^2+o(1)\}$. Observe that

$$\frac{\sigma^2_{\mathrm{A}_n}}{\mu^2_{\mathrm{A}_n}} \le \frac{1}{n}\frac{E\{\epsilon_0^4\psi_0^4(X^{\mathrm{T}}X)^2\}}{E^2\{\epsilon_0^2\psi_0^2(X^{\mathrm{T}}X)\}} \asymp n^{-2\delta}.$$

Regarding the higher order terms in (A.5), for $k \ge 1$,

$$(\frac{A_n-\mu_{\mathrm{A}_n}}{\mu_{\mathrm{A}_n}})^k(\frac{U_n-\mu_{\mathrm{U}_n}}{\mu_{\mathrm{A}_n}}) = (\frac{A_n-\mu_{\mathrm{A}_n}}{\sigma_{\mathrm{A}_n}})^k(\frac{U_n-\mu_{\mathrm{U}_n}}{\sigma_{\mathrm{U}_n}})\frac{\sigma^k_{\mathrm{A}_n}}{\mu^k_{\mathrm{A}_n}}\frac{\sigma_{\mathrm{U}_n}}{\mu_{\mathrm{A}_n}} = O_p(n^{-k\delta}\frac{\sigma_{\mathrm{U}_n}}{\mu_{\mathrm{A}_n}}) = O_p(n^{-k\delta}\sigma_{\mathrm{G}}).$$

Notice that under the "fixed" alternatives $\mathscr{L}^{\mathrm{F}}_\beta$

$$\frac{\sigma^2_{\mathrm{A}_n}}{\mu^2_{\mathrm{A}_n}}\frac{\mu_{\mathrm{U}_n}}{\sigma_{\mathrm{U}_n}} = O(n^{-2\delta})\frac{n\Delta^{\mathrm{T}}_{\beta,\beta_0}\Delta_{\beta,\beta_0}}{\{2\mathrm{tr}(\Xi^2_{\beta,\beta_0})\}^{1/2}}\{1+o(1)\} = O(n^{-\delta}) = o(1).$$

Then, for $k \ge 2$,

$$\mu^{-1}_{\mathrm{A}_n}\mu_{\mathrm{U}_n}(\frac{A_n-\mu_{\mathrm{A}_n}}{\mu_{\mathrm{A}_n}})^k = (\frac{A_n-\mu_{\mathrm{A}_n}}{\sigma_{\mathrm{A}_n}})^k(\frac{\sigma^{k-2}_{\mathrm{A}_n}}{\mu^{k-2}_{\mathrm{A}_n}})(\frac{\sigma^2_{\mathrm{A}_n}}{\mu^2_{\mathrm{A}_n}}\frac{\mu_{\mathrm{U}_n}}{\sigma_{\mathrm{U}_n}})\frac{\sigma_{\mathrm{U}_n}}{\mu_{\mathrm{A}_n}} = o_p(n^{-(k-2)\delta}\frac{\sigma_{\mathrm{U}_n}}{\mu_{\mathrm{A}_n}}) = o_p(\sigma_{\mathrm{G}}).$$

It follows that,

$$\widehat{S}_n = 1 + \mu^{-1}_{\mathrm{A}_n}\mu_{\mathrm{U}_n} - \mu_{\mathrm{U}_n}\mu^{-2}_{\mathrm{A}_n}(A_n-\mu_{\mathrm{A}_n}) + \mu^{-1}_{\mathrm{A}_n}(U_n-\mu_{\mathrm{U}_n}) + o_p(\sigma_{\mathrm{G}}).$$

From the joint asymptotic normalities of $A_n-\mu_{\mathrm{A}_n}$ and $U_n-\mu_{\mathrm{U}_n}$, we have

$$\sigma^{-1}_{\mathrm{G}}(\widehat{S}_n - 1 - \mu^{-1}_{\mathrm{A}_n}\mu_{\mathrm{U}_n}) \xrightarrow{d} N(0,1)$$

where $\sigma^2_{\mathrm{G}} = \sigma^2_{\mathrm{U}_n}\mu^{-2}_{\mathrm{A}_n}\{1+\tau^2+o(1)\}$.

Analogous to the proof of Lemma A.3, we can show that under the "fixed" alternatives $\mathscr{L}^{\mathrm{F}}_\beta$,

$$\sigma^{-1}_{\mathrm{G}}\left[2\mathrm{tr}\{\widehat{\Sigma^2_\beta(\beta_0)}\}/\mathrm{tr}^2\{\widehat{\Sigma_\beta(\beta_0)}\}\right]^{1/2} \xrightarrow{p} \frac{1}{(1+\tau^2)^{1/2}} \quad \text{as } n\to\infty.$$

Together with the asymptotic normality of $\widehat{S}_n$, we complete the proof. □

Proof of Theorem 5

The proofs of Theorem 5 and Lemma A.6 show that the proposed test procedure is invariant to scale transformation of Y, hence, invariant to the dispersion parameter ϕ in the multiplicative variance function $V(\mu, \phi) = \phi V(\mu)$. For a general form of the variance function, we can use the similar technique as in Zhang and Zhang (2014) and Lockhart et al. (2014). Hence, we decide to just consider the nuisance parameter $\theta = \beta^{(1)}$, otherwise, the already lengthy proof would be even complicated. We divide the proof into the following two lemmas.

Lemma A.4. *Suppose Assumptions 3.1, 3.3, 5.1-5.4 hold, then under the H_{01},*

$$\frac{\widetilde{U}_n}{[2\mathrm{tr}\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]^{1/2}} \xrightarrow{d} N(0,1) \quad \textit{as } n \to \infty \textit{ and } p_2 \to \infty.$$

Lemma A.5. *Suppose Assumptions 3.1, 3.3, 5.1-5.4 hold, under the "local" alternatives $\mathscr{L}_{\beta^{(2)}}$,*

$$\frac{\widetilde{U}_n - n\Delta_{\beta,\beta_0^*}^{(2)\mathrm{T}}\Delta_{\beta,\beta_0^*}^{(2)}}{[2\mathrm{tr}\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]^{1/2}} \xrightarrow{d} N(0,1) \quad \textit{as } n \to \infty \textit{ and } p_2 \to \infty.$$

Proof of Lemma A.4

Recall that $\hat{\beta}_0^{(1)}$ is the maximum quasi-likelihood estimator of $\beta^{(1)}$ under H_{01} and $\beta^{*(1)} = \beta^{(1)}$. For notational convenience, we let $\hat{\beta}_0 = (\hat{\beta}_0^{(1)\mathrm{T}}, \beta_0^{(2)\mathrm{T}})^\mathrm{T}$, $\beta_0 = (\beta^{(1)\mathrm{T}}, \beta_0^{(2)\mathrm{T}})^\mathrm{T}$ and

$$\hat{\mu}_{0i} = g(X_i^\mathrm{T}\hat{\beta}_0), \quad \mu_{0i} = g(X_i^\mathrm{T}\beta_0), \quad \hat{\psi}_{0i} = g'(X_i^\mathrm{T}\hat{\beta}_0)/V\{g(X_i^\mathrm{T}\hat{\beta}_0)\}, \quad \psi_{0i} = g'(X_i^\mathrm{T}\beta_0)/V\{g(X_i^\mathrm{T}\beta_0)\};$$

$$\widehat{\mu}_0 = (\hat{\mu}_{01}, \ldots, \hat{\mu}_{0n})^\mathrm{T}, \quad \mu_0 = (\mu_{01}, \ldots, \mu_{0n})^\mathrm{T}, \quad \widehat{\Psi}_0 = (\hat{\psi}_{01}, \ldots, \hat{\psi}_{0n})^\mathrm{T}, \quad \Psi_0 = (\psi_{01}, \ldots, \psi_{0n})^\mathrm{T};$$

$$g'_{0i} = \partial g(t)/\partial t \mid_{t=X_i^\mathrm{T}\beta_0}, \quad \psi'_{0i} = \partial\psi\{g(t)\}/\partial g \mid_{t=X_i^\mathrm{T}\beta_0}.$$

Define $\widehat{\mathbb{D}} = (\mathbb{Y} - \widehat{\mu}_0) \circ \widehat{\Psi}_0 = \{(Y_1 - \hat{\mu}_{01})\hat{\psi}_{01}, \ldots, (Y_n - \hat{\mu}_{0n})\hat{\psi}_{0n}\}^\mathrm{T}$. Then we can write $\widetilde{U}_n$ as

$$\widetilde{U}_n = n^{-1}\{(\mathbb{Y} - \widehat{\mu}_0) \circ \widehat{\Psi}_0\}^\mathrm{T}(\mathbb{X}^{(2)}\mathbb{X}^{(2)\mathrm{T}} - \mathbb{M})\{(\mathbb{Y} - \widehat{\mu}_0) \circ \widehat{\Psi}_0\} = n^{-1}\widehat{\mathbb{D}}^\mathrm{T}(\mathbb{X}^{(2)}\mathbb{X}^{(2)\mathrm{T}} - \mathbb{M})\widehat{\mathbb{D}}$$

where $\mathbb{M}$ is the diagonal matrix with diagonal elements being those of $\mathbb{X}^{(2)}\mathbb{X}^{(2)\mathrm{T}}$.

Following the approach in Le Cessie and Van Houwelingen (1991), we have

$$\widehat{\mathbb{D}} = [\mathbb{I}_n + (\mathbb{W}_2 - \mathbb{W}_1)\mathbb{X}^{(1)}\{\mathbb{I}(\beta^{(1)})\}^{-1}\mathbb{X}^{(1)\mathrm{T}}]\mathbb{D} \tag{A.7}$$

where $\mathbb{I}_n$ is the $n \times n$ identity matrix, $\mathbb{W}_1$ and $\mathbb{W}_2$ are two diagonal matrices defined as

$$\mathbb{W}_1 = \mathrm{diag}\{\psi_{01}^2 E(\epsilon_{01}^2|X_1), \ldots, \psi_{0n}^2 E(\epsilon_{0n}^2|X_n)\} \quad \text{and}$$

$$\mathbb{W}_2 = \mathrm{diag}\{\psi'_{01}g'_{01}(Y_1 - g_{01}), \ldots, \psi'_{0n}g'_{0n}(Y_n - g_{0n})\}.$$

Moreover, $\mathbb{D} = (\mathbb{Y} - \mu_0) \circ \Psi_0$ and $\mathbb{I}(\beta^{(1)})$ is a $p_1 \times p_1$ matrix given by $\mathbb{I}(\beta^{(1)}) = \mathbb{X}^{(1)\mathrm{T}}\mathbb{W}_1\mathbb{X}^{(1)}$.

In order to simplify the notations, let

$$\mathbb{A} = (\mathbb{X}^{(2)}\mathbb{X}^{(2)\mathrm{T}} - \mathbb{M}) = (a_{ij})_{n\times n} \quad \text{and} \quad \mathbb{B} = \mathbb{X}^{(1)}\{\mathbb{I}(\beta^{(1)})\}^{-1}\mathbb{X}^{(1)\mathrm{T}} = (b_{ij})_{n\times n}.$$

Therefore, by (A.7), we can decompose the statistic $\widetilde{U}_n$ as

$$\begin{aligned}\widetilde{U}_n &= n^{-1}\mathbb{D}^{\mathrm{T}}\mathbb{A}\mathbb{D} + n^{-1}\mathbb{D}^{\mathrm{T}}\mathbb{B}\mathbb{W}_1\mathbb{A}\mathbb{W}_1\mathbb{B}\mathbb{D} + n^{-1}\mathbb{D}^{\mathrm{T}}\mathbb{B}\mathbb{W}_2\mathbb{A}\mathbb{W}_2\mathbb{B}\mathbb{D} \\ &\quad + 2n^{-1}\mathbb{D}^{\mathrm{T}}\mathbb{A}\mathbb{W}_2\mathbb{B}\mathbb{D} - 2n^{-1}\mathbb{D}^{\mathrm{T}}\mathbb{B}\mathbb{W}_1\mathbb{A}\mathbb{W}_2\mathbb{B}\mathbb{D} - 2n^{-1}\mathbb{D}^{\mathrm{T}}\mathbb{A}\mathbb{W}_1\mathbb{B}\mathbb{D} \\ &= T_{n1} + T_{n2} + T_{n3} + 2T_{n4} - 2T_{n5} - 2T_{n6}, \quad \text{say.}\end{aligned}$$

Notice that under H_{01}, by the properties of conditional expectation and Assumption 3.3,

$$E(\epsilon_{0i}|X_i^{(2)}) = E\{E(\epsilon_{0i}|X_i)|X_i^{(2)}\} = 0, \qquad E(\epsilon_{0i}^2|X_i^{(2)}) = E\{E(\epsilon_{0i}^2|X_i)|X_i^{(2)}\} \geq K_1$$

$$\text{and} \quad E(\epsilon_{0i}^8|X_i^{(2)}) = E\{E(\epsilon_{0i}^8|X_i)|X_i^{(2)}\} \leq K_2.$$

From Assumption 3.1, we can partition X_i and Γ respectively as

$$X_i = \begin{pmatrix} X^{(1)}_{i_{p_1\times 1}} \\ \\ X^{(2)}_{i_{p_2\times 1}} \end{pmatrix} = \begin{pmatrix} \Gamma_1 Z_i \\ \Gamma_2 Z_i \end{pmatrix} \quad \text{and} \quad \Gamma = \begin{pmatrix} \Gamma_{1_{p_1\times m}} \\ \Gamma_{2_{p_2\times m}} \end{pmatrix}.$$

Furthermore, we have $\Sigma_{\mathrm{X}^{(2)}} = \Gamma_2\Gamma_2^{\mathrm{T}}$. This indicates that the model in Assumption 3.1 still holds for $X_i^{(2)}$, except we replace Σ_{X} as $\Sigma_{\mathrm{X}^{(2)}}$, Γ as Γ_2.

Under the null hypothesis, by Assumptions 3.1, 3.3, 5.1 and 5.4, the same technique used in the proof of Theorem 3 leads to

$$\frac{T_{n1}}{[2\mathrm{tr}\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]^{1/2}} \xrightarrow{d} N(0,1) \quad \text{as } n\to\infty \text{ and } p_2\to\infty.$$

In the following proofs, we denote all the constants by C which may vary from place to place.

Observe that

$$|T_{n2}| \leq n^{-1}|\mathbb{D}^{\mathrm{T}}\mathbb{B}\mathbb{W}_1\mathbb{A}\mathbb{W}_1\mathbb{B}\mathbb{D}| \leq n^{-1}(|\lambda_{\max}(\mathbb{A})| \wedge |\lambda_{\min}(\mathbb{A})|)\mathbb{D}^{\mathrm{T}}\mathbb{B}\mathbb{W}_1^2\mathbb{B}\mathbb{D}. \tag{A.8}$$

By the method of Lan et al. (2014), we can show that

$$|\lambda_{\max}(\mathbb{A})| = O_p\big(n^{3/4}\mathrm{tr}^{1/2}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big) \tag{A.9}$$

and the same order holds for $|\lambda_{\min}(\mathbb{A})|$.

From the independence among the observations and $E(\epsilon_{0i}|X_i) = 0$,

$$E(\mathbb{D}^{\mathrm{T}}\mathbb{B}\mathbb{W}_1^2\mathbb{B}\mathbb{D}) = \sum_{i=1}^n\sum_{k=1}^n E(b_{ik}^2\psi_{0i}^2\psi_{0k}^2 g_{0i}'^2\epsilon_{0k}^2) \leq C\sum_{i=1}^n\sum_{k=1}^n E(b_{ik}^2) = CE\{\mathrm{tr}(\mathbb{B}^2)\}. \tag{A.10}$$

Notice that from Assumptions 3.3 and 5.4, we have

$$\mathbb{I}(\beta^{(1)}) = \mathbb{X}^{(1)\mathrm{T}}\mathbb{W}_1\mathbb{X}^{(1)} = \sum_{i=1}^{n}\{\psi_{0i}^2 E(\epsilon_{0i}^2|X_i)X_i^{(1)}X_i^{(1)\mathrm{T}}\} \geq K_1c_1\sum_{i=1}^{n}X_i^{(1)}X_i^{(1)\mathrm{T}} = K_1c_1\mathbb{X}^{(1)\mathrm{T}}\mathbb{X}^{(1)}.$$

Together with the matrix inequality from Seber (2008), we get

$$E\{\mathrm{tr}(\mathbb{B}^2)\} = E\big(\mathrm{tr}[\mathbb{X}^{(1)}\{\mathbb{I}(\beta^{(1)})\}^{-1}\mathbb{X}^{(1)\mathrm{T}}\mathbb{X}^{(1)}\{\mathbb{I}(\beta^{(1)})\}^{-1}\mathbb{X}^{(1)\mathrm{T}}]\big) \leq K_1^{-2}c_1^{-2}p_1. \tag{A.11}$$

Thus, the order of T_{n2} is

$$T_{n2} = O(n^{-1})O_p\big(n^{3/4}\mathrm{tr}^{1/2}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big)O_p(p_1) = o_p\big(\mathrm{tr}^{1/2}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big).$$

Applying the same technique, we can show

$$T_{n3} = o_p\big(\mathrm{tr}^{1/2}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big).$$

For the order of T_{n4}, to simplify the notations, we define

$$\mathbb{D}^{\mathrm{T}}\mathbb{A}\mathbb{W}_2 = \left(\sum_{k=1}^{n}a_{k1}\psi_{01}'g_{01}'\psi_{0k}\epsilon_{0k}\epsilon_{01},\ldots,\sum_{k=1}^{n}a_{kn}\psi_{0n}'g_{0n}'\psi_{0k}\epsilon_{0k}\epsilon_{0n}\right) = (c_{01},\ldots,c_{0n}), \quad \text{say;}$$

$$\mathbb{B}\mathbb{D} = \left(\sum_{k=1}^{n}b_{1k}\psi_{0k}\epsilon_{0k},\ldots,\sum_{k=1}^{n}b_{nk}\psi_{0k}\epsilon_{0k}\right)^{\mathrm{T}} = (f_{01},\ldots,f_{0n})^{\mathrm{T}}, \quad \text{say.}$$

Then

$$E(\mathbb{D}^{\mathrm{T}}\mathbb{A}\mathbb{W}_2\mathbb{B}\mathbb{D})^2 = \sum_{i=1}^{n}E(c_{0i}^2f_{0i}^2) + \sum_{i_1\neq i_2}^{n}E(c_{0i_1}c_{0i_2}f_{0i_1}f_{0i_2}). \tag{A.12}$$

We can write

$$\sum_{i=1}^{n}E(c_{0i}^2f_{0i}^2) = T_{41} + 2T_{42} + 2T_{43} \tag{A.13}$$

where

$$T_{41} = \sum_{i=1}^{n}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}E(a_{k_1i}^2b_{ik_2}^2\psi_{0i}'^2g_{0i}'^2\psi_{0k_1}^2\psi_{0k_2}^2\epsilon_{0k_1}^2\epsilon_{0k_2}^2\epsilon_{0i}^2),$$

$$T_{42} = \sum_{i=1}^{n}\sum_{k=1}^{n}E(a_{ki}^2b_{ii}b_{ik}\psi_{0k}^3\psi_{0i}\psi_{0i}'^2g_{0i}'^2\epsilon_{0k}^3\epsilon_{0i}^3) \quad \text{and}$$

$$T_{43} = \sum_{i=1}^{n}\sum_{k_1\neq k_2}^{n}E(a_{k_1i}a_{k_2i}b_{ik_1}b_{ik_2}\psi_{0k_1}^2\psi_{0k_2}^2\psi_{0i}'^2g_{0i}'^2\epsilon_{0k_1}^2\epsilon_{0k_2}^2\epsilon_{0i}^2).$$

Notice that

$$T_{41} \leq C\sum_{i=1}^{n}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}E(a_{k_1i}^2b_{ik_2}^2).$$

From the Cauchy-Schwartz inequality,

$$\sum_{i=1}^{n}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n} E(a_{k_1 i}^2 b_{ik_2}^2) \le E^{1/2}\Big\{\sum_{i=1}^{n}\Big(\sum_{k_1=1}^{n} a_{k_1 i}^2\Big)^2\Big\} E^{1/2}\Big\{\sum_{i=1}^{n}\Big(\sum_{k_2=1}^{n} b_{ik_2}^2\Big)^2\Big\}. \tag{A.14}$$

Recall that $a_{ii} = 0$, from Lemma A.2 in Chen and Guo (2014), we have

$$E\Big\{\sum_{i=1}^{n}\Big(\sum_{k_1=1}^{n} a_{k_1 i}^2\Big)^2\Big\} = O\big(n^3 \mathrm{tr}^2[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big). \tag{A.15}$$

On the other hand, $\mathrm{rank}(\mathbb{B}) \le p_1$ and employing the same technique as we used in the derivation of (A.11), we have

$$E\Big\{\sum_{i=1}^{n}\Big(\sum_{k_2=1}^{n} b_{ik_2}^2\Big)^2\Big\} \le E\{\mathrm{tr}^2(\mathbb{B}^2)\} \le E\{\mathrm{rank}(\mathbb{B})\mathrm{tr}(\mathbb{B}^4)\} = O(p_1^2). \tag{A.16}$$

Combining (A.14), (A.15) and (A.16),

$$T_{41} = O(n^{3/2})O_p(p_1)O_p\big(\mathrm{tr}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big) = o_p\big(n^{7/4}\mathrm{tr}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big).$$

Applying the same method to that for T_{41}, we can show

$$T_{42} = o_p\big(n^{7/4}\mathrm{tr}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big) \quad \text{and} \quad T_{43} = o_p\big(n^{7/4}\mathrm{tr}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big).$$

Then from (A.13), we have

$$\sum_{i=1}^{n} E(c_{0i}^2 f_{0i}^2) = o\big(n^{7/4}\mathrm{tr}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big). \tag{A.17}$$

Derivations given in Chen and Guo (2014) show

$$\sum_{i_1 \ne i_2}^{n} E(c_{0i_1} c_{0i_2} f_{0i_1} f_{0i_2}) = o\big(n^2 \mathrm{tr}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big). \tag{A.18}$$

Combining (A.12), (A.17) and (A.18), we have

$$T_{n4} = o_p\big(\mathrm{tr}^{1/2}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big).$$

From (A.9), (A.10) and (A.11), we can show

$$|T_{n5}| \le n^{-1}(|\lambda_{\max}(\mathbb{A})| \wedge |\lambda_{\min}(\mathbb{A})|)(\mathbb{D}^{\mathrm{T}}\mathbb{B}\mathbb{W}_1^2\mathbb{B}\mathbb{D})^{1/2}(\mathbb{D}^{\mathrm{T}}\mathbb{B}\mathbb{W}_2^2\mathbb{B}\mathbb{D})^{1/2} = o_p\big(\mathrm{tr}^{1/2}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big).$$

For the order of T_{n6}, notice that $\mathbb{B}$ is a non-negative matrix, then

$$E|T_{n6}| = n^{-1}E|\mathbb{D}^{\mathrm{T}}\mathbb{A}\mathbb{W}_1\mathbb{B}\mathbb{D}| \le n^{-1}E^{1/2}(\mathbb{D}^{\mathrm{T}}\mathbb{A}\mathbb{W}_1\mathbb{B}\mathbb{W}_1\mathbb{A}\mathbb{D})E^{1/2}(\mathbb{D}^{\mathrm{T}}\mathbb{B}\mathbb{D}).$$

By the definitions of $\mathbb{W}_1$ and $\mathbb{D}$, it is straightforward to see

$$\mathbb{W}_1 = \mathrm{diag}\{\psi_{01}^2 E(\epsilon_{01}^2|X_1), \cdots, \psi_{0n}^2 E(\epsilon_{0n}^2|X_n)\} = E(\mathbb{D}\mathbb{D}^{\mathrm{T}}|\mathbb{X}).$$

Applying some basic matrix inequalities, we have

$$E(\mathbb{D}^{\mathrm{T}}\mathbb{A}\mathbb{W}_1\mathbb{B}\mathbb{W}_1\mathbb{A}\mathbb{D}) = \mathrm{tr}E(\mathbb{W}_1\mathbb{A}\mathbb{W}_1\mathbb{B}\mathbb{W}_1\mathbb{A}) \leq E^{1/2}\{\mathrm{tr}(\mathbb{A}\mathbb{W}_1\mathbb{A}\mathbb{W}_1^2\mathbb{A}\mathbb{W}_1\mathbb{A})\}E^{1/2}\{\mathrm{tr}(\mathbb{B}\mathbb{W}_1^2\mathbb{B})\}$$

$$\text{and} \quad E\{\mathrm{tr}(\mathbb{A}\mathbb{W}_1\mathbb{A}\mathbb{W}_1^2\mathbb{A}\mathbb{W}_1\mathbb{A})\} \leq CE\{\mathrm{tr}(\mathbb{A}^4)\} = O\big(n^3\mathrm{tr}^2[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big).$$

It can be shown that $E\{\mathrm{tr}(\mathbb{B}\mathbb{W}_1^2\mathbb{B})\} = O(p_1)$ and $E(\mathbb{D}^{\mathrm{T}}\mathbb{B}\mathbb{D}) = O(p_1)$. Thus,

$$T_{n6} = O(n^{-1})O_p\big(n^{3/4}\mathrm{tr}^{1/2}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big)O(p_1^{1/4})O_p(p_1^{1/2}) = o_p\big(\mathrm{tr}^{1/2}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big).$$

Therefore, the asymptotic normality of T_{n1} and the orders of $T_{n2}, \ldots, T_{n6}$ lead to

$$\frac{\widetilde{U}_n}{[2\mathrm{tr}\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]^{1/2}} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty \text{ and } p_2 \to \infty.$$

□

Proof of Lemma A.5

Define

$$\beta_0^* = (\beta^{*(1)\mathrm{T}}, \beta_0^{(2)\mathrm{T}})^{\mathrm{T}}, \quad \beta = (\beta^{(1)\mathrm{T}}, \beta^{(2)\mathrm{T}})^{\mathrm{T}}, \quad g_{0i}^* = g(X_i^{\mathrm{T}}\beta_0^*), \quad g_i = g(X_i^{\mathrm{T}}\beta),$$

$$\psi_{0i}^* = g'(X_i^{\mathrm{T}}\beta_0^*)/V\{g(X_i^{\mathrm{T}}\beta_0^*)\}, \quad g_{0i}^{*\prime} = \partial g(t)/\partial t \mid_{t=X_i^{\mathrm{T}}\beta_0^*},$$

$$\psi_{0i}^{*\prime} = \partial\psi\{g(t)\}/\partial g \mid_{t=X_i^{\mathrm{T}}\beta_0^*}, \quad \epsilon_i = Y_i - g_i \quad \text{and} \quad \epsilon_{0i}^* = Y_i - g_{0i}^*.$$

Similar derivations to those used in the proof of Lemma A.4 show that, under the "local" alternatives $\mathscr{L}_{\beta^{(2)}}$, we have

$$\widehat{\mathbb{D}} = \{\mathbb{I}_n + (\mathbb{W}_{20}^* - \mathbb{W}_{10}^*)\mathbb{B}^*\}\mathbb{D}^*$$

where

$$\mathbb{D}^* = (\epsilon_{01}^*\psi_{01}^*, \ldots, \epsilon_{0n}^*\psi_{0n}^*)^{\mathrm{T}}, \quad \mathbb{B}^* = \mathbb{X}^{(1)}\{\mathbb{I}(\beta^{*(1)})\}^{-1}\mathbb{X}^{(1)\mathrm{T}}, \quad \mathbb{I}(\beta^{*(1)}) = \mathbb{X}^{(1)\mathrm{T}}\mathbb{W}_{10}^*\mathbb{X}^{(1)},$$

$$\mathbb{W}_{10}^* = \mathrm{diag}\{\psi_{01}^{*2}E(\epsilon_{01}^{*2}|X_1), \ldots, \psi_{0n}^{*2}E(\epsilon_{0n}^{*2}|X_n)\} \quad \text{and} \quad \mathbb{W}_{20}^* = \mathrm{diag}\{\psi_{01}^{*\prime}g_{01}^{*\prime}\epsilon_{01}^*, \ldots, \psi_{0n}^{\prime}g_{0n}^*\epsilon_{0n}^*\}.$$

Hence,

$$\begin{aligned}\widetilde{U}_n =& n^{-1}[\mathbb{D}^{*\mathrm{T}}\{\mathbb{I}_n + (\mathbb{W}_{20}^* - \mathbb{W}_{10}^*)\mathbb{B}^*\}^{\mathrm{T}}\mathbb{A}\{\mathbb{I}_n + (\mathbb{W}_{20}^* - \mathbb{W}_{10}^*)\mathbb{B}^*\}\mathbb{D}^*] \\ =& n^{-1}\mathbb{D}_1^{*\mathrm{T}}\mathbb{A}^*\mathbb{D}_1^* + n^{-1}\mathbb{D}_2^{*\mathrm{T}}\mathbb{A}^*\mathbb{D}_2^* + 2n^{-1}\mathbb{D}_1^{*\mathrm{T}}\mathbb{A}^*\mathbb{D}_2^* = T_1 + T_2 + T_3, \quad \text{say,}\end{aligned} \tag{A.19}$$

where $\mathbb{D}_1^* = (\epsilon_1\psi_{01}^*, \ldots, \epsilon_n\psi_{0n}^*)^{\mathrm{T}}$, $\mathbb{D}_2^* = \{(g_1 - g_{01}^*)\psi_{01}^*, \ldots, (g_n - g_{0n}^*)\psi_{0n}^*\}^{\mathrm{T}}$ and

$$\mathbb{A}^* = \{\mathbb{I}_n + (\mathbb{W}_{20}^* - \mathbb{W}_{10}^*)\mathbb{B}^*\}^{\mathrm{T}}\mathbb{A}\{\mathbb{I}_n + (\mathbb{W}_{20}^* - \mathbb{W}_{10}^*)\mathbb{B}^*\}.$$

Derivations in Chen and Guo (2014) demonstrate that

$$\frac{T_1}{[2\mathrm{tr}\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]^{1/2}} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty \text{ and } p_2 \to \infty. \tag{A.20}$$

$$T_2 - n\Delta_{\beta,\beta_0^*}^{(2)\mathrm{T}}\Delta_{\beta,\beta_0^*}^{(2)} = o_p\big(\mathrm{tr}^{1/2}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big) \quad \text{and} \tag{A.21}$$

$$T_3 = o_p\big(\mathrm{tr}^{1/2}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big). \tag{A.22}$$

Combining the results in (A.19)-(A.22), under the "local" alternatives $\mathscr{L}_{\beta^{(2)}}$, we have

$$\frac{\widetilde{U}_n - n\Delta_{\beta,\beta_0^*}^{(2)\mathrm{T}}\Delta_{\beta,\beta_0^*}^{(2)}}{[2\mathrm{tr}\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]^{1/2}} \xrightarrow{d} N(0,1) \quad \text{as } n \to \infty \text{ and } p_2 \to \infty.$$

□

Lemma A.6. *Under Assumptions 3.1, 3.3, 5.1-5.4 and H_{01},*

$$\frac{\widehat{R}_n}{\mathrm{tr}\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2} \xrightarrow{p} 1 \quad \textit{as } n \to \infty.$$

Proof of Lemma A.6

Derivations given in Chen and Guo (2014) show that

$$\max_{1\le i\le n} |\hat{\mu}_{0i} - \mu_{0i}| = o_p(1) \quad \text{and} \quad \max_{1\le i\le n} |\hat{\psi}_{0i}^2 - \psi_{0i}^2| = o_p(1). \tag{A.23}$$

Note that

$$(Y_i - \hat{\mu}_{0i})^2 = \epsilon_{0i}^2 + (\hat{\mu}_{0i} - \mu_{0i})^2 + 2(\hat{\mu}_{0i} - \mu_{0i})\epsilon_{0i} \quad \text{and} \quad \hat{\psi}_{0i}^2 = \psi_{0i}^2 + \hat{\psi}_{0i}^2 - \psi_{0i}^2. \tag{A.24}$$

Thus we can write

$$\widehat{R}_n = \frac{1}{n(n-1)}\sum_{i\neq j}^{n} \epsilon_{0i}^2\epsilon_{0j}^2\psi_{0i}^2\psi_{0j}^2(X_i^{(2)\mathrm{T}}X_j^{(2)})^2 + \Delta R_n = R_n + \Delta R_n, \quad \text{say.}$$

ΔR_n is straightforward to obtain from (A.24) and the definition of $\widehat{R}_n$ and hence is omitted here. Similar to the proofs in Lemma A.3, we have $R_n = \mathrm{tr}\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2 + o_p\big(\mathrm{tr}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big)$. Analogous to Lemma A.1, we can show $\Delta R_n = o_p\big(\mathrm{tr}[\{\Sigma_\beta^{(2)}(\beta_0^*)\}^2]\big)$ using (A.23). Hence we complete the proof of the proposition. □

Proof of Corollary 3

Similar to the proofs in Proposition A.6, under the "local" alternatives $\mathscr{L}_{\beta^{(2)}}$,

$$\max_{1\le i\le n} |\hat{\mu}_{0i} - \mu^*_{0i}| = o_p(1) \quad \text{and} \quad \max_{1\le i\le n} |\hat{\psi}^2_{0i} - \psi^{*2}_{0i}| = o_p(1),$$

where $\mu^*_{0i} = g(X_i^{\mathrm{T}}\beta^*_0)$, $\psi^*_0 = \psi(X, \beta^*_0)$ and $\beta^*_0 = (\beta^{*(1)\mathrm{T}}, \beta^{(2)\mathrm{T}}_0)^{\mathrm{T}}$. Define $\Sigma^*_0 = E(\epsilon^{*2}_0 \psi^{*2}_0 X^{(2)} X^{(2)\mathrm{T}})$ where $\epsilon^*_0 = Y - \mu^*_0$. Employing the same technique as we used in Proposition A.6,

$$\widehat{R}_n / \{\mathrm{tr}(\Sigma^{*2}_0)\} \xrightarrow{p} 1 \quad \text{as } n \to \infty \text{ and } p_2 \to \infty. \tag{A.25}$$

Under the "local" alternatives $\mathscr{L}_{\beta^{(2)}}$, it can be shown that

$$\mathrm{tr}(\Sigma^{*2}_0) - \mathrm{tr}[\{\Sigma^{(2)}_\beta(\beta^*_0)\}^2] = o\big(\mathrm{tr}[\{\Sigma^{(2)}_\beta(\beta^*_0)\}^2]\big).$$

This together with (A.25) implies

$$\frac{\widehat{R}_n}{\mathrm{tr}[\{\Sigma^{(2)}_\beta(\beta^*_0)\}^2]} \xrightarrow{p} 1 \quad \text{as } n \to \infty.$$

Hence, the power of the test is

$$\widetilde{\Omega}^{(2)}(\beta, \beta^*_0) = \Phi\left(-z_\alpha + \frac{n\Delta^{(2)\mathrm{T}}_{\beta,\beta^*_0}\Delta^{(2)}_{\beta,\beta^*_0}}{\left[2\mathrm{tr}\{\Sigma^{(2)}_\beta(\beta^*_0)\}^2\right]^{1/2}}\right)\{1 + o(1)\}.$$

This completes the proof the corollary. □

References

Auer, P. L. and Doerge, R. W. (2010). Statistical design and analysis of RNA sequencing data. *Genetics*, **185**, 405-416.

Bai, Z. D. and Saranadasa, H. (1996). Effect of high dimension: by an example of two sample problem. *Statistica Sinica*, **6**, 311-329.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **57**, 289-300.

Berk, R., Brown, L., Buja, A., Zhang, K. and Zhao, L. (2013). Valid post-selection inference. *The Annals of Statistics*, **41**, 802-837.

Chang, J., Tang, C. Y. and Wu, Y. (2013). Marginal empirical likelihood and sure independence feature screening. *The Annals of Statistics*, **41**, 2123-2148.

Chen, S. X. and Cui, H. J. (2003). An extended empirical likelihood for generalized linear models. *Statistica Sinica*, **13**, 69-81.

Chen, S. X. and Guo, B. (2014). Tests for high dimensional generalized linear models. Technical report, Guanghua School of Managment, Peking University.

Chen, S. X., Peng, L. and Qin, Y. L. (2009). Effects of data dimension on empirical likelihood. *Biometrika*, **96**, 711-722.

Chen, S. X. and Qin, Y. L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *The Annals of Statistics*, **38**, 808-835.

Chen, S. X., Zhang, L. X. and Zhong, P. S. (2010). Tests for high-dimensional covariance matrices. *Journal of the American Statistical Association*, **105**, 810-819.

Chiaretti, S., Li, X., Gentleman, R., Vitale, A., Vignetti, M., Mandelli, F., Ritz, J. and Foa, R. (2004). Gene expression profile of adult T-cell acute lymphocytic leukemia identifies distinct subsets of patients with different response to therapy and survival. *Blood*, **103**, 2771-2778.

Dudoit, S., Keles, S. and van der Laan, M. J. (2008) Multiple tests of association with biological annotation metadata. *Institute of Mathematical Statistics. Collections*, **2**, 153-218.

Efron, B. and Tibshirani, R. (2007). On testing the significance of sets of genes. *The Annals of Applied Statistics*, **1**, 107-129.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate statistical modelling based on generalized linear models* (2nd edition). Springer, New York.

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

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

Gentleman, R., Irizarry, R. A., Carey, V. J., Dudoit, S. and Huber, W. (2005). *Bioinformatics and Computational Biology Solutions Using R and Bioconductor.* Springer, New York.

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 I error control. *Biometrika*, **98**, 381-390.

Hall, P., and Heyde, C. C. (1980). *Martingale limit theory and its application.* Academic Press.

Lan, W., Wang, H. and Tsai, C. L. (2014). Testing covariates in high-dimensional regression. *Annals of Institute of Statistical Mathematics*, DOI: 10.1007/s10463-013-0414-0.

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

Lee, J. D., Sun, D. L., Sun, Y. and Taylor, J. E. (2014). Exact inference after model selection via the Lasso. *arXiv:1311.6238*.

Li, J. and Chen, S. X. (2012). Two sample tests for high-dimensional covariance matrices. *The Annals of Statistics*, **40**, 908-940.

Lockhart, R., Taylor, J., Tibshirani, J. R. and Tibshirani, R. (2014). A significance test for the lasso (with discussion). *The Annals of Statistics*.

Lund, S., Nettleton, D., McCarthy, D. and Smyth, G. (2012). Detecting differential expression in RNA-sequence data using quasi-likelihood with shrunken dispersion estimates. *Statistical applications in genetics and molecular biology*, **11**, 8.

McCullagh, P. (1983). Quasi-likelihood functions. *The Annals of Statistics*, **11**, 59-67.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear models* (2nd edition). Chapman and Hall.

Pan, W. (2009). Asymptotic tests of association with multiple SNPs in linkage disequilibrium. *Genetic epidemiology*, **33**, 497-507.

Rahmatallah, Y., Emmert-Streib, F. and Glazko, G. (2012). Gene set analysis for self-contained tests: complex null and specific alternative hypotheses. *Bioinformatics*, **28**, 3073-3080.

Seber G. A. (2008). *A matrix handbook for statisticians*. Wiley, New York.

Serfling, R. J. (1980). *Approximate theorems of mathematical statistics.* Wiley, New York.

Taylor, J., Lockhart, R., Tibshirani, J. R. and Tibshirani, R. (2014). Post-selection adaptive inference for Least Angle Regression and the Lasso. *arXiv:1401.3889*.

van de Geer, S. A. (2008). High-dimensional generalized linear models and the lasso. *The Annals of Statistics*, **36**, 614-645.

van de Geer, S., Bühlmann, P., Ritov, Y. and Dezeure, R. (2013). On asymptotically optimal confidence regions and tests for high-dimensional models. *arXiv:1303.0518.*

van der Vaart, A.W. (2000). *Asymptotic statistics.* Cambridge university press.

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

Wedderburn, R. W. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 439-447.

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.

Zhang, C.-H. and Zhang, 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.

Table 1: Two by two classifications on the number (proportion) of gene-sets rejected/not rejected in the tests with nuisance parameter under the logistic and probit models.

	Probit model					
Logistic model	Rejected	Not rejected	Rejected	Not rejected	Rejected	Not rejected
	Biological Processes		Cellular Components		Molecular Functions	
Rejected	981(0.44)	103(0.05)	140(0.43)	14(0.04)	113(0.28)	40(0.10)
Not rejected	0(0.00)	1161(0.51)	0(0.00)	174(0.53)	19(0.05)	230(0.57)

Figure 1: Empirical power profiles, for testing the global hypothesis, of the proposed test (solid lines with triangles) and the test of Goeman et al. (2011) (dashed lines with circles).

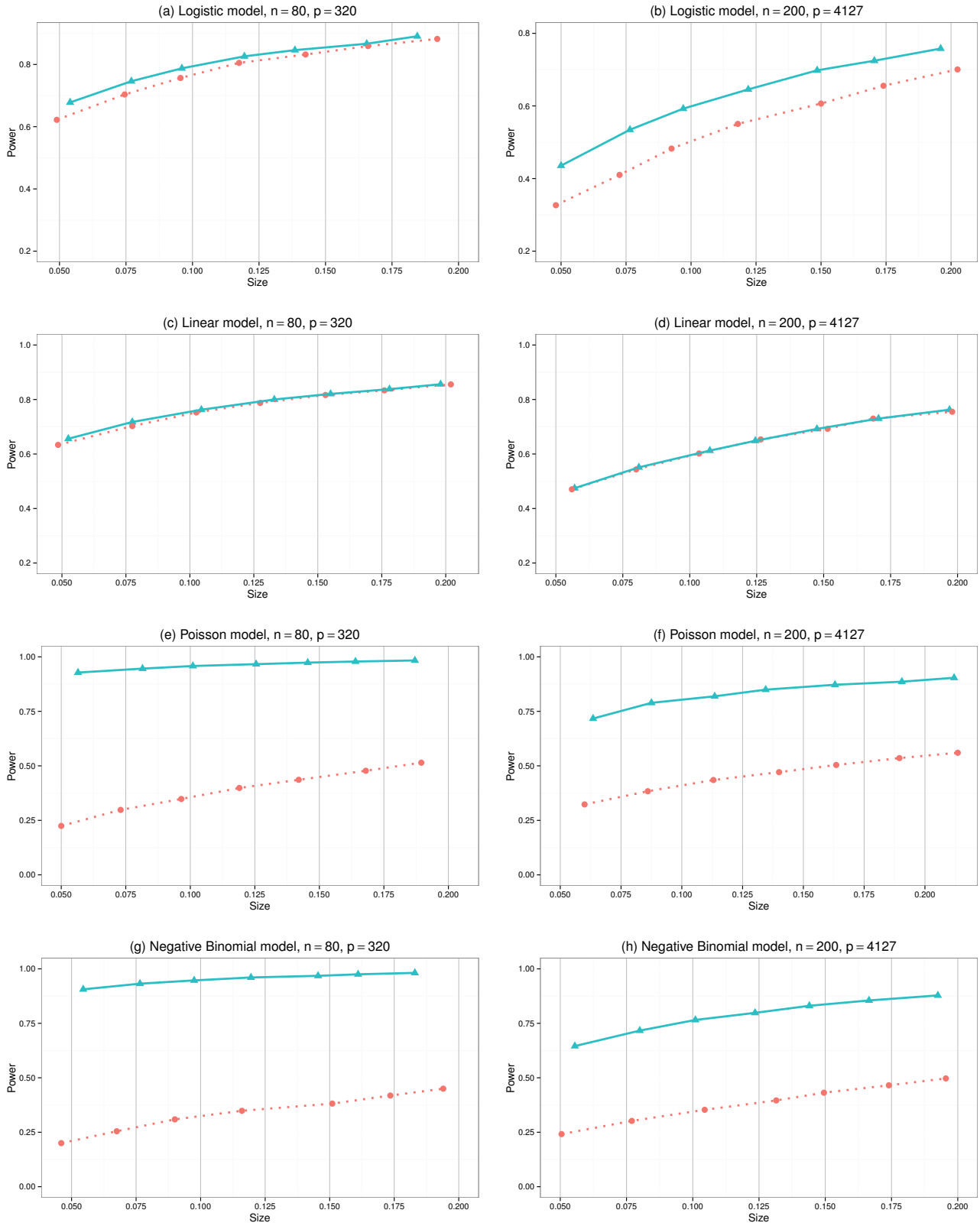

Figure 2: Empirical power profiles, for testing the hypothesis with nuisance parameters, of the proposed test (solid lines with triangles) and the test of Goeman et al. (2011) (dashed lines with circles).

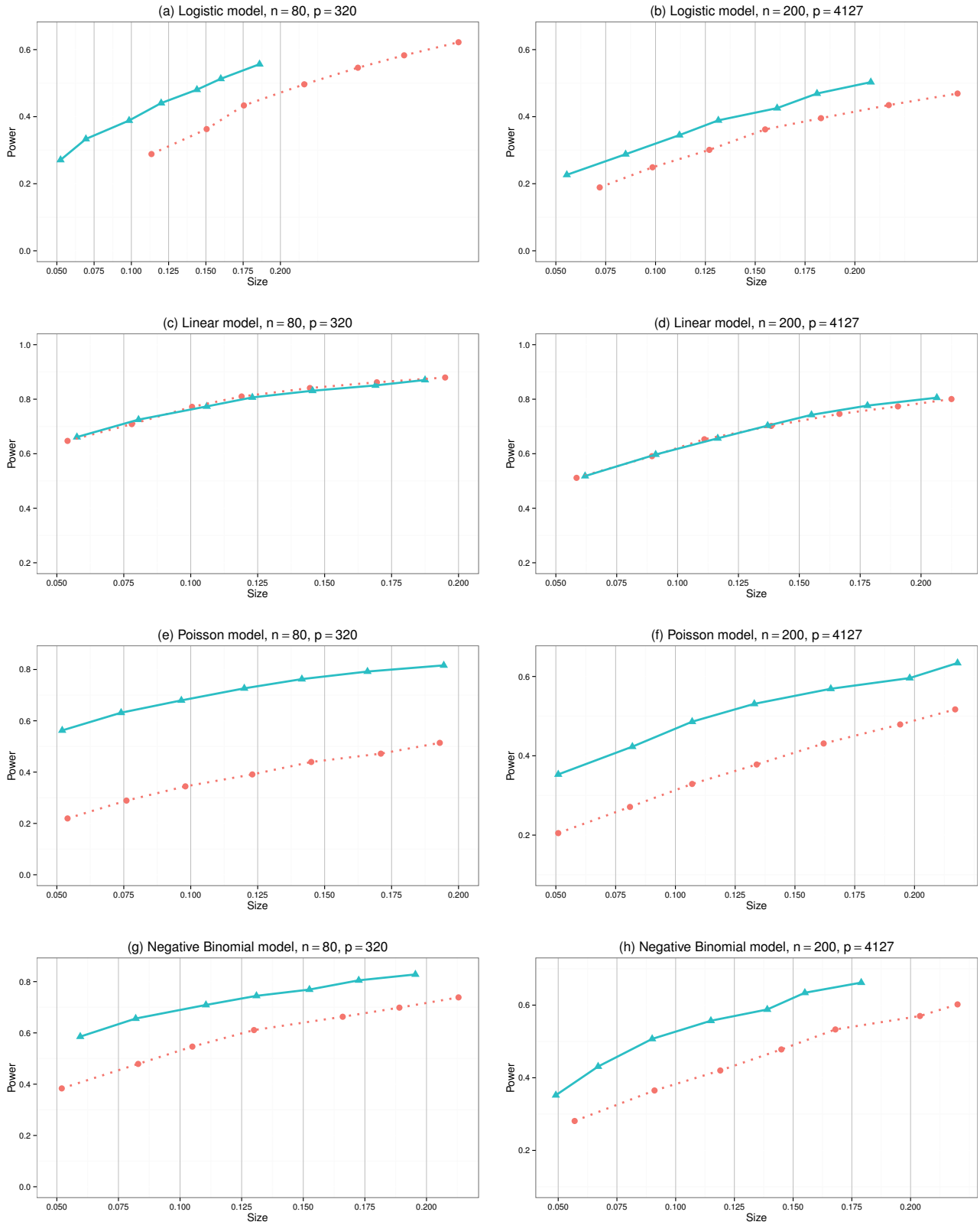

Figure 3: Histograms of p-values (left panels) and the standardized test statistic under the null hypothesis (right panels) for the global hypothesis.

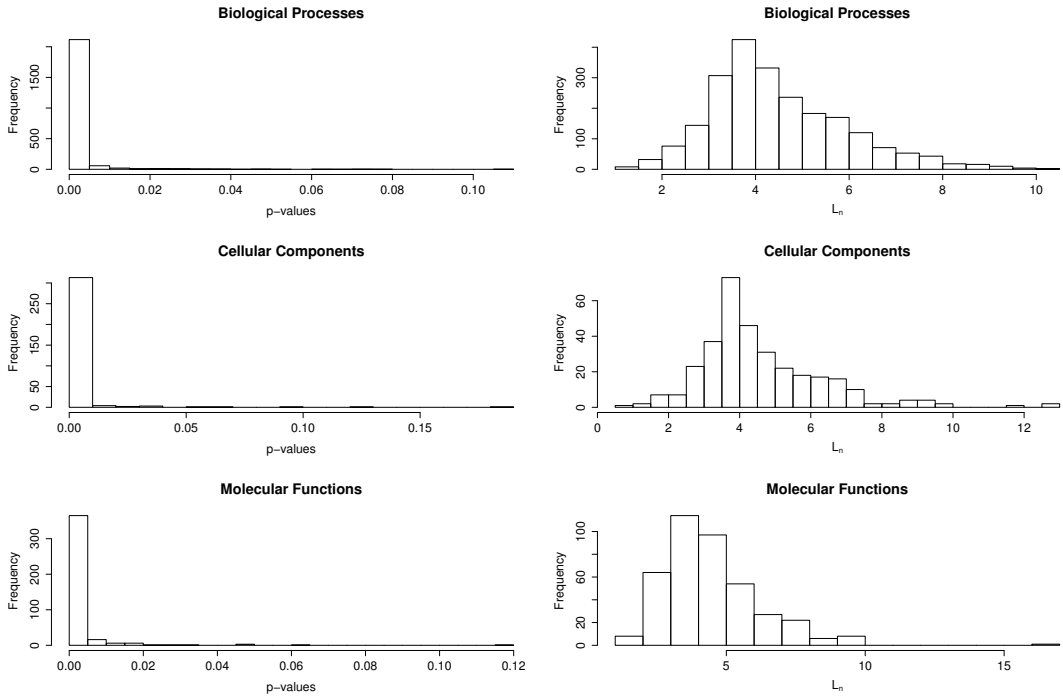

Figure 4: Histograms of p-values (left panels) and the standardized test statistics (right panels) of the proposed test in the presence of the nuisance parameter under the logistic model.

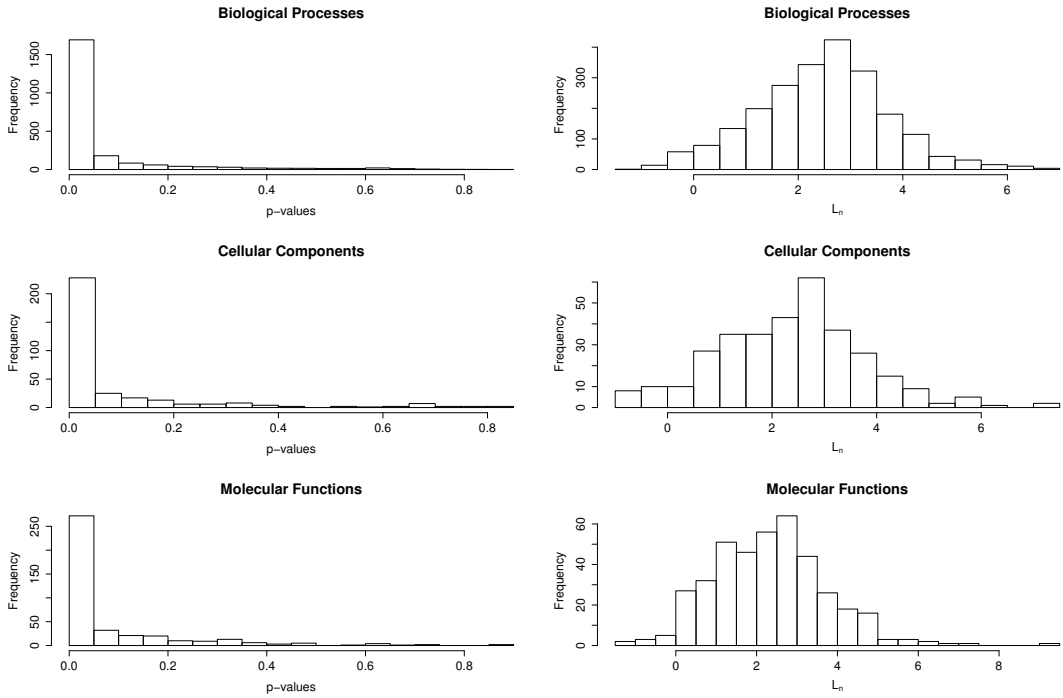

Figure 5: Histograms of p-values (left panels) and the standardized test statistics (right panels) of the proposed test in the presence of the nuisance parameters under the probit model.