Testing Covariates in High Dimensional Regression

Wei Lan, Hansheng Wang

Guanghua School of Management, Peking University

Beijing, P. R. China, 100871

and Chih-Ling Tsai

Graduate School of Management, University of California–Davis

Davis, California 95616

This Version: May 3, 2013

Abstract

In a high dimensional linear regression model, we propose a new procedure for testing statistical significance of a subset of regression coefficients. Specifically, we employ the partial covariances between the response variable and the tested covariates to obtain a test statistic. The resulting test is applicable even if the predictor dimension is much larger than the sample size. Under the null hypothesis, together with boundedness and moment conditions on the predictors, we show that the proposed test statistic is asymptotically standard normal, which is further supported by Monte Carlo experiments. A similar test can be extended to generalized linear models. The practical usefulness of the test is illustrated via an empirical example on paid search advertising.

KEY WORDS: Generalized Linear Model; High Dimensional Data; Hypotheses Testing; Paid Search Advertising; Partial Covariance; Partial F-Test

[†]The research of Wang and Lan were supported in part by National Natural Science Foundation of China (NSFC, 11131002, 11271032), Fox Ying Tong Education Foundation, the Business Intelligence Research Center at Peking University, and the Center for Statistical Science at Peking University. The authors are grateful to the Editor, the AE, and two referees for their helpful comments and advices.

1. INTRODUCTION

Linear regression is arguably one of the mostly important and widely used statistical techniques (Draper and Smith, 1998; Seber and Lee, 2003; Weisberg, 2005). A good summary of various applications can be found in Yandel (1997), Milliken and Johnson (2009), and Vittinghoff et al. (2010), among others. One of the major goals of regression analysis is to model a linear relationship between a response and a set of predictors. To this end, under the assumption that the predictor's dimension (p) is smaller than the sample size (n), we estimate unknown regression coefficients and then test their significances (Lehmann, 1998; Shao, 2003). In addition, we are able to employ the F-test to assess the utility of a model, which allows one to determine whether a significant relationship exists between the dependent variable and the set of all the independent ones (i.e., the full model).

Although the F-test is useful, it cannot be directly applied to testing a subset of variables. Hence, when two competing models are nested, one generally employs the partial F-test (Ravishanker and Dey, 2001; Chatterjee and Hadi, 2006) to check the significance of the additional variables present only under the larger model. This test has been widely used across various fields (e.g., biology, economics, engineering, medicine, psychology, and sociology), and is straightforward to calculate in many software packages (e.g., SAS, SPSS, Minitab, and R). In high-dimensional situations ($n < p$), however, the partial F-test is not applicable. This is because the usual ordinary least squares (OLS) estimator no longer exists, and the OLS estimator is needed for the computation of the classical partial F-test statistics. To solve the problem, Zhong and Chen (2011) proposed a novel test based on a diverging factor model (Bai and Saranadasa, 1996). Their method is useful for linear regression models augmented with a factorial design. Because extending their method to the situation with a general random design matrix is not straightforward, this motivates us to develop a new test to fulfill this theoretical gap.

In this paper, we follow the spirit of the partial F-test (i.e., the partial covariance) to develop a new test statistic. The resulting test enjoys a simple and elegant asymptotic null distribution, namely the standard normal distribution. Accordingly, the proposed test can be easily implemented in practice with a standard normal table. Adopting similar techniques used for linear regression, we also extend the test to generalized linear models with canonical link functions. The rest of this article is organized as follows. Section 2 introduces the model, notation, and technical conditions. Section 3 develops the test statistic and then obtains its asymptotic property. Section 4 presents Monte Carlo studies and an empirical example. Section 5 concludes the article by extending the proposed test to generalized linear models. All technical details are left to the Appendix.

2. MODEL STRUCTURE AND CONDITIONS

2.1. Models and Notations

Let (Y_i, X_i) be the observation collected from the ith subject, where $Y_i \in \mathbb{R}^1$ is the response variable and $X_i \in \mathbb{R}^p$ is the associated predictor for $1 \le i \le n$. We assume that X_i can be decomposed as $X_i = (X_{ia}^\top, X_{ib}^\top)^\top$ with $X_{ia} = (X_{i1}, \cdots, X_{iq})^\top \in \mathbb{R}^q$ and $X_{ib} = (X_{i(q+1)}, \cdots, X_{ip})^\top \in \mathbb{R}^{p-q}$, where q is smaller than the sample size n, and p is much larger than n. For the sake of simplicity, we also assume that $E(X_i) = 0$. To establish the relationship between Y_i and X_i, we consider the following standard linear regression model,

$$Y_i = X_i^\top \beta + \varepsilon_i = X_{ia}^\top \beta_a + X_{ib}^\top \beta_b + \varepsilon_i, \tag{1}$$

where $\beta = (\beta_a^\top, \beta_b^\top)^\top \in \mathbb{R}^p$, $\beta_a \in \mathbb{R}^q$, and $\beta_b \in \mathbb{R}^{p-q}$ are unknown regression coefficient vectors. In addition, we assume that ε_i in (1) is random noise with $E(\varepsilon_i) = 0$, $\text{var}(\varepsilon_i) = \sigma^2$, and $E(\varepsilon_i^4) = (3 + \Delta)\sigma^4$ for some finite constant $\Delta > -3$.

For the sake of convenience, let $\mathbb{Y} = (Y_1, \cdots, Y_n)^\top \in \mathbb{R}^n$ be the response vector, and let $\mathbb{X}_a = (X_{1a}, \cdots, X_{na})^\top \in \mathbb{R}^{n\times q}$ and $\mathbb{X}_b = (X_{1b}, \cdots, X_{nb})^\top \in \mathbb{R}^{n\times(p-q)}$ be the matrices associated with the X_{ia}'s and X_{ib}'s, respectively. Let $\mathbb{X} = (\mathbb{X}_a, \mathbb{X}_b) = (X_1, \cdots, X_n)^\top \in \mathbb{R}^{n\times p}$ be the matrix including all predictive variables and $\mathcal{E} = (\varepsilon_1, \cdots, \varepsilon_n)^\top \in \mathbb{R}^n$ be the noise vector. Then model (1) can be re-expressed as follows.

$$\mathbb{Y} = \mathbb{X}\beta + \mathcal{E} = \mathbb{X}_a\beta_a + \mathbb{X}_b\beta_b + \mathcal{E}. \tag{2}$$

In practice, $\mathbb{X}_a$ often contains a small set of relevant predictors via prior knowledge or preliminary analysis. In contrast, $\mathbb{X}_b$ collects a large number of predictors, whose statistical significance is still not clear and thus needs to be investigated. Accordingly, we consider the following statistical hypotheses,

$$H_0 : \beta_b = 0 \quad \text{vs.} \quad H_1 : \beta_b \neq 0. \tag{3}$$

When X_a is a null vector, equation (3) is equivalent to test $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$. It is noteworthy that, under H_0, model (2) reduces to

$$\mathbb{Y} = \mathbb{X}_a\beta_a + \mathcal{E}, \tag{4}$$

where we slightly abuse notation by using $\mathcal{E}$ to represent the random error vector in both the full and reduced models. In the rest of this paper, we will use it to stand for the random error in the reduced model only.

When $n > p$, one commonly uses the conventional partial F-test given below to test the null hypothesis in (3).

$$F = \frac{\mathbb{Y}^\top \widetilde{\mathbb{X}}_b (\widetilde{\mathbb{X}}_b^\top \widetilde{\mathbb{X}}_b)^{-1} \widetilde{\mathbb{X}}_b^\top \mathbb{Y}/(p-q)}{\mathbb{Y}^\top \{I_n - \mathbb{X}(\mathbb{X}^\top \mathbb{X})^{-1}\mathbb{X}^\top\}\mathbb{Y}/(n-p)},$$

where $I_n \in \mathbb{R}^{n\times n}$ stands for a $n \times n$ identity matrix, $\widetilde{\mathbb{X}}_b = (I_n - H_a)\mathbb{X}_b$, $H_a = \mathbb{X}_a(\mathbb{X}_a^\top \mathbb{X}_a)^{-1}\mathbb{X}_a^\top$. Under $n < p$, however, neither $\mathbb{X}^\top\mathbb{X}$ nor $\widetilde{\mathbb{X}}_b^\top \widetilde{\mathbb{X}}_b$ is invertible and hence this test is not applicable to high dimensional data. It is noteworthy that, under H_0, the contribution of $\mathbb{X}_b$ in explaining the variation of $\mathbb{Y}$ should be 0 after controlling for the effect of $\mathbb{X}_a$. As a result, we should have that $E\{(\mathbb{Y} - \mathbb{X}_a\beta_a)^\top \mathbb{X}_b\} = E\{\mathcal{E}^\top \mathbb{X}_b\} = 0$. This motivates the new testing procedure presented in this paper.

2.2. Boundedness and Moment Conditions

Before presenting the detailed procedure, we need to investigate a number of important and reasonable technical conditions. To this end, define $\Sigma_{b|a} = E\{\mathrm{cov}(X_{ib}|X_{ia})\} = (\sigma^*_{j_1 j_2}) \in \mathbb{R}^{(p-q)\times(p-q)}$. Without loss of generality, we also assume that $\sigma^*_{jj} = 1$ for any $j \in \mathcal{S} = \{q+1, \cdots, p\}$. Then, we introduce the following boundedness condition.

(C1) *Boundedness Condition.* Assume that there exist two positive constants $\tau_{\min}$ and $\tau_{\max}$ such that $\tau_{\min} < \lambda_{\min}(\Sigma_{b|a}) \le \lambda_{\max}(\Sigma_{b|a}) < \tau_{\max}$, where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ represent the smallest and largest eigenvalues of an arbitrary semi-positive definite matrix A, respectively.

Condition (C1) assures the model identifiability. Specifically, (C1) indicates that, conditional on X_{ia}, none of the predictors in X_{ib} (or $\mathcal{S}$) can be linearly represented by other predictors in $\mathcal{S}$. A similar condition has been widely used in the literature; see, for example, Fan et al. (2008), Zhang and Huang (2008), Wang (2009), and many others. However, Condition (C1) is typically insufficient for establishing the asymptotic normality of test statistics. Hence, we next introduce moment conditions on the conditional predictor,

$$X_{ib}^* = X_{ib} - BX_{ia}, \tag{5}$$

where $X_{ib}^* \in \mathbb{R}^{p-q}$ is the residual vector obtained by regressing X_{ib} on X_{ia} and $B = \text{cov}(X_{ib}, X_{ia})\text{cov}^{-1}(X_{ia}) \in \mathbb{R}^{(p-q)\times q}$. By our previous assumptions, we immediately have that $E(X_{ib}^*) = 0$ and $\text{cov}(X_{ib}^*) = \Sigma_{b|a}$. Define a collective set of X_{ib}^* as $\mathbb{X}_b^* = (X_{1b}^*, \cdots, X_{nb}^*)^\top \in \mathbb{R}^{n\times(p-q)}$. We then request the following moment conditions, which are driven by the diverging number of predictors (i.e., $p \to \infty$).

(C2) *Moment Conditions.* Assume that $q/p \to 0$, and, for any $1 \le i \le n$ and $1 \le i_1 \ne i_2 \le n$, the following moment conditions are satisfied.

(C2.a) $E(p^{-1}\sum_{j\in\mathcal{S}} X_{ij}^{*2} - 1)^4 = O(p^{-2})$,

(C2.b) $E(p^{-1}\sum_{j\in\mathcal{S}} X_{i_1j}^* X_{i_2j}^*)^4 = O(p^{-2})$.

By condition (C2.a) and Cauchy's inequality, we obtain that

$$\text{var}(p^{-1}\sum_{j\in\mathcal{S}} X_{ij}^{*2}) = p^{-2}\sum_{j_1,j_2\in\mathcal{S}}\{E(X_{ij_1}^{*2}X_{ij_2}^{*2}) - 1\} = O(1/p). \tag{6}$$

Furthermore, by condition (C2.b), we have that

$$\begin{aligned}
&p^2E(p^{-1}\sum_j X_{i_1j}^* X_{i_2j}^*)^4 \\
&= p^{-2}E\{\sum_{j_1,j_2,j_3,j_4}(X_{i_1j_1}^*X_{i_1j_2}^*X_{i_1j_3}^*X_{i_1j_4}^*)(X_{i_2j_1}^*X_{i_2j_2}^*X_{i_2j_3}^*X_{i_2j_4}^*)g\}
\end{aligned}$$

$$
\begin{aligned}
&= p^{-2}\{\sum_{j_1,j_2,j_3,j_4} E(X^*_{i_1j_1}X^*_{i_1j_2}X^*_{i_1j_3}X^*_{i_1j_4})E(X^*_{i_2j_1}X^*_{i_2j_2}X^*_{i_2j_3}X^*_{i_2j_4})\} \\
&= p^{-2}\sum_{j_1,j_2,j_3,j_4}\{E(X^*_{ij_1}X^*_{ij_2}X^*_{ij_3}X^*_{ij_4})\}^2 = O(1). \qquad (7)
\end{aligned}
$$

Both (6) and (7) will be used in technical appendices.

If the X^*_{ij}s are mutually independent for a fixed i, then both $p^{-1}\sum_j X^{*2}_{ij} - 1$ and $p^{-1}\sum_j X^*_{i_1j}X^*_{i_2j}$ are of the order $O_p(p^{-1/2})$. Accordingly, the fourth moment condition (C2) holds. In practice, however, we cannot expect the X^*_{ij}s to be independent of each other. Hence, two known assumptions, namely a multivariate normal distribution and a diverging factor model (Bai and Saranadasa, 1996) have often been considered in the literature. Under the boundedness condition (C1), we are able to demonstrate that both the multivariate normal distribution and the diverging factor model lead to (C2); see the following two propositions.

Proposition 1. *Assume that X^*_{ib} follows a multivariate normal distribution with mean 0 and covariance matrix $\Sigma_{b|a}$. In addition, assume that $\Sigma_{b|a}$ satisfies condition (C1). Then, condition (C2) must hold.*

The proof is given in Appendix B. We next consider the diverging factor model, which assumes that X^*_{ib} can be written as $X^*_{ib} = \Gamma Z_i$, where $\Gamma = (\gamma_{jk}) \in \mathbb{R}^{(p-q)\times m}$ for some $m \geq p-q$, $Z_i = (Z_{i1}, \cdots, Z_{im})^\top \in \mathbb{R}^m$, $E(Z_{ij}) = 0$, $\text{var}(Z_{ij}) = 1$, $E(Z^4_{ij}) = 3 + \Delta_z$ for some finite constant Δ_z, and $E(Z^8_{ij}) < \infty$. In addition, it is also required that $E(Z^{s_1}_{ij_1}Z^{s_2}_{ij_2}\cdots Z^{s_r}_{ij_r}) = E(Z^{s_1}_{ij_1})\cdots E(Z^{s_r}_{ij_r})$ for any integers $s_v \geq 0$ with $\sum_{v=1}^r s_v \leq 8$ and for different indices $j_1, j_2, \cdots, j_r \in \{1, 2, \cdots, m\}$.

Proposition 2. *Assume that X^*_{ib} follows a diverging factor model structure and its covariance matrix satisfies condition (C1). Then, (C2) holds.*

The proof is given in Appendix B. Propositions 1 and 2 indicate that the conditions

(C2.a) and (C2.b) are rather mild.

3. METHODOLOGY DEVELOPMENT

3.1. An Initial Test Statistic

After introducing the two regularity conditions (C1) and (C2), we propose a test statistic for testing the hypotheses (3). To this end, we first estimate β_a under H_0. Since we assume that the dimension of X_{ia} is low, the unknown regression coefficient can be estimated via the OLS approach. The resulting estimator is $\hat{\beta}_a = (n^{-1}\mathbb{X}_a^\top \mathbb{X}_a)^{-1}(n^{-1}\mathbb{X}_a^\top \mathbb{Y})$. Subsequently, the residual calculated through (4) is $\hat{\mathcal{E}} = \mathbb{Y} - \mathbb{X}_a\hat{\beta}_a$. If the sample size is large, one naturally expects that $n^{-1}\hat{\mathcal{E}}^\top \mathbb{X}_b \approx n^{-1}E\{\mathcal{E}^\top \mathbb{X}_b\} = 0$. This leads to the following test statistic

$$T_1 = n^{-1}\|\mathbb{X}_b^\top \hat{\mathcal{E}}\|^2 = n^{-1}\hat{\mathcal{E}}^\top \mathbb{X}_b \mathbb{X}_b^\top \hat{\mathcal{E}} = n^{-1}\sum_{j\in\mathcal{S}} \hat{\mathcal{E}}^\top \mathbb{X}_j \mathbb{X}_j^\top \hat{\mathcal{E}} = n^{-1}\sum_{j\in\mathcal{S}} \widetilde{\mathbb{Y}}^\top \widetilde{\mathbb{X}}_j \widetilde{\mathbb{X}}_j^\top \widetilde{\mathbb{Y}},$$

where $\mathbb{X}_j = (X_{1j}, X_{2j}, \cdots, X_{nj})^\top \in \mathbb{R}^n$, $\widetilde{\mathbb{X}}_j = (I_n - H_a)\mathbb{X}_j = (\widetilde{X}_{1j}, \cdots, \widetilde{X}_{nj})^\top \in \mathbb{R}^n$, and $\widetilde{\mathbb{Y}} = (I_n - H_a)\mathbb{Y}$. It is of note that $\widetilde{\mathbb{Y}}^\top \widetilde{\mathbb{X}}_j \widetilde{\mathbb{X}}_j^\top \widetilde{\mathbb{Y}}/n$ is the partial covariance of $\mathbb{Y}$ and $\mathbb{X}_j$, for $j \in \mathcal{S}$, after controlling for $\mathbb{X}_a$. Hence, T_1 is the sum of partial covariances across the explanatory variables being tested. In general, one naturally rejects the null hypothesis when T_1 is sufficiently large.

Theorem 1. *Assume the null hypothesis of (3), and conditions (C1) and (C2). If* $\min\{n, p\} \to \infty$, $qn^{-1/4} \to 0$, *and* $n/p \to 0$, *then we have*

$$p^{-1}\sigma^{-2}E_*(T_1) \to_p 1 \quad \text{and} \quad np^{-2}\sigma^{-4}var_*(T_1) \leq 2 + |\Delta| \tag{8}$$

with probability tending to 1, where $E_*(\cdot) = E(\cdot|\mathbb{X})$ *and* $var_*(\cdot) = var(\cdot|\mathbb{X})$.

The proof is given in Appendix C. By Theorem 1, we have that $E_*(T_1)/\text{var}_*^{1/2}(T_1) \to_p \infty$. This implies that, under the null hypothesis of (3) and conditional on $\mathbb{X}$, the normalized test statistic $T_1/\text{var}_*^{1/2}(T_1)$ cannot be asymptotically distributed as any non-degenerate distribution. As a result, we modify T_1 by using its conditionally bias-corrected estimator.

3.2. The Bias-Corrected Test Statistic

Under the null hypothesis in (3) and the fact that $(I_n - H_a)\mathbb{X}_b = (I_n - H_a)\mathbb{X}_b^*$, we know that the conditional bias of T_1 from (8) is $E_*(T_1) = \sigma^2(n^{-1}tr\{\mathbb{X}_b^*\mathbb{X}_b^{*\top}\} - n^{-1}tr\{\mathbb{X}_b^{*\top}H_a\mathbb{X}_b^*\}) = \sigma^2 n^{-1}(p-q)tr(\mathcal{M}\mathcal{Q})$, where $\mathcal{M} = (p-q)^{-1}\sum_{j\in\mathcal{S}}\widetilde{\mathbb{X}}_j\widetilde{\mathbb{X}}_j^\top$ and $\mathcal{Q} = I_n - H_a$. Conditional on $\mathbb{X}$, both quantities $\mathcal{M}$ and $\mathcal{Q}$ are known. This motivates us to correct the bias of T_1 by replacing σ^2 in $E_*(T_1)$ with its unbiased estimator, $\hat{\sigma}^2 = (n-q)^{-1}\hat{\mathcal{E}}^\top\hat{\mathcal{E}}$, which yields the following bias-corrected test statistic

$$T_2 = T_1 - \hat{\sigma}^2 n^{-1}(p-q)tr(\mathcal{M}\mathcal{Q}). \tag{9}$$

It can easily be seen that $E_*(T_2) = 0$, which immediately implies that $E(T_2) = 0$. To obtain a standardized test statistic, we next compute the variance of T_2 without conditioning on $\mathbb{X}$.

Theorem 2. *Under the same conditions and assumptions as those in Theorem 1, we have* $var(T_2) = 2\sigma^4 tr(\Sigma_{b|a}^2)\{1+o(1)\}$.

The proof is given in Appendix D. By Theorem 2, the asymptotic variance of T_2 is given by $2\sigma^4 tr(\Sigma_{b|a}^2)$. Accordingly, we can construct a test statistic,

$$Z = T_2/\{2\sigma^4 tr(\Sigma_{b|a}^2)\}^{1/2}, \tag{10}$$

whose asymptotic null distribution is given below.

Theorem 3. *Under the same conditions and assumptions as those in Theorem 1, we have that* $Z \to_d N(0,1)$.

The proof is given in Appendix E. This theorem allows us to test a subset of regression coefficients in high dimensional data. It is noteworthy that the numerator of Z is a bias-corrected term from T_1, and a larger Z tends to reject the null hypothesis. Accordingly, Theorem 3 indicates that, for a given significance level α, we reject the null hypothesis if $Z > z_{1-\alpha}$, where z_α stands for the αth quantile of a standard normal distribution. Based on the above theorem, one can calibrate the size of the proposed test by an usual standard normal distribution table.

Remark 1. To employ the proposed test statistic, we need to estimate the unknown quantity $2\sigma^4 tr(\Sigma_{b|a}^2)$. It seems natural to use $2\hat{\sigma}^4 tr(\hat{\Sigma}_{b|a}^2)$, where $\hat{\Sigma}_{b|a} = n^{-1}\widetilde{\mathbb{X}}_b^\top \widetilde{\mathbb{X}}_b$. However, as demonstrated by Srivastava (2005), $tr(\hat{\Sigma}_{b|a}^2)$ is not a consistent estimator of $tr(\Sigma_{b|a}^2)$ when $q = 0$; see their Remark 2.1 on page 253 as well as some relevant discussions in Chen and Qin (2010) and Chen et al. (2010). To this end, we adopt the approach of Srivastava (2005) and consider the following bias-corrected estimator

$$\widehat{tr(\Sigma_{b|a}^2)} = n^2(n+1-q)^{-1}(n-q)^{-1}\{tr(\hat{\Sigma}_{b|a}^2) - tr^2(\hat{\Sigma}_{b|a})/(n-q)\}.$$

Under the normality assumption with $q = 0$, Srivastava (2005) show that it is ratio consistent, i.e., $\widehat{tr(\Sigma_{b|a}^2)}/tr(\Sigma_{b|a}^2) \to_p 1$. For the case of non-normal data with $q > 0$, our simulation experiences indicate that this estimator also performs fairly well; see Examples 3.1 and 3.2 in the next section.

Remark 2. For the sake of simplicity, we assumed that $E(X_{ij}) = \alpha_j = 0$ for every j. In practice, this assumption may not be valid, i.e., $\alpha_j \neq 0$ for some j. To resolve this problem, we can simply include an intercept term in X_{ia} to redefine $X_{ia} := (1, X_{ia}^\top)^\top$. Accordingly, $(I_n - H_a)\mathbf{1} = 0$, where $\mathbf{1} = (1, 1, \cdots, 1)^\top \in \mathbb{R}^n$. This leads to $\widetilde{\mathbb{X}}_j = (I_n - H_a)\mathbb{X}_j = (I_n - H_a)\mathbb{X}_j^+$, where $\mathbb{X}_j^+ = (X_{1j} - \alpha_j, \cdots, X_{nj} - \alpha_j)^\top \in \mathbb{R}^n$ is the centralized predictor. Because our test statistic is based on $\widetilde{\mathbb{X}}_j$, it makes no difference to use $\mathbb{X}_j$ (non-centralized predictor) or $\mathbb{X}_j^+$ (centralized predictor), as long as the intercept is included in X_{ia}. Consequently, the asymptotic theory given in Theorem 3 (established for centralized design matrix) is still applicable, even though $E(X_{ij}) = \alpha_j \neq 0$. This conclusion is further confirmed by simulation studies; see Example 3.1 in the next section.

4. NUMERICAL STUDIES

4.1. Simulation Results

In this subsection, we present two simulation examples that evaluate the finite sample performance of the proposed test. The first example considers weakly correlated predictors (Tibshirani, 1996), while the second example studies the case in which a strong relationship exists between $\mathbb{X}_a$ and $\mathbb{X}_b$ (Fan et al., 2008).

Example 3.1. We generate the data from (2), where the regression coefficients β_j for $j \in \{1, 2, \cdots, q\}$ are simulated from a standard normal distribution, and then we set $\beta_j = 0$ for $j > q$. In addition, the predictor vector is given by $X_i = \Sigma^{1/2} Z_i^*$ for $i = 1 \cdots, n$, and each component of Z_i^* is independently generated from a s-tandard exponential distribution, exp(1). Moreover, the random error ε_i is independently generated from a standard normal distribution or a mixture distribution $0.1N(0, 3^2) + 0.9N(0, 1)$. We then consider $\text{cov}(X_i) = \Sigma = (\sigma_{j_1 j_2}) \in \mathbb{R}^{p \times p}$ with $\sigma_{j_1 j_2} = 0.5^{|j_1 - j_2|}$ (Tibshirani, 1996; Fan and Li, 2001). Hence, X_{ij_1} and X_{ij_2} are approximately uncorrelated when the difference $|j_1 - j_2|$ is sufficiently large. It is noteworthy that $E(X_i) \neq 0$, which violates our model assumption of $E(X_i) = 0$. According to Remark 2, we add an intercept term into X_{ia} to adjust for non-central predictors, and then redefine $X_{ia} := (1, X_{ia}^\top)^\top$.

We consider two different sample sizes (n=100, 200), three dimensions of predictors in the full model (p=200, 500, 1000), and three dimensions of predictors in the reduced model (q=0, 8, and 15). For each fixed parameter setting (i.e., n, p, and q), a total of 1,000 realizations are conducted with a nominal level of 5%. Table 1 presents the size of the bias-corrected test. For the sake of comparison, the test proposed by Zhong and Chen (2011) is also included, and we name it the ZC-test. A well-behaved test should have an empirical size around 0.05. Table 1 indicates that both methods perform quite well.

We next study the power of the bias-corrected test. To this end, we follow the settings of Zhong and Chen (2011) and consider two different types of alternative hypotheses. The first type is a non-sparse alternative, where $\beta_b = \kappa(\beta_{b1}, \beta_{b2}, \cdots, \beta_{b(p-q)})^\top \in \mathbb{R}^{p-q}$, the β_{bj}s are simulated from a standard normal distribution, and κ is selected so that the signal strength $\beta_b^\top \Sigma_{b|a} \beta_b$ ranges from 0 to 1.5. The second type is a sparse alternative, where β_{bj} $(1 \leq j \leq 5)$ are generated from a standard normal distribution with β_{bj} being set to be 0 for every $j > 5$. In addition, the signal strengths are the same as those of the first type. For the sake of illustration, we consider only the situation where the random error is normally distributed with $q = 0$, $n = 100$, and $p = 200$. Figure 1 depicts the empirical powers of the bias-corrected test and ZC-test, which indicates they steadily increase towards 100% as the signal strength gets larger. In sum, both tests perform satisfactorily and comparably in both sparse alternative and non-sparse alternative scenarios.

Example 3.2. We consider a case in which X_{ia} and X_{ib} are heavily correlated.

More specifically, we generate the data according to the factor model (5), where the variables X_{ij} with $1 \leq j \leq q$ and ε_i are randomly generated from the standard normal distribution. In addition,we have the variables $X_{ib} = BX_{ia} + X_{ib}^*$ for $1 \leq i \leq n$, where each element of the factor loading $B \in \mathbb{R}^{(p-q)\times q}$ is simulated from a standard normal distribution. Moreover, $X_{ib}^* \in \mathbb{R}^{p-q}$ is generated from a multivariate normal distribution with mean 0 and covariance matrix $\Sigma_{b|a} = (\sigma_{j_1 j_2}^*) \in \mathbb{R}^{(p-q)\times(p-q)}$ with $\sigma_{j_1 j_2}^* = 0.5^{|j_1 - j_2|}$. The regression coefficients, sample sizes, full model sizes, and reduced model sizes, as well as the number of realizations, are the same as those in Example 3.1. It can be verified that the technical conditions (C1) and (C2) imposed on $\Sigma_{b|a}$ (instead of Σ) are satisfied. Hence, the results presented in Table 2 indicate that the bias-corrected test performs reasonably well and is qualitatively similar to that in the previous example. Because the condition for the ZC-test is invalid under this simulation setting, it is not surprising that ZC-test does not perform well. Specifically, one can show that $tr(\Sigma^4) = tr(BB^\top)^4\{1+o(1)\} = tr(B^\top B)^4\{1+o(1)\} = p^4 tr(I_q)\{1+o_p(1)\} = qp^4\{1+o(1)\}$ and $tr(\Sigma^2) = qp^2\{1+o(1)\}$. As a result, $tr(\Sigma^4)/tr^2(\Sigma^2) \to q^{-1} \neq 0$ if q is fixed; this violates condition (2.8) in Zhong and Chen (2011). Finally, Figure 2 depicts the empirical power of the bias-corrected test. It is not surprising that the non-sparse alternative performs better than the sparse alternative, and their overall performances are qualitatively similar to those of Figure 1.

4.2. Real Data Analysis

To further demonstrate the practical usefulness of our proposed method, we consider an empirical example using data from an online mobile phone retailer. The data set can be obtained from the authors upon request, and will be made available for research purposes only. The data set contains a total of $n = 98$ daily records. The response is the revenue from the retailer's online sales, and the explanatory variables are the advertising spending on each of $p = 1,048$ different keywords that were bid for on Baidu (www.baidu.com), the leading domestic search engine in China. In practice, allocating the advertising spending on profitable keywords is critical for online sales. We therefore start by ranking $1,048$ keywords according to their relative importance measured by the coefficient of variation (CV) for each keyword. This is because a keyword with a weaker CV is typically associated with larger spending but smaller variability; empirical experience suggests that those keywords are more likely to be associated with online sales. As a result, a keyword with a weak CV is more important than one with a strong CV.

We next denote the sorted predictors as $V_{(1)}, V_{(2)}, \cdots, V_{(p)}$. Since sales vary with the day of the week, we introduce the 6-dimensional indicator variables $W \in \mathbb{R}^6$ to represent Sunday to Friday. For a fixed k, we define $X_a = (W, V_{(1)}, \cdots, V_{(k)})$, and then test whether the advertising spending on the rest of keywords, $X_b = (V_{(j)} : j > k)$, could provide a significant contribution to online sales by controlling for the effect of X_a. To this end, the bias-corrected test procedure is applied sequentially with $k = 1, 2, \cdots$, until the resulting p-value is larger than the 5% level of significance. The testing procedure stops with $k = 8$; this suggests that, after controlling for advertising spending on the first eight keywords, the others are not statistically significant to the response.

After carefully examining those eight keywords, we find that they can be classified into three different categories. The first category contains a single keyword, the brand name of this particular online retailer. People generally would not search for such a keyword if they were not already familiar with this retailer. Hence, identifying this keyword is a highly desirable result. The second category contains a keyword that is the name of a Chinese version of iPad ("one-person-one-book" directly translated from

Chinese). Since it is considered to be the most important competitor for iPad in the domestic Chinese market, targeting this keyword is also an expected result. The last category consists of six keywords that are related to mobile phones designed specifically for "senior people" (directly translated from Chinese). Since the percentage of seniors in China has increased steadily in the past few years as a result of the one child policy, it is not surprising that they play an important role in the mobile phone market.

Based on the experience of a field practitioner, the eight keywords identified by our bias-corrected test are highly interpretable and useful. In addition, offline data confirms that the product categories represented by those eight search keywords are economically important; they account for more than 65% of the entire online sales. It is worth noting that the simulation studies in Section 3.1 indicate that the bias-corrected test performs well when $n = 100$, $p = 1,000$, and $q = 8$; our empirical example is similar to this case. In sum, our test is able to identify 8 critical keywords from the 1,048 keywords and 98 observations; this method efficiently utilizes the high volume of data available to online retailers.

5. CONCLUDING REMARKS

To broaden the usefulness of our proposed test, we conclude this article by extending the test statistic to generalized linear models (McCullagh and Nelder, 1989). Consider

$$E(Y_i|X_i) = g^{-1}(X_i^\top \beta) = g^{-1}(X_{ia}^\top \beta_a + X_{ib}^\top \beta_b),$$

where $g(\cdot)$ is the canonical link function. The resulting log-likelihood function, after omitting the irrelevant constants, is given by $\sum_i \{(Y_i \cdot X_i^\top \beta) - nb(X_i^\top \beta)\}$ for some smooth function $b(\cdot)$. By maximizing the log-likelihood function under the null hypothesis of (3), we obtain the maximum likelihood estimator (MLE) of β_a, which is denoted $\hat{\beta}_a$. Then, with slight abuse of the notation for random errors and their corresponding residuals used in the linear model, we denote $\varepsilon_i = Y_i - g^{-1}(X_{ia}^\top \beta_a)$ and $\hat{\varepsilon}_i = Y_i - g^{-1}(X_{ia}^\top \hat{\beta}_a)$. As a result, the estimator of σ^2 is $\hat{\sigma}^2 = \hat{\mathcal{E}}^\top \hat{\mathcal{E}}/(n-q)$, where $\hat{\mathcal{E}} = (\hat{\varepsilon}_1, \cdots, \hat{\varepsilon}_n)^\top$.

Under the null hypothesis of (3), we have $E(\varepsilon_i \widetilde{X}_{ij}) = 0$ for any $j > q$, where $\widetilde{X}_{ij}$ is the i-th element of $\widetilde{\mathbb{X}}_j$ defined in Section 3. Similarly, an initial test statistic can be constructed as $T_1^g = n^{-1} \sum_{j>q} (\hat{\mathcal{E}}^\top \widetilde{\mathbb{X}}_j)^2 = n^{-1} \hat{\mathcal{E}}^\top \widetilde{\mathbb{X}}_b \widetilde{\mathbb{X}}_b^\top \hat{\mathcal{E}}$. Under the null hypothesis, we have $E_*(\hat{\mathcal{E}}^\top \widetilde{\mathbb{X}}_b \widetilde{\mathbb{X}}_b^\top \hat{\mathcal{E}}) \approx E_*(\mathcal{E}^\top \widetilde{\mathbb{X}}_b \widetilde{\mathbb{X}}_b^\top \mathcal{E}) = \sum_{i=1}^n \sigma_i^2 \omega_i$, where $\sigma_i^2 = E(\varepsilon_i^2)$ and ω_i is the ith diagonal element of $\widetilde{\mathbb{X}}_b \widetilde{\mathbb{X}}_b^\top \in \mathbb{R}^{n\times n}$. Accordingly, we can estimate $E_*(\mathcal{E}^\top \widetilde{\mathbb{X}}_b \widetilde{\mathbb{X}}_b^\top \mathcal{E})$ by $\hat{\mathcal{E}}^\top \Omega \hat{\mathcal{E}}$, where $\Omega = \mathrm{diag}\{\omega_1, \cdots, \omega_n\}$. This leads us to propose the following bias-corrected test statistic $T_2^g = n^{-1} \hat{\mathcal{E}}^\top (\widetilde{\mathbb{X}}_b \widetilde{\mathbb{X}}_b^\top - \Omega) \hat{\mathcal{E}}$. Then, employing similar techniques as in the linear model, we are able to show that $\mathrm{var}(T_2^g) = 2\bar{\sigma}^4 tr(\Sigma_{b|a}^2)\{1 + o(1)\}$, where $\bar{\sigma}^2 = n^{-1} \sum_{i=1}^n \sigma_i^2$. Accordingly, we obtain a test statistic, $Z_g = T_2^g / \{2\hat{\sigma}^4 \widehat{tr(\Sigma_{b|a}^2)}\}^{1/2}$, for testing the null hypothesis of (3) in generalized linear models. Our unreported numerical results suggest that the test statistic Z_g works fairly well in terms of both size and power.

To conclude the article, we discuss two interesting topics for future research. The first is to obtain a test statistic for testing the null hypothesis $H_0 : \beta_a = 0$. This is a challenging task since the total number of unknown parameters in β_b remains large even under H_0. The second is to employ the Pearson residual or deviance residual, as proposed by an anonymous referee, to derive a test statistic for generalized linear models. We believe these efforts would strengthen the use of hypothesis testing for making inferences in high dimensional data analysis.

APPENDIX

Appendix A. Technical Lemmas

Before proving four theorems, we present the following three lemmas. Lemma 1 can be found in Bendat and Piersol (1966) and Lemma 2 can be derived directly from Bao and Ullah (2010). Accordingly, we only provide the details for Lemma 3.

Lemma 1. *Let* $(U_1, U_2, U_3, U_4)^\top \in \mathbb{R}^4$ *be a 4-dimensional normal random vector with* $E(U_j) = 0$ *and* $var(U_j) = 1$ *for* $1 \le j \le 4$*. We then have* $E(U_1U_2U_3U_4) = \delta_{12}\delta_{34} + \delta_{13}\delta_{24} + \delta_{14}\delta_{23}$*, where* $\delta_{ij} = E(U_iU_j)$*.*

Lemma 2. *Let* $V = (V_1, \cdots, V_m)^\top \in \mathbb{R}^m$ *be a random vector with* $E(V) = 0$ *and* $cov(V) = I_m$*. We further assume that* $E(V_{i_1}^{g_1} V_{i_2}^{g_2} \cdots V_{i_r}^{g_r}) = E(V_{i_1}^{g_1}) \cdots E(V_{i_r}^{g_r})$ *for indices* $i_1, i_2, \cdots, i_r \in \{1, 2, \cdots, m\}$ *and for any integers* $g_v \ge 0$ *with* $\sum_{v=1}^r g_v \le 8$*. Then, for any symmetric* $m \times m$ *matrix* A_1 *and any* $m \times m$ *positive definite matrix* A_2*, we have that (i.)* $E(V^\top A_1 V)^2 = tr^2(A_1) + 2tr(A_1^2) + \bar{\Delta} tr(A_1^{\otimes 2})$*, where* $A_1 = (a_{j_1j_2})$*,* $A_1^{\otimes 2} = (a_{j_1j_2}^2)$*, and* $\bar{\Delta} = E(V_i^4) - 3$*; (ii.) there is a finite constant* C *such that* $E\{V^\top A_2 V - tr(A_2)\}^4 \le C tr^2(A_2^2)$*.*

Lemma 3. *Assume that* $W_i = \tilde{\Gamma}\tilde{Z}_i \in \mathbb{R}^l$*, where* $\tilde{\Gamma} = (\tilde{\gamma}_{jk}) \in \mathbb{R}^{l\times m}$*,* $\tilde{Z}_i \in \mathbb{R}^m$*,* $E(\tilde{Z}_i) = 0$*,* $cov(\tilde{Z}_i) = I_m$*, and* $E(\tilde{Z}_{ij}^8) < \infty$ *for* $j = 1, \cdots, m$*. In addition, assume that* $E(\tilde{Z}_{ij_1}^{l_1} \tilde{Z}_{ij_2}^{l_2} \cdots \tilde{Z}_{ij_r}^{l_r}) = E(\tilde{Z}_{ij_1}^{l_1}) \cdots E(\tilde{Z}_{ij_r}^{l_r})$ *for indices* $j_1, j_2, \cdots, j_r \in \{1, 2, \cdots, m\}$ *and for any integers* $l_v \ge 0$ *with* $\sum_{v=1}^r l_v \le 8$*. We then have that* $E(W_{ij_1}W_{ij_2}W_{ij_3}W_{ij_4}) = \tilde{\sigma}_{j_1j_2}\tilde{\sigma}_{j_3j_4} + \tilde{\sigma}_{j_1j_3}\tilde{\sigma}_{j_2j_4} + \tilde{\sigma}_{j_1j_4}\tilde{\sigma}_{j_2j_3} + \tilde{\Delta}\sum_{k=1}^m \tilde{\gamma}_{j_1k}\tilde{\gamma}_{j_2k}\tilde{\gamma}_{j_3k}\tilde{\gamma}_{j_4k}$*, where* $cov(W_i) = (\tilde{\sigma}_{j_1j_2}) \in \mathbb{R}^{l\times l}$ *and* $\tilde{\Delta} = E(\tilde{Z}_{ij}^4) - 3$*.*

Proof. From $W_i = \tilde{\Gamma}\tilde{Z}_i$, we have $W_{ij} = \sum_{k=1}^m \tilde{\gamma}_{jk}\tilde{Z}_{ik}$ for $1 \le j \le l$. As a result, we obtain that

$$\begin{aligned}
E(W_{ij_1}W_{ij_2}W_{ij_3}W_{ij_4}) &= \sum_{k_1,k_2,k_3,k_4} \tilde{\gamma}_{j_1k_1}\tilde{\gamma}_{j_2k_2}\tilde{\gamma}_{j_3k_3}\tilde{\gamma}_{j_4k_4}E(\tilde{Z}_{ik_1}\tilde{Z}_{ik_2}\tilde{Z}_{ik_3}\tilde{Z}_{ik_4}) \\
&= E(\tilde{Z}_{ik}^4)\sum_k \tilde{\gamma}_{j_1k}\tilde{\gamma}_{j_2k}\tilde{\gamma}_{j_3k}\tilde{\gamma}_{j_4k} + \sum_{k_1\neq k_2} \tilde{\gamma}_{j_1k_1}\tilde{\gamma}_{j_2k_1}\tilde{\gamma}_{j_3k_2}\tilde{\gamma}_{j_4k_2} \\
&\quad + \sum_{k_1\neq k_2} \tilde{\gamma}_{j_1k_1}\tilde{\gamma}_{j_2k_2}\tilde{\gamma}_{j_3k_1}\tilde{\gamma}_{j_4k_2} + \sum_{k_1\neq k_2} \tilde{\gamma}_{j_1k_1}\tilde{\gamma}_{j_2k_2}\tilde{\gamma}_{j_3k_2}\tilde{\gamma}_{j_4k_1} \\
&= \{E(\tilde{Z}_{ik}^4)-3\}\sum_k \tilde{\gamma}_{j_1k}\tilde{\gamma}_{j_2k}\tilde{\gamma}_{j_3k}\tilde{\gamma}_{j_4k} + \sum_{k_1,k_2} \tilde{\gamma}_{j_1k_1}\tilde{\gamma}_{j_2k_1}\tilde{\gamma}_{j_3k_2}\tilde{\gamma}_{j_4k_2} \\
&\quad + \sum_{k_1,k_2} \tilde{\gamma}_{j_1k_1}\tilde{\gamma}_{j_2k_2}\tilde{\gamma}_{j_3k_1}\tilde{\gamma}_{j_4k_2} + \sum_{k_1,k_2} \tilde{\gamma}_{j_1k_1}\tilde{\gamma}_{j_2k_2}\tilde{\gamma}_{j_3k_2}\tilde{\gamma}_{j_4k_1} \\
&= \tilde{\sigma}_{12}\tilde{\sigma}_{34} + \tilde{\sigma}_{13}\tilde{\sigma}_{24} + \tilde{\sigma}_{14}\tilde{\sigma}_{23} + \tilde{\Delta}\sum_{k=1}^{m} \tilde{\gamma}_{j_1k}\tilde{\gamma}_{j_2k}\tilde{\gamma}_{j_3k}\tilde{\gamma}_{j_4k}. \qquad (11)
\end{aligned}$$

The last equality is due to the fact that $\tilde{\sigma}_{j_1j_2} = \sum_k \tilde{\gamma}_{j_1k}\tilde{\gamma}_{j_2k}$ for any $1 \leq j_1, j_2 \leq l$. This completes the proof. □

Appendix B. Proof of Proposition 1-2

The normal distribution assumed in Proposition 1 implies the diverging factor model in Proposition 2. Hence, we only present proofs for Proposition 2, where (C2.a) and (C2.b) are satisfied.

Proof of (C2.a). The moment condition (C2.a) can be obtained directly from Lemma 2(ii); we thus omitted it.

Proof of (C2.b). Using Lemma 3, we have that $E\{X^*_{ij_1}X^*_{ij_2}X^*_{ij_3}X^*_{ij_4}\} = \sigma^*_{j_1j_2}\sigma^*_{j_3j_4} + \sigma^*_{j_1j_3}\sigma^*_{j_2j_4} + \sigma^*_{j_1j_4}\sigma^*_{j_2j_3} + \Delta_z\sum_{k=1}^m \gamma_{j_1k}\gamma_{j_2k}\gamma_{j_3k}\gamma_{j_4k}$. This, together with Cauchy's inequality, condition (C1), and $\Sigma_{b|a} = \Gamma\Gamma^\top$, implies that

$$\begin{aligned}
& p^{-2} \sum_{j_1,j_2,j_3,j_4} \{E(X^*_{ij_1} X^*_{ij_2} X^*_{ij_3} X^*_{ij_4})\}^2 \\
= \ & p^{-2} \sum_{j_1,j_2,j_3,j_4} (\sigma^*_{j_1j_2}\sigma^*_{j_3j_4} + \sigma^*_{j_1j_3}\sigma^*_{j_2j_4} + \sigma^*_{j_1j_4}\sigma^*_{j_2j_3} + \Delta_z \sum_k \gamma_{j_1k}\gamma_{j_2k}\gamma_{j_3k}\gamma_{j_4k})^2 \\
\leq \ & 2p^{-2} \sum_{j_1,j_2,j_3,j_4} (\sigma^*_{j_1j_2}\sigma^*_{j_3j_4} + \sigma^*_{j_1j_3}\sigma^*_{j_2j_4} + \sigma^*_{j_1j_4}\sigma^*_{j_2j_3})^2 \\
& +2p^{-2}\Delta_z^2 \sum_{j_1,j_2,j_3,j_4} (\sum_k \gamma_{j_1k}\gamma_{j_2k}\gamma_{j_3k}\gamma_{j_4k})^2 \\
= \ & 2p^{-2}\{3tr^2(\Sigma^2_{b|a}) + 6tr(\Sigma^4_{b|a})\} + 2p^{-2}\Delta_z^2 \sum_{j_1,j_2,j_3,j_4} (\sum_k \gamma_{j_1k}\gamma_{j_2k}\gamma_{j_3k}\gamma_{j_4k})^2 \\
= \ & O(1) + 2p^{-2}\Delta_z^2 \sum_{j_1,j_2,j_3,j_4} \{\sum_{k_1,k_2} (\gamma_{j_1k_1}\gamma_{j_2k_1}\gamma_{j_3k_1}\gamma_{j_4k_1})(\gamma_{j_1k_2}\gamma_{j_2k_2}\gamma_{j_3k_2}\gamma_{j_4k_2})\} \\
= \ & O(1) + 2p^{-2}\Delta_z^2 \sum_{k_1,k_2}\{ \sum_{j_1,j_2,j_3,j_4} (\gamma_{j_1k_1}\gamma_{j_1k_2})(\gamma_{j_2k_1}\gamma_{j_2k_2})(\gamma_{j_3k_1}\gamma_{j_3k_2})(\gamma_{j_4k_1}\gamma_{j_4k_2})\} \\
= \ & O(1) + 2p^{-2}\Delta_z^2 \sum_{k_1,k_2} (\sum_j \gamma_{jk_1}\gamma_{jk_2})^4 \leq O(1) + 2p^{-2}\Delta_z^2\{\sum_{k_1,k_2}(\sum_j \gamma_{jk_1}\gamma_{jk_2})^2\}^2 \\
\leq \ & O(1) + 2p^{-2}\Delta_z^2 tr^2(\Gamma^\top\Gamma)^2 = O(1) + 2p^{-2}\Delta_z^2 tr^2(\Gamma\Gamma^\top)^2 \\
= \ & O(1) + 2p^{-2}\Delta_z^2 tr^2(\Sigma^2_{b|a}) = O(1).
\end{aligned}$$

From (7), we complete the proof. □

Appendix C. Proof of Theorem 1

To prove the theorem, we consider two steps; Step (1) shows the first part in (8) and Step (2) demonstrates the second part.

Step (1). Under the null hypothesis of (3), we have that $\hat{\mathcal{E}} = (I_n - H_a)\mathcal{E}$ and

$$\begin{aligned}
p^{-1}\sigma^{-2}E_*(T_1) &= n^{-1}p^{-1}tr\{(I_n - H_a)\mathbb{X}_b\mathbb{X}_b^\top(I_n - H_a)\} \\
&= n^{-1}p^{-1}tr\{(I_n - H_a)\mathbb{X}^*_b\mathbb{X}^{*\top}_b(I_n - H_a)\} \\
&= n^{-1}p^{-1}tr(\mathbb{X}^*_b\mathbb{X}^{*\top}_b) - n^{-1}p^{-1}tr(\mathbb{X}^{*\top}_b H_a\mathbb{X}^*_b). \qquad (12)
\end{aligned}$$

Define $\mathcal{T} = n^{-1}p^{-1}tr(\mathbb{X}_b^*\mathbb{X}_b^{*\top}) = n^{-1}\sum_{i=1}^n(p^{-1}\sum_{j\in\mathcal{S}}X_{ij}^{*2})$. It is obvious that $E(\mathcal{T}) = 1 + o(1)$. Applying (6) and condition (C2.a), we further show that

$$\text{var}(\mathcal{T}) = n^{-1}\text{var}(p^{-1}\sum_{j\in\mathcal{S}}X_{ij}^{*2}) = o(1).$$

Accordingly, the first term of (12) is $n^{-1}p^{-1}tr\{\mathbb{X}_b^*\mathbb{X}_b^{*\top}\} = 1 + o_p(1)$. We next demonstrate that the order of the second term in (A.2) is $o_p(1)$. Using $tr(H_a) = q$, then

$$n^{-1}p^{-1}tr\{\mathbb{X}_b^{*\top}H_a\mathbb{X}_b^*\} \leq n^{-1}p^{-1}tr(H_a)\lambda_{\max}(\mathbb{X}_b^*\mathbb{X}_b^{*\top}) = \{n^{-1}q\}\lambda_{\max}(p^{-1}\mathbb{X}_b^*\mathbb{X}_b^{*\top}). \quad (13)$$

Define $\bar{\mathcal{H}} = \mathbb{X}_b^*\mathbb{X}_b^{*\top} - (p-q)I_n = (\bar{h}_{i_1i_2}) \in \mathbb{R}^{n\times n}$. By Chebyshev's inequality, for any arbitrarily large constant t, we obtain that

$$P(\lambda_{\max}(\bar{\mathcal{H}}) > n^{3/4}p^{1/2}t) \leq n^{-3}p^{-2}t^{-4}E\{\lambda_{\max}^4(\bar{\mathcal{H}})\} \leq n^{-3}p^{-2}t^{-4}E\{tr\bar{\mathcal{H}}^4\}. \quad (14)$$

It is noteworthy that $tr(\bar{\mathcal{H}}^4) = \sum_{i_1,i_2,i_3,i_4}\bar{h}_{i_1i_2}\bar{h}_{i_2i_3}\bar{h}_{i_3i_4}\bar{h}_{i_4i_1}$. Hence,

$$\begin{aligned} E\{tr(\bar{\mathcal{H}}^4)\} &= E\{\sum_{i_1,i_2,i_3,i_4}\bar{h}_{i_1i_2}\bar{h}_{i_2i_3}\bar{h}_{i_3i_4}\bar{h}_{i_4i_1}\} \\ &= E\{\sum_{\mathcal{A}}\bar{h}_{i_1i_2}\bar{h}_{i_2i_3}\bar{h}_{i_3i_4}\bar{h}_{i_4i_1}\} + E\{\sum_{\mathcal{A}^c}\bar{h}_{i_1i_2}\bar{h}_{i_2i_3}\bar{h}_{i_3i_4}\bar{h}_{i_4i_1}\}, \quad (15) \end{aligned}$$

where $\mathcal{A} = \{(i_1, i_2, i_3, i_4), i_1 \neq i_2 \neq i_3 \neq i_4\}$. After algebraic simplification with condition (C1), the first term of (15) becomes

$$
\begin{aligned}
& E\{\sum_{\mathcal{A}} \bar{h}_{i_1i_2}\bar{h}_{i_2i_3}\bar{h}_{i_3i_4}\bar{h}_{i_4i_1}\} \\
&= E\{\sum_{i_1\neq i_2\neq i_3\neq i_4}\sum_{j_1\in\mathcal{S}} X^*_{i_1j_1}X^*_{i_2j_1}\sum_{j_2\in\mathcal{S}} X^*_{i_2j_2}X^*_{i_3j_2}\sum_{j_3\in\mathcal{S}} X^*_{i_3j_3}X^*_{i_4j_3}\sum_{j_4\in\mathcal{S}} X^*_{i_1j_4}X^*_{i_4j_4}\} \\
&= \sum_{i_1\neq i_2\neq i_3\neq i_4}\sum_{j_1,j_2,j_3,j_4\in\mathcal{S}} E\{X^*_{i_1j_1}X^*_{i_1j_4}X^*_{i_2j_1}X^*_{i_2j_2}X^*_{i_3j_2}X^*_{i_3j_3}X^*_{i_4j_3}X^*_{i_4j_4}\} \\
&= \sum_{i_1\neq i_2\neq i_3\neq i_4}\sum_{j_1,j_2,j_3,j_4\in\mathcal{S}} \sigma^*_{j_1j_4}\sigma^*_{j_1j_2}\sigma^*_{j_2j_3}\sigma^*_{j_3j_4} \\
&= n(n-1)(n-2)(n-3)tr(\Sigma^4_{b|a}) = O(n^4p). \qquad (16)
\end{aligned}
$$

Using the fact that $|\mathcal{A}^c| \le n^3$ and Cauchy's inequality, the second term of (15) satisfies

$$
E\{\sum_{A^c} \bar{h}_{i_1i_2}\bar{h}_{i_2i_3}\bar{h}_{i_3i_4}\bar{h}_{i_4i_1}\} \le 4^{-1}n^3\{E(\bar{h}^4_{i_1i_2}) + E(\bar{h}^4_{i_2i_3}) + E(\bar{h}^4_{i_3i_4}) + E(\bar{h}^4_{i_4i_1})\}. \qquad (17)
$$

To further simplify (17), we consider two cases, $i_1 \neq i_2$ and $i_1 = i_2$. When $i_1 \neq i_2$, we employ condition (C2.b) and equation (7), and then obtain that

$$
\begin{aligned}
E(\bar{h}^4_{i_1i_2}) &= \sum_{j_1,j_2,j_3,j_4\in\mathcal{S}} E\{X^*_{i_1j_1}X^*_{i_1j_2}X^*_{i_1j_3}X^*_{i_1j_4}X^*_{i_2j_1}X^*_{i_2j_2}X^*_{i_2j_3}X^*_{i_2j_4}\} \\
&= \sum_{j_1,j_2,j_3,j_4\in\mathcal{S}} \{E(X^*_{i_1j_1}X^*_{i_1j_2}X^*_{i_1j_3}X^*_{i_1j_4})\}^2 = O(p^2).
\end{aligned}
$$

For the case with $i_1 = i_2$, the same conclusion can be established via condition (C2.a). The above results, together with (16) and (17), imply that $E\{tr(\bar{\mathcal{H}}^4)\} = O(n^4p) + O(n^3p^2)$. Hence, as long as n/p tends to 0 and t is sufficiently large, we have that $n^{-3}p^{-2}t^{-4}E\{tr\bar{\mathcal{H}}^4\} \to 0$. This result and (14) lead to

$$
\lambda_{\max}(\bar{\mathcal{H}}) = O_p(p^{1/2}n^{3/4}). \qquad (18)
$$

Using (13), (18), and the assumption of $qn^{-1/4} \to 0$, we thus obtain that, with probability tending to 1, the second term of (12) is $n^{-1}p^{-1}tr\{\mathbb{X}_b^{*\top} H_a \mathbb{X}_b^*\} \le \{n^{-1}q\}\{1 + O_p(n^{3/4}p^{-1/2})\} = o_p(1)$. Consequently, $p^{-1}\sigma^{-2}E_*(T_1) = 1 + o_p(1)$, which completes the proof of the first part in (8).

STEP (2). We next consider the conditional variance $\text{var}_*(T_1)$. After algebraic simplification and employing Lemma 2(i), we have that

$$
\begin{aligned}
np^{-2}\text{var}_*(T_1) &= n^{-1}p^{-2}[E_*\{\mathcal{E}^\top(I_n - H_a)\mathbb{X}_b^*\mathbb{X}_b^{*\top}(I_n - H_a)\mathcal{E}\}^2 \\
&\quad -\{E_*\{\mathcal{E}^\top(I_n - H_a)\mathbb{X}_b^*\mathbb{X}_b^{*\top}(I_n - H_a)\mathcal{E}\}\}^2] \\
&= 2n^{-1}p^{-2}\sigma^4 tr\{(I_n - H_a)\mathbb{X}_b^*\mathbb{X}_b^{*\top}(I_n - H_a)\mathbb{X}_b^*\mathbb{X}_b^{*\top}(I_n - H_a)\} \\
&\quad +\Delta n^{-1}p^{-2}\sigma^4 tr\{(I_n - H_a)\mathbb{X}_b^*\mathbb{X}_b^{*\top}(I_n - H_a)\}^{\otimes 2} \\
&\le 2n^{-1}p^{-2}\sigma^4 tr\{(I_n - H_a)\mathbb{X}_b^*\mathbb{X}_b^{*\top}(I_n - H_a)\mathbb{X}_b^*\mathbb{X}_b^{*\top}(I_n - H_a)\} \\
&\quad +|\Delta| n^{-1}p^{-2}\sigma^4 tr\{(I_n - H_a)\mathbb{X}_b^*\mathbb{X}_b^{*\top}(I_n - H_a)\}^2, \\
&= (2+\Delta)n^{-1}p^{-2}\sigma^4 tr\{(I_n - H_a)\mathbb{X}_b^*\mathbb{X}_b^{*\top}(I_n - H_a)\mathbb{X}_b^*\mathbb{X}_b^{*\top}(I_n - H_a)\} \\
&\le (2+|\Delta|)n^{-1}p^{-2}\sigma^4 tr\{\mathbb{X}_b^*\mathbb{X}_b^{*\top}(I_n - H_a)\mathbb{X}_b^*\mathbb{X}_b^{*\top}\} \\
&\le (2+|\Delta|)n^{-1}p^{-2}\sigma^4 tr\{(\mathbb{X}_b^*\mathbb{X}_b^{*\top})^2\} = (2+|\Delta|)n^{-1}\sigma^4 tr\{p^{-1}(\mathbb{X}_b^*\mathbb{X}_b^{*\top})^2\}.
\end{aligned}
$$

Using similar techniques to those used above, we can verify that $n^{-1}p^{-2}tr\{(\mathbb{X}_b^*\mathbb{X}_b^{*\top})^2\} \to_p 1$. As a result, $np^{-2}\sigma^{-4}\text{var}_*(T_1) \le 2(1+\Delta)$ in probability. This completes the entire proof of Theorem 1. □

Appendix D. Proof of Theorem 2

To prove this theorem, we introduce the statistic, $\tilde{T}_2 = n^{-1}\bar{T}_1 - (p-q)n^{-1}\mathcal{E}^\top\mathcal{E}$, where $\bar{T}_1 = \mathcal{E}^\top\mathbb{X}_b^*\mathbb{X}_b^{*\top}\mathcal{E}$. We then consider two steps, namely computing the variance of $\tilde{T}_2$ and showing that the difference between $\tilde{T}_2$ and T_2 is negligible.

Step (1). By condition (C1), we have $\mathrm{var}(\mathbb{X}_j^*) = \sigma_{jj}^* = 1$. Under the null hypothesis of (3), one can easily verify that $E(\tilde{T}_2) = 0$. With algebraic simplification, we obtain that $n^2\mathrm{var}(\tilde{T}_2) = E(\bar{T}_1^2) + (p-q)^2 E\{\mathcal{E}^\top\mathcal{E}\}^2 - 2(p-q)E\{\bar{T}_1\mathcal{E}^\top\mathcal{E}\}$. We next evaluate the three terms on the right-hand side of this equation. Since $E(\varepsilon_i^4) = (3+\Delta)\sigma^4$,

$$\begin{aligned}
E(\bar{T}_1^2) &= E(\mathcal{E}^\top\mathbb{X}_b^*\mathbb{X}_b^{*\top}\mathcal{E})^2 = Eg\{\sum_{j\in\mathcal{S}}(\sum_i X_{ij}^*\varepsilon_i)^2 g\}^2 = E(\sum_{j\in\mathcal{S}}\sum_{i_1,i_2} X_{i_1j}^*X_{i_2j}^*\varepsilon_{i_1}\varepsilon_{i_2})^2 \\
&= \sum_{j_1,j_2\in\mathcal{S}}\sum_{i_1,i_2,i_3,i_4} E\{\varepsilon_{i_1}\varepsilon_{i_2}\varepsilon_{i_3}\varepsilon_{i_4}X_{i_1j_1}^*X_{i_2j_1}^*X_{i_3j_2}^*X_{i_4j_2}^*\} \\
&= \sum_{j_1,j_2\in\mathcal{S}}\sum_{i=1}^{n} E\{\varepsilon_i^4 X_{ij_1}^{*2}X_{ij_2}^{*2}\} + \sum_{j_1,j_2\in\mathcal{S}}\sum_{i_1\neq i_2} E\{\varepsilon_{i_1}^2\varepsilon_{i_2}^2X_{i_1j_1}^{*2}X_{i_2j_2}^{*2}\} \\
&\quad +2\sum_{j_1,j_2\in\mathcal{S}}\sum_{i_1\neq i_2} E\{\varepsilon_{i_1}^2\varepsilon_{i_2}^2X_{i_1j_1}^*X_{i_1j_2}^*X_{i_2j_1}^*X_{i_2j_2}^*\} \\
&= (3+\Delta)\sigma^4 n\sum_{j_1,j_2\in\mathcal{S}} E\{X_{ij_1}^{*2}X_{ij_2}^{*2}\} \\
&\quad +\sigma^4 n(n-1)(p-q)^2 + 2\sigma^4 n(n-1)tr(\Sigma_{b|a}^2). \qquad (19)
\end{aligned}$$

By the definition of Δ, one can demonstrate that $\sigma^{-4}E(\mathcal{E}^\top\mathcal{E})^2 = n(n+2)+\Delta n$, and then show that

$$\begin{aligned}
\sigma^{-4}E\{\bar{T}_1\mathcal{E}^\top\mathcal{E}\} &= \sigma^{-4}E\{\mathcal{E}^\top\mathbb{X}_b^*\mathbb{X}_b^{*\top}\mathcal{E}\mathcal{E}^\top\mathcal{E}\} = \sigma^{-4}(p-q)E\{\mathcal{E}^\top\mathcal{E}\}^2 \\
&= (p-q)\{n(n+2)+n\Delta\}.
\end{aligned}$$

The above results, together with condition (C2.a) and equation (6), imply that

$$\begin{aligned}
n^2\sigma^{-4}\mathrm{var}(\tilde{T}_2) &= n(3+\Delta)\sum_{j_1,j_2\in\mathcal{S}}\{E(X_{ij_1}^{*2}X_{ij_2}^{*2})-1\} + 2n(n-1)tr(\Sigma_{b|a}^2) \\
&= O(np) + 2n(n-1)tr(\Sigma_{b|a}^2) = 2n^2 tr(\Sigma_{b|a}^2)(1+o(1)), \qquad (20)
\end{aligned}$$

which completes the proof of Step (1).

Step (2). After simple calculation, we have that $\text{var}(\tilde{T}_2) \geq 2\sigma^4(p-q)\tau_{\min}^2\{1+o(1)\}$, which is of order $O(p)$ by condition (C1) and the assumption of $q/p \to 0$. As a result, it suffices to show that $p^{-1/2}(T_2 - \tilde{T}_2) = o_p(1)$ to complete the proof. Note that

$$p^{-1/2}(T_2 - \tilde{T}_2) = p^{-1/2}\{T_1 - n^{-1}\bar{T}_1 + n^{-1}(p-q)\mathcal{E}^\top H_a \mathcal{E}\}$$

$$-p^{-1/2}\{\hat{\sigma}^2 n^{-1}(p-q)tr(\tilde{\mathcal{M}}\mathcal{Q}) - n^{-1}(p-q)\mathcal{E}^\top \mathcal{Q}\mathcal{E}\}, \tag{21}$$

where $\tilde{\mathcal{M}} = (p-q)^{-1}\mathbb{X}_b^* \mathbb{X}_b^{*\top} \in \mathbb{R}^{n\times n}$. We next demonstrate that the two terms on the right-hand side of the above equation are of order $o_p(1)$. By $\hat{\mathcal{E}} = (I_n - H_a)\mathcal{E}$, we have

$$\begin{aligned} & n\{T_1 - n^{-1}\bar{T}_1 + n^{-1}(p-q)\mathcal{E}^\top H_a \mathcal{E}\} \\ = \;& \mathcal{E}^\top (I_n - H_a)\mathbb{X}_b^* \mathbb{X}_b^{*\top}(I_n - H_a)\mathcal{E} - \mathcal{E}^\top \mathbb{X}_b^* \mathbb{X}_b^{*\top}\mathcal{E} + (p-q)\mathcal{E}^\top H_a \mathcal{E} \\ = \;& \mathcal{E}^\top H_a \bar{\mathcal{H}} H_a \mathcal{E} - 2\mathcal{E}^\top H_a \bar{\mathcal{H}}\mathcal{E}. \end{aligned} \tag{22}$$

The first term in (22) satisfies $\mathcal{E}^\top H_a \bar{\mathcal{H}} H_a \mathcal{E} \leq \lambda_{\max}(\bar{\mathcal{H}})\mathcal{E}^\top H_a \mathcal{E}$. In addition, (18) indicates that $\lambda_{\max}(\bar{\mathcal{H}}) = O_p(p^{1/2}n^{3/4})$. By $tr(H_a^{\otimes 2}) \leq tr(H_a^2) = q$, we have that $\sigma^{-4}\text{var}\{n^{-1/4}\mathcal{E}^\top H_a \mathcal{E}\} = n^{-1/2}\{2q + \Delta tr(H_a^{\otimes 2})\} \leq n^{-1/2}q\{2+\Delta\} = o(1)$. This, in conjunction with the fact that $\sigma^{-2}E\{n^{-1/4}\mathcal{E}^\top H_a \mathcal{E}\} = n^{-1/4}q = o(1)$, yields $\mathcal{E}^\top H_a \mathcal{E} = o_p(n^{1/4})$. Hence, $\mathcal{E}^\top H_a \bar{\mathcal{H}} H_a \mathcal{E} = o_p(np^{1/2})$. Analogously, we can demonstrate that the second term in (22), $\mathcal{E}^\top H_a \bar{\mathcal{H}}\mathcal{E}$, is also of order $o_p(np^{1/2})$. Accordingly, the first term on the right-hand side of (21) is of order $o_p(1)$. Finally, applying similar techniques as those used in the above proofs, we are able to show that the second term on the right-hand side of (21) is also of order $o_p(1)$, which completes the entire proof. □

Appendix E. Proof of Theorem 3

According to the proof of Theorem 2, we only need to demonstrate that $\tilde{T}_2/\text{var}^{1/2}(\tilde{T}_2) \to_d$

$N(0,1)$. To this end, denote

$$\tilde{T}_2 = n^{-1}\sum_{1\le i_1\ne i_2\le n}\sum_{j\in\mathcal{S}}\varepsilon_{i_1}\varepsilon_{i_2}X^*_{i_1j}X^*_{i_2j} + n^{-1}\sum_{i=1}\sum_{j\in\mathcal{S}}\varepsilon_i^2(X^{*2}_{ij}-1) \doteq \Pi_1+\Pi_2.$$

Then, it suffices to show that

$$\Pi_2/\mathrm{var}^{1/2}(\tilde{T}_2) = o_p(1) \;\text{ and }\; \Pi_1/\mathrm{var}^{1/2}(\tilde{T}_2) \to_d N(0,1), \tag{23}$$

and the detailed proofs are given in the following two steps, respectively.

Step (1). After algebraic simplification with condition (C2.a), we obtain

$$\begin{aligned}
&n^2\sigma^{-4}\mathrm{var}(\Pi_2)\\
&= \sigma^{-4}\sum_{1\le i_1,i_2\le n}\sum_{j_1,j_2\in\mathcal{S}}E\{\varepsilon_{i_1}^2\varepsilon_{i_2}^2(X^{*2}_{i_1j_1}-1)(X^{*2}_{i_2j_2}-1)\}\\
&= \sum_{i_1\ne i_2}\sum_{j_1,j_2\in\mathcal{S}}E\{(X^{*2}_{i_1j_1}-1)(X^{*2}_{i_2j_2}-1)\} + (3+\Delta)\sum_{i=1}^{n}\sum_{j_1,j_2\in\mathcal{S}}E\{(X^{*2}_{ij_1}-1)(X^{*2}_{ij_2}-1)\}\\
&= (3+\Delta)\sum_{i=1}^{n}\sum_{j_1,j_2\in\mathcal{S}}\{E(X^{*2}_{ij_1}X^{*2}_{ij_2})-1\}\\
&= (3+\Delta)n\sum_{j_1,j_2\in\mathcal{S}}\{E(X^{*2}_{ij_1}X^{*2}_{ij_2})-1\} = O(np),
\end{aligned}$$

where the last equality is due to (6). Therefore, $\mathrm{var}(\Pi_2) = o(p)$. Recalling that $E(\Pi_2)=0$ and $\mathrm{var}(\tilde{T}_2) = 2\sigma^4 tr(\Sigma^2_{b|a})\{1+o(1)\} \ge \sigma^4 p\tau^2_{\min}$, this completes the proof of the first term of (23).

Step (2). Applying similar techniques used in the proof of Theorem 2, we have that $\mathrm{var}(\Pi_1) = 2\sigma^4 tr(\Sigma^2_{b|a})\{1+o(1)\}$. Then, $\mathrm{var}(\Pi_1)^{-1}\mathrm{var}(\tilde{T}_2)$ tends to 1 as n goes to infinity. Hence, we only need to show that

$$\Pi_1/\mathrm{var}^{1/2}(\Pi_1) \to_d N(0,1). \tag{24}$$

To this end, define $\mathcal{F}_r = \sigma\{\varepsilon_{i_1}, X^*_{i_2 j}, 1 \le j \le p, 1 \le i_1, i_2 \le r, i_1 \ne i_2\}$, the σ-field generated by $\{\varepsilon_{i_1}, X^*_{i_2 j}\}$, where $1 \le j \le p, 1 \le i_1, i_2 \le r$ for $r = 1, 2, \cdots, n$ and $i_1 \ne i_2$. In addition, define

$$T_{n,r} = n^{-1} \sum_{1 \le i_1 \ne i_2 \le r} \sum_{j \in \mathcal{S}} \varepsilon_{i_1} \varepsilon_{i_2} X^*_{i_1 j} X^*_{i_2 j}.$$

Obviously, $T_{n,r} \in \mathcal{F}_r$. Then, set $\Delta_{n,r} = T_{n,r} - T_{n,r-1}$ with $\Delta_0 = 0$. One can easily verify that $E(\Delta_{n,r}|\mathcal{F}_q) = 0$ and $E(T_r|\mathcal{F}_q) = T_q$ for any $q < r$. This implies that, for an arbitrary fixed n, $\{\Delta_{n,r}, 0 \le r \le n\}$ is a martingale difference sequence with respect to $\{\mathcal{F}_r, 0 \le r \le n\}$ with $\mathcal{F}_0 = \emptyset$. Accordingly, by the Martingale Central Limit Theorem (Hall and Heyde, 1980), for the proof of (24), it suffices to show that

$$\frac{\sum_{r=1}^n \sigma^{*2}_{n,r}}{\text{var}(\Pi_1)} \to_p 1 \quad \textit{and} \quad \frac{\sum_{r=1}^n E(\Delta^4_{n,r})}{\text{var}^2(\Pi_1)} \to_p 0 \tag{25}$$

where $\sigma^{*2}_{n,r} = E(\Delta^2_{n,r}|\mathcal{F}_{r-1})$.

We begin by showing the first term of (25). After algebraic simplification, we have that $\Delta_{n,r} = 2n^{-1} \sum_{i<r} \sum_{j \in \mathcal{S}} X^*_{ij} X^*_{rj} \varepsilon_i \varepsilon_r$ and

$$\begin{aligned}
\sum_{r=1}^n \sigma^{*2}_{n,r} &= \sum_{r=1}^n 4n^{-2} E(\sum_{i_1<r} \sum_{i_2<r} \sum_{j_1,j_2 \in \mathcal{S}} X^*_{i_1 j_1} X^*_{i_2 j_2} X^*_{r j_1} X^*_{r j_2} \varepsilon_{i_1} \varepsilon_{i_2} \varepsilon_r^2 | \mathcal{F}_{r-1}) \\
&= 4n^{-2}\sigma^2 \sum_{r=1}^n \sum_{i_1<r} \sum_{i_2<r} \sum_{j_1,j_2 \in \mathcal{S}} \varepsilon_{i_1} \varepsilon_{i_2} X^*_{i_1 j_1} X^*_{i_2 j_2} \sigma^*_{j_1 j_2} \\
&= 4n^{-2}\sigma^2 (\sum_{r=1}^n \sum_{i_1 \ne i_2 < r} \sum_{j_1,j_2 \in \mathcal{S}} \varepsilon_{i_1} \varepsilon_{i_2} X^*_{i_1 j_1} X^*_{i_2 j_2} \sigma^*_{j_1 j_2} \\
&\quad + \sum_{r=1}^n \sum_{i<r} \sum_{j_1,j_2 \in \mathcal{S}} \varepsilon_i^2 X^*_{i j_1} X^*_{i j_2} \sigma^*_{j_1 j_2}).
\end{aligned}$$

Using $tr\{\Sigma_{b|a}^4\} = O(p)$, we then obtain

$$\begin{aligned}
&\mathrm{var}(\sum_{r=1}^{n}\sum_{i_1\neq i_2<r}\sum_{j_1,j_2\in\mathcal{S}}\varepsilon_{i_1}\varepsilon_{i_2}X_{i_1j_1}^*X_{i_2j_2}^*\sigma_{j_1j_2}^*) \\
&= 4E(\sum_{r_1,r_2}\sum_{i_1<i_2<r_1}\sum_{i_3<i_4<r_2}\sum_{\{j_1,j_2,j_3,j_4\}\in\mathcal{S}}\varepsilon_{i_1}\varepsilon_{i_2}\varepsilon_{i_3}\varepsilon_{i_4}X_{i_1j_1}^*X_{i_2j_2}^*X_{i_3j_3}^*X_{i_4j_4}^*\sigma_{j_1j_2}^*\sigma_{j_3j_4}^*) \\
&= 4\sigma^4\sum_{r_1,r_2}\sum_{i_1<i_2<\min(r_1,r_2)}\sum_{\{j_1,j_2,j_3,j_4\}\in\mathcal{S}}\sigma_{j_1j_3}^*\sigma_{j_2j_4}^*\sigma_{j_1j_2}^*\sigma_{j_3j_4}^* \\
&= 4\sigma^4\sum_{r_1,r_2}\sum_{i_1<i_2<\min(r_1,r_2)} tr(\Sigma_{b|a}^4) = O(n^4p) = o(n^4p^2). \qquad (26)
\end{aligned}$$

Furthermore, by condition (C2.b) and Cauchy's inequality, we have

$$\begin{aligned}
&\sigma^{-4}E\{\sum_{r=1}^{n}\sum_{i<r}\sum_{j_1,j_2\in\mathcal{S}}\varepsilon_i^2X_{ij_1}^*X_{ij_2}^*\sigma_{j_1j_2}^*\}^2 \\
&= \sigma^{-4}E(\sum_{r_1,r_2}\sum_{i_1<r_1}\sum_{i_2<r_2}\sum_{\{j_1,j_2,j_3,j_4\in\mathcal{S}\}}\varepsilon_{i_1}^2\varepsilon_{i_2}^2X_{i_1j_1}^*X_{i_1j_2}^*X_{i_2j_3}^*X_{i_2j_4}^*\sigma_{j_1j_2}^*\sigma_{j_3j_4}^*) \\
&= \sum_{r_1,r_2}\sum_{i_1<r_1,i_2<r_2,i_1\neq i_2}\sum_{\{j_1,j_2,j_3,j_4\in\mathcal{S}\}}\sigma_{j_1j_2}^{*2}\sigma_{j_3j_4}^{*2} \\
&\quad +(3+\Delta)\sum_{r_1,r_2}\sum_{i<\min(r_1,r_2)}\sum_{\{j_1,j_2,j_3,j_4\in\mathcal{S}\}}E\{X_{ij_1}^*X_{ij_2}^*X_{ij_3}^*X_{ij_4^*}\}\sigma_{j_1j_2}^*\sigma_{j_3j_4}^* \\
&\leq \sum_{r_1,r_2}\sum_{i_1<r_1,i_2<r_2,i_1\neq i_2} tr^2(\Sigma_{b|a}^2) \\
&\quad +2^{-1}(3+\Delta)\sum_{r_1,r_2}\sum_{i<\min(r_1,r_2)}\sum_{j_1,j_2,j_3,j_4\in\mathcal{S}}\{E(X_{ij_1}^*X_{ij_2}^*X_{ij_3}^*X_{ij_4}^*)\}^2 \\
&\quad +2^{-1}(3+\Delta)\sum_{r_1,r_2}\sum_{i<\min(r_1,r_2)}\sum_{j_1,j_2,j_3,j_4\in\mathcal{S}}\sigma_{j_1j_2}^{*2}\sigma_{j_3j_4}^{*2} \\
&= \sum_{r_1,r_2}\sum_{i_1<r_1,i_2<r_2} tr^2(\Sigma_{b|a}^2) + O(n^3p^2),
\end{aligned}$$

where the last equality is due to the fact that

$$\sum_{r_1,r_2}\sum_{i<\min(r_1,r_2)}\sum_{j_1,j_2,j_3,j_4\in\mathcal{S}}\{E(X_{ij_1}^*X_{ij_2}^*X_{ij_3}^*X_{ij_4}^*)\}^2 = O(n^3p^2) \qquad (27)$$

obtained via condition (C2.b) and

$$\sum_{r_1,r_2}\sum_{i<\min(r_1,r_2)}\sum_{j_1,j_2,j_3,j_4\in\mathcal{S}}\sigma^{*2}_{j_1j_2}\sigma^{*2}_{j_3j_4}=\sum_{r_1,r_2}\sum_{i<\min(r_1,r_2)}tr^2(\Sigma^2_{b|a})=O(n^3p^2). \quad (28)$$

This further implies that

$$\begin{aligned}
&\text{var}(\sum_{r=1}^{n}\sum_{i<r}\sum_{j_1,j_2\in\mathcal{S}}\varepsilon_i^2X^*_{ij_1}X^*_{ij_2}\sigma^*_{j_1j_2})\\
&= E\{\sum_{r=1}^{n}\sum_{i<r}\sum_{j_1,j_2\in\mathcal{S}}\varepsilon_i^2X^*_{ij_1}X^*_{ij_2}\sigma^*_{j_1j_2}\}^2-\{E(\sum_{r=1}^{n}\sum_{i<r}\sum_{j_1,j_2\in\mathcal{S}}\varepsilon_i^2X^*_{ij_1}X^*_{ij_2}\sigma^*_{ij})\}^2\\
&\le \sigma^4\sum_{r_1,r_2}\sum_{i_1<r_1,i_2<r_2}tr^2(\Sigma^2_{b|a})+O(n^3p^2)-\sigma^4\sum_{r_1,r_2}\sum_{i_1<r_1,i_2<r_2}tr^2(\Sigma^2_{b|a})\\
&= O(n^3p^2)=o(n^4p^2). \quad (29)
\end{aligned}$$

Moreover, it can easily be shown that $E(\sum_{r=1}^n\sigma^{*2}_{n,r})=\text{var}(\Pi_1)$. This, together with the fact that $\text{var}(\Pi_1)=2(1-n^{-1})\sigma^4tr(\Sigma^2_{b|a})\{1+o(1)\}\ge\sigma^4p\tau^2_{\min}$, as well as (26) and (29), yields $\text{var}(\sum_{r=1}^n\sigma^{*2}_{n,r})=o\{\text{var}^2(\Pi_1)\}$. This completes the proof of the first part of (25).

We next consider the second term of (25). Since $\Delta_{n,r}=2n^{-1}\sum_{i<r}\sum_{j\in\mathcal{S}}X^*_{ij}X^*_{rj}\varepsilon_i\varepsilon_r$, we have the following

$$\begin{aligned}
&2^{-4}n^4\sum_{r=1}^{n}E(\Delta^4_{n,r})\\
&= Eg\{\sum_{r=1}^{n}\sum_{i_1,i_2,i_3,i_4<r}\sum_{j_1,j_2,j_3,j_4\in\mathcal{S}}X^*_{i_1j_1}X^*_{i_2j_2}X^*_{i_3j_3}X^*_{i_4j_4}X^*_{rj_1}X^*_{rj_2}X^*_{rj_3}X^*_{rj_4}\varepsilon_r^4\varepsilon_{i_1}\varepsilon_{i_2}\varepsilon_{i_3}\varepsilon_{i4}g\}\\
&= Eg\{\sum_{r=1}^{n}\sum_{i<r}\sum_{j_1,j_2,j_3,j_4\in\mathcal{S}}X^*_{ij_1}X^*_{ij_2}X^*_{ij_3}X^*_{ij_4}X^*_{rj_1}X^*_{rj_2}X^*_{rj_3}X^*_{rj_4}\varepsilon_r^4\varepsilon_i^4g\}\\
&+ 6Eg\{\sum_{r=1}^{n}\sum_{i_1\ne i_2<r}\sum_{j_1,j_2,j_3,j_4\in\mathcal{S}}X^*_{i_1j_1}X^*_{i_1j_2}X^*_{i_2j_3}X^*_{i_2j_4}X^*_{rj_1}X^*_{rj_2}X^*_{rj_3}X^*_{rj_4}\varepsilon_r^4\varepsilon_{i_1}^2\varepsilon_{i_2}^2g\}. \quad (30)
\end{aligned}$$

It is noteworthy that $E(\varepsilon_i^4) = (3+\Delta)\sigma^4$. Then, by condition (C2.b), the first term on the right-hand side of (30) is smaller than the following quantity:

$$\{3+\Delta\}^2\sigma^8 \sum_{r=1}^{n} \sum_{i<r} \sum_{j_1,j_2,j_3,j_4 \in \mathcal{S}} \{E(X_{ij_1}^* X_{ij_2}^* X_{ij_3}^* X_{ij_4}^*)\}^2 = O(n^2 p^2). \tag{31}$$

By Cauchy's inequality, we know that $2\varepsilon_{i_1}^2 \varepsilon_{i_2}^2 \le \varepsilon_{i_1}^4 + \varepsilon_{i_2}^4$; this enables us to show that the second term on the right-hand side of (30) is less than the quantity given below.

$$6\{3+\Delta\}^2\sigma^8 \sum_{r=1}^{n} \sum_{i_1 \neq i_2 < r} \sum_{j_1,j_2,j_3,j_4 \in \mathcal{S}} \sigma_{j_1 j_2}^* \sigma_{j_3 j_4}^* E\{X_{ij_1}^* X_{ij_2}^* X_{ij_3}^* X_{ij_4}^*\} = O(n^3 p^2), \tag{32}$$

where the last equality above is due to (27) and (28). Equations (30), (31), and (32) lead to $\sum_{r=1}^n E(\Delta_{n,r}^4) = O(n^{-1}p^2)$. This, in conjunction with the fact that $\mathrm{var}(\Pi_1) \ge \sigma^4 p \tau_{\min}^2$, shows the second part of (25); the entire proof is complete. □

REFERENCES

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

Bao, Y. and Ullah, A. (2010), "Expectation of quadratic forms in normal and nonnormal variables with applications," *Journal of Statistical Planning and Inference*, 140, 1193–1205.

Bendat, J. S. and Piersol, A. G. (1966), *Measurement and Analysis of Random Data*, New York: Wiley.

Chatterjee, S. and Hadi, A. S. (2006), *Regression Analysis by Example, 4th Edition*, New Work: Wiley.

Chen, S. X. and Qin, Y. L. (2010), "A two sample test for high dimensional data with application 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.

Draper, N. R. and Smith, H. (1998), *Applied Regression Analysis, 3rd Edition*, New York: Wiley.

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

Fan, J., Fan, Y., and Lv, J. (2008), "High dimensional covariance matrix estimation using a factor model," *Journal of Econometrics*, 147, 186–197.

Hall, P. and Heyde, C. C. (1980), *Martingale Limit Theory and its Application*, New York: Academic Press.

Lehmann, E. L. (1998), *Theory of Point Estimation, 2nd Edition*, New York: Springer.

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, New York: Chapman and Hall.

Milliken, G. A. and Johnson, D. E. (2009), *Analysis of Messy Data, Volume I: Designed Experiments, 2nd Edition*, New York: Chapman and Hall.

Ravishanker, N. and Dey, D. K. (2001), *A First Course in Linear Model Theory*, New York: Chapman and Hall/CRC.

Seber, G. A. F. and Lee, A. J. (2003), *Linear Regression Analysis, 2nd Edition*, New York: Wiley.

Shao, J. (2003), *Mathematical Statistics, 2nd Edition*, New York: Springer-Verlag.

Srivastava, M. S. (2005), "Some tests concerning the covariance matrix in high dimensional data," *Journal of Japan Statistical Society*, 35, 251–272.

Tibshirani, R. J. (1996), "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society, Series B*, 58, 267–288.

Vittinghoff, E., Glidden, D., Shiboski, S., and McCulloch, C. E. (2010), *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*, New York: Springer.

Wang, H. (2009), "Forward regression for ultra-high dimensional variable screening," *Journal of the American Statistical Association*, 104, 1512–1524.

Weisberg, S. (2005), *Applied Linear Regression, 3rd Edition*, New York: Wiley.

Yandel, B. S. (1997), *Practical Data Analysis for Designed Experiments*, New York: CRC Press.

Zhang, C. H. and Huang, J. (2008), "The sparsity and bias of the lasso selection in high-dimensional linear regression," *The Annals of Statistics*, 36, 1567–1594.

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.

Table 1: Size of the bias-corrected test and the ZC-test for Example 3.1.

		Normal			Mixture		
			q			q	
n	p	0	8	15	0	8	15
Bias-Corrected Test							
100	200	0.066	0.041	0.045	0.062	0.059	0.057
	500	0.059	0.041	0.036	0.067	0.052	0.039
	1000	0.065	0.047	0.041	0.064	0.056	0.044
200	200	0.061	0.050	0.048	0.062	0.062	0.059
	500	0.057	0.050	0.042	0.068	0.062	0.060
	1000	0.053	0.052	0.055	0.065	0.056	0.051
ZC-Test							
100	200	0.045	0.035	0.024	0.052	0.049	0.034
	500	0.072	0.046	0.035	0.048	0.031	0.027
	1000	0.060	0.046	0.031	0.051	0.030	0.022
200	200	0.049	0.049	0.049	0.049	0.042	0.038
	500	0.060	0.054	0.044	0.046	0.046	0.040
	1000	0.060	0.060	0.057	0.064	0.059	0.043

Table 2: Size of the bias-corrected test and the ZC-test for Example 3.2.

		Normal			Mixture		
			q			q	
n	p	0	8	15	0	8	15
Bias-Corrected Test							
100	200	0.062	0.054	0.038	0.071	0.050	0.039
	500	0.070	0.045	0.035	0.072	0.047	0.042
	1000	0.062	0.040	0.035	0.057	0.056	0.035
200	200	0.070	0.055	0.057	0.066	0.053	0.051
	500	0.057	0.056	0.046	0.054	0.061	0.035
	1000	0.052	0.062	0.044	0.062	0.049	0.046
ZC-Test							
100	200	0.053	0.144	0.216	0.072	0.146	0.217
	500	0.059	0.203	0.458	0.062	0.228	0.484
	1000	0.063	0.356	0.762	0.041	0.321	0.743
200	200	0.070	0.096	0.134	0.063	0.084	0.149
	500	0.056	0.116	0.208	0.048	0.129	0.199
	1000	0.047	0.163	0.295	0.049	0.166	0.303

Figure 1: Power of the bias-corrected test (BC-test) and the ZC-test for Example 3.1 under the normal setting with $n = 100$, $p = 200$ and $q = 0$.

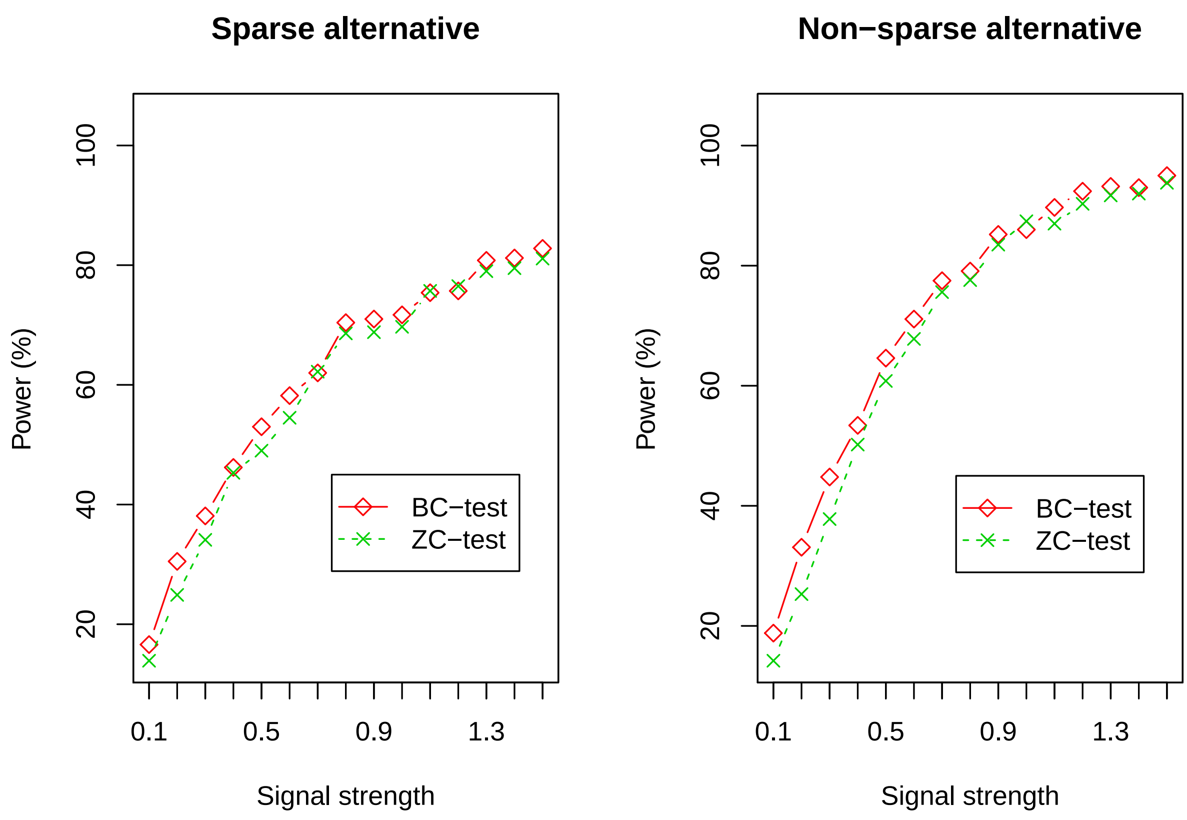

Figure 2: Power of the bias-corrected test (BC-test) for Example 3.2 under the normal setting for both sparse alternative (S) and non-sparse alternative (N) with $n = 100$, $p = 200$ and $q = 0$.