The Annals of Statistics
2018, Vol. 46, No. 3, 958–988
https://doi.org/10.1214/17-AOS1573

TEST FOR HIGH-DIMENSIONAL REGRESSION COEFFICIENTS USING REFITTED CROSS-VALIDATION VARIANCE ESTIMATION

BY HENGJIAN CUI*,[1], WENWEN GUO* AND WEI ZHONG†,[2]

Capital Normal University and Xiamen University*†

Testing a hypothesis for high-dimensional regression coefficients is of fundamental importance in the statistical theory and applications. In this paper, we develop a new test for the overall significance of coefficients in high-dimensional linear regression models based on an estimated U-statistics of order two. With the aid of the martingale central limit theorem, we prove that the asymptotic distributions of the proposed test are normal under two different distribution assumptions. Refitted cross-validation (RCV) variance estimation is utilized to avoid the overestimation of the variance and enhance the empirical power. We examine the finite-sample performances of the proposed test via Monte Carlo simulations, which show that the new test based on the RCV estimator achieves higher powers, especially for the sparse cases. We also demonstrate an application by an empirical analysis of a microarray data set on Yorkshire gilts.

1. Introduction. Conventional multivariate statistical approaches are generally derived for low-dimensional data where the number of covariates (p) is smaller than the sample size (n). However, high-dimensional data are increasingly collected in many applications of statistics such as genetical, biological and financial studies. This so-called "large p, small n" phenomenon not only brings many challenges to traditional multivariate statistical methods but also creates opportunities for statisticians to derive new methods. For example, the Hotelling's T^2 test is the conventional test for testing whether the two populations have the same mean when the dimension p is smaller than n. However, its power was proved by [1] to be adversely affected due to the nearly singularity of the sample covariance matrix for high-dimensional data. Several extensions of the Hotelling's T^2 test have been introduced in high-dimensional settings in the literature. Bai and Saranadasa [1] proposed a test statistic based on the squared Euclidean norm of two sample means to avoid the inverse of the sample covariance matrix. This test statistic was

Received February 2016; revised April 2017.

[1]Supported by National Natural Science Foundation of China (NNSFC) Grants 11471223, 11231010, 11071022, BCMIIS and the Key project of Beijing Municipal Educational Commission KZ201410028030.

[2]Supported by NNSFC Grants 11671334, 11301435, 11401497 and the Fundamental Research Funds for the Central Universities 20720140034.

MSC2010 subject classifications. 62F03, 62H15.

Key words and phrases. High-dimensional regression, hypothesis testing, martingale central limit theorem, refitted cross-validation variance estimation, U-statistics.

further improved by [3] via removing the cross-product terms. Srivastava and Du [17] replaced the sample covariance matrix in the Hotelling's T^2 test statistic by its diagonal to ensure the invertibility. Cai et al. [2] considered a test statistic based on a linear transformation of the data by the precision matrix. Wang et al. [19] developed a nonparametric test based on the spatial sign transformation under the elliptical distribution assumption.

Linear regression analysis is the most commonly used statistical approach to model the relationship between a response variable and many covariates. It is of fundamental importance to test the overall significance of linear regression coefficients. When the number of covariates p is fixed and less than the sample size n, the F-test [15] is the conventional method for testing the overall significance of linear regression coefficients. However, if p is greater than n in the high-dimensional problems, the F-test is no longer applicable. Zhong and Chen [24] showed that the power of the F-test is affected adversely by an increasing dimension even when $p < n-1$. Wang and Cui [20] proposed a generalized F-test and studied its asymptotic normality when $p/n \to \rho$ with $0 < \rho < 1$. However, the generalized F-test also fails when $p > n$ because the sample covariance matrix is not invertible. To deal with the high dimensionality, [24] developed a test based on a U-statistic for high-dimensional linear regression coefficients for both simple random or factorial designs. They derived its asymptotic normality under either the null hypothesis or the local alternatives. Goeman et al. [10, 11] proposed an empirical Bayes test by replacing a Mahalanobis norm in F-statistic by an Euclidian norm for high-dimensional linear regression and generalized linear models, respectively. Wang and Cui [21] further proposed a U-statistic like [24] for testing part of regression coefficients in high-dimensional linear models. Yata and Aoshima [22] developed a test by applying an extended cross-data-matrix methodology for testing high-dimensional correlations.

Motivated by an empirical study of the gene-set testing, we propose a new test for high-dimensional regression coefficients in this paper. Although this new test statistic shares the same spirit as the work in [24], it has three distinguishing features. First, the new test statistic based on an estimated U-statistic of order two can substantially reduce computational complexity compared with the U-statistic of order four in [24]. Second, we derive the asymptotic normality of the proposed test statistic under two different distribution assumptions: (I) the pseudo-independence assumption; (II) the elliptical contoured distribution assumption. Third, the idea of Refitted Cross-Validation (RCV) method introduced by [6] is used in the estimation of the test statistic to reduce the bias of the sample variance. As a result, the empirical power performance of the new test can be substantially enhanced, especially for the sparse cases.

The rest of the paper is organized as follows. Section 2 introduces the model settings and the new test for testing the significance of high-dimensional regression coefficients. In Section 3, the asymptotic distributions of the test statistic are derived under under the null hypothesis or the local alternatives. Section 4 reports empirical results from Monte Carlo simulations and Section 5 studies an empirical analysis of a microarray data set on Yorkshire gilts. The discussion is included in Section 6. Some useful lemmas and technical proofs as well as some figures are given in the Appendix.

2. A new test statistic.

2.1. *Model settings*. Let $(\mathbf{X}_i, Y_i)$ be the ith observation in a random sample for $i = 1, 2, \dots, n$, where Y_i denotes the ith response variable of interest and $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^{\mathrm{T}} \in \mathbb{R}^p$ denotes the p-dimensional vector of covariates with mean $E(\mathbf{X}_i) = \boldsymbol{\mu}$ and covariance matrix $\operatorname{cov}(\mathbf{X}_i) = \Sigma$. To model the regression relationship, we consider a linear regression

$$Y_i = \alpha + \mathbf{X}_i^{\mathrm{T}} \boldsymbol{\beta} + \varepsilon_i, \tag{2.1}$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^{\mathrm{T}} \in \mathbb{R}^p$ is a p-dimensional vector of regression coefficients of interest, α is a nuisance intercept parameter and $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ are independent and identically distributed random errors with mean zero and variance σ^2.

Our interest is testing the high-dimensional regression coefficients simultaneously in (2.1). That is,

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0 \quad \text{versus} \quad H_1 : \boldsymbol{\beta} \neq \boldsymbol{\beta}_0, \tag{2.2}$$

for some $\boldsymbol{\beta}_0 \in \mathbb{R}^p$. In particular, when $\boldsymbol{\beta}_0 = \mathbf{0}$, it tests the overall significance of linear regression coefficients.

2.2. *Test statistic*. For testing $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ in the linear regression (2.1) when $\alpha = 0$, we naturally consider the difference between the ordinary least squares estimator $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}} Y$ and $\boldsymbol{\beta}_0$, where $Y = (Y_1, Y_2, \dots, Y_n)^{\mathrm{T}}$ is the vector of response values and $\mathbf{X} = (\mathbf{X}_1^{\mathrm{T}}, \dots, \mathbf{X}_n^{\mathrm{T}})$ is the $n \times p$ design matrix. Note that $\widehat{\boldsymbol{\beta}}$ is practically infeasible when $p > n$ but it serves as a motivation for the test statistic. By noting that $\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0$ is equivalent to $\mathbf{X}^{\mathrm{T}}(Y - \mathbf{X}\boldsymbol{\beta}_0) = 0$, $E\|\mathbf{X}_i(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0)\|^2$ can be utilized as an effective measure of the discrepancy between $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$. [24] considered a U-statistic with $\mathbf{X}_i^{\mathrm{T}}\mathbf{X}_j(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0)(Y_j - \mathbf{X}_j^{\mathrm{T}}\boldsymbol{\beta}_0)$ for $i \neq j$ as the kernel to estimate $E\|\mathbf{X}_i(Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0)\|^2$ when $\alpha = 0$. To remove the effect of α when $\alpha \neq 0$ and $\boldsymbol{\mu} = E(\mathbf{X}_i) \neq \mathbf{0}$, they further proposed the following test statistic based on a U-statistic of order four for testing (2.2):

$$Z_{n,p} = \frac{(n-4)!}{4n!} \sum^{*} (\mathbf{X}_{i_1} - \mathbf{X}_{i_2})^{\mathrm{T}} (\mathbf{X}_{i_3} - \mathbf{X}_{i_4}) \Delta_{i_1,i_2} \Delta_{i_3,i_4}, \tag{2.3}$$

where $\Delta_{i,j} = (Y_i - \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta}_0) - (Y_j - \mathbf{X}_j^{\mathrm{T}}\boldsymbol{\beta}_0)$ and $\sum^{*}$ denotes summations over distinct indices.

In this paper, without loss of generality, we focus on testing the overall significance of the linear regression model (2.1), that is, we test $H_0 : \boldsymbol{\beta} = \mathbf{0}$. In this case, $E\|\mathbf{X}_i Y_i\|^2$ serves as an effective measure of the discrepancy between $\boldsymbol{\beta}$ and $\mathbf{0}$. When $\boldsymbol{\mu} = \mathbf{0}$ and $\alpha = 0$, it can be estimated by a U-statistic with $\mathbf{X}_i^{\mathrm{T}}\mathbf{X}_j Y_i Y_j$ for $i \neq j$ as the kernel. To eliminate the effect of nonzero $\boldsymbol{\mu}$ and α, we suggest to replace $\mathbf{X}_i$ and Y_i by their centralized counterparts $\mathbf{X}_i - \overline{\mathbf{X}}$ and $Y_i - \overline{Y}$, where $\overline{\mathbf{X}}$ and $\overline{Y}$ are respective means of $\mathbf{X}_i$'s and Y_i's. Although its computation is simple, the centralization brings in the bias for the estimator. To further correct this bias, we note the fact that $E[(\mathbf{X}_i - \overline{\mathbf{X}})^{\mathrm{T}}(\mathbf{X}_j - \overline{\mathbf{X}})] = -\operatorname{tr}(\Sigma)/n$ for $i \neq j$ and propose the following statistic by subtracting an unbiased estimator of the bias:

$$\Delta_{i,j}(\mathbf{X}) =: (\mathbf{X}_i - \overline{\mathbf{X}})^{\mathrm{T}}(\mathbf{X}_j - \overline{\mathbf{X}}) + \frac{\|\mathbf{X}_i - \mathbf{X}_j\|^2}{2n},$$

where $\|\cdot\|$ denotes the Euclidean norm. Similarly, we define

$$\Delta_{i,j}(Y) =: (Y_i - \overline{Y})(Y_j - \overline{Y}) + \frac{(Y_i - Y_j)^2}{2n}.$$

We can show that $E[\Delta_{i,j}(\mathbf{X})] = 0$, $E[\Delta_{i,j}(Y)] = 0$. This motivates us to construct the following test statistic:

$$T_{n,p} = \left(1 - \frac{2}{n}\right)^{-2} \binom{n}{2}^{-1} \sum_{i=2}^{n} \sum_{j=1}^{i-1} \Delta_{i,j}(\mathbf{X}) \Delta_{i,j}(Y) \tag{2.4}$$

for testing (2.2), where $(1 - 2/n)^{-2}$ is a modified constant which makes the mean of the test statistic $T_{n,p}$ have a simple dominant term $\boldsymbol{\beta}^{\mathrm{T}}\Sigma^2\boldsymbol{\beta}$ in Theorems 3.1 and 3.3. The test statistic $T_{n,p}$ serves as an effective measure of the difference between $\boldsymbol{\beta}$ and $\mathbf{0}$ and is invariant to location shifts in both $\mathbf{X}_i$ and Y_i. It can be viewed as an estimated U-statistic of order two. Compared with the U-statistic of order four in [24], the new test statistic can substantially reduce computational complexity. This feature will be confirmed by Monte Carlo simulations in Section 4.

3. Main results. In this section, we derive the asymptotic distributions of the new test statistic $T_{n,p}$ under two different kinds of covariates assumptions: (I) the pseudo-independence assumption; (II) the elliptical contoured distribution assumption; how to implement the new test is also discussed. In particular, the idea of refitted cross-validation (RCV) technique in [6] is applied for the variance estimation of $T_{n,p}$ to reduce the bias of the sample variance and enhance the empirical power.

We first impose the condition on the dimension of the covariates, which describes the "large p, small n" paradigm:

(C1) $p \to \infty$ as $n \to \infty$; $\Sigma > 0$, $\operatorname{tr}(\Sigma^4) = o\{\operatorname{tr}^2(\Sigma^2)\}$.

This condition is also imposed by [3, 19, 24] to derive the asymptotic properties of their test statistics. It does not assume any explicit relationship between p and n. Thus, it can accommodate the high dimensionality. The positive definiteness of Σ

ensures the identification of the linear regression coefficients. $\operatorname{tr}(\Sigma^4) = o\{\operatorname{tr}^2(\Sigma^2)\}$ holds trivially if all eigenvalues of Σ are bounded away from 0 and ∞. [3] also showed $\operatorname{tr}(\Sigma^4) = o\{\operatorname{tr}^2(\Sigma^2)\}$ holds under some general conditions even if some of the eigenvalues are unbounded. Thus, it is milder than the Riesz condition which is often assumed in the high-dimensional literature, such as [23]. In fact, (C1) imposes some limitations on p and Σ but it be considered as mild.

For the local power analysis, we consider the following local alternative hypothesis:

(C2) $\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta} = o(1)$, and $\boldsymbol{\beta}^{\mathrm{T}}\Sigma^3\boldsymbol{\beta} = o\{n^{-1}\operatorname{tr}(\Sigma^2)\}$.

Condition (C2) was also considered in [24]. It prescribes a small discrepancy between $\boldsymbol{\beta}$ and $\mathbf{0}$, thus it specifies a class of the local alternatives.

3.1. *Results on pseudo-independence assumption.* We first assume that the covariates satisfy the following pseudo-independence assumption.

(C3) [The pseudo-independence assumption.] Suppose the random vector $\mathbf{X}_i$ follows the general multivariate model $\mathbf{X}_i = \Gamma\mathbf{Z}_i + \mu$, where $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{im})^{\mathrm{T}}$ is a m-variate random vector for some $m \geq p$ satisfying $E(\mathbf{Z}_i) = 0$ and $\operatorname{Var}(\mathbf{Z}_i) = \mathbf{I}_m$, Γ is a $p \times m$ matrix such that $\Gamma\Gamma^{\mathrm{T}} = \Sigma$. Furthermore, we assume that $E(Z_{il}^4) = 3 + \Delta < \infty$ for some constant Δ and $E(Z_{1i_1}^{l_1} Z_{1i_2}^{l_2} \cdots Z_{1i_d}^{l_d}) = E(Z_{1i_1}^{l_1}) E(Z_{1i_2}^{l_2}) \cdots E(Z_{1i_d}^{l_d})$ for any $\sum_{v=1}^{d} l_v \leq 4$ and $i_1 \neq \cdots \neq i_d$, where d is a positive integer.

Condition (C3) which resembles a factor model structure assumes that the covariates $\mathbf{X}_i$ are generated linearly by a larger dimensional factor vector $\mathbf{Z}_i$. We note that the number of factors m in (C3) is assumed to be at least as large as p while the traditional factor model assumes m should be far less than p. The similar versions of the pseudo-independence assumptions were assumed in the literature. Bai and Saranadasa [1] require each Z_{il} is independent from one another while [3, 24] assume that $\sum_{v=1}^{d} l_v \leq 8$.

The theoretical properties of the test statistic $T_{n,p}$ under the pseudo-independence assumption are presented in the following theorems.

THEOREM 3.1. *Suppose conditions* (C1) *and* (C3) *hold*, *we have*:

(i) *The expectation of the test statistic* (2.4) *is*

$$E(T_{n,p}) = \boldsymbol{\beta}^{\mathrm{T}}\Sigma^2\boldsymbol{\beta} + \frac{\Delta}{2n(n-2)}\big(\boldsymbol{\beta}^{\mathrm{T}}\Gamma\operatorname{diag}(\Gamma^{\mathrm{T}}\Gamma)\Gamma^{\mathrm{T}}\boldsymbol{\beta}\big).$$

(ii) *Under condition* (C2), *the reconstruction of the test statistic* (2.4) *is as follows*:

$$T_{n,p} - \boldsymbol{\beta}^{\mathrm{T}}\Sigma^2\boldsymbol{\beta} = \binom{n}{2}^{-1}\sum_{i=2}^{n}\sum_{j=1}^{i-1}(\mathbf{X}_i - \boldsymbol{\mu})^{\mathrm{T}}(X_j - \boldsymbol{\mu})\varepsilon_i\varepsilon_j + o_P\big(n^{-1}\sqrt{\operatorname{tr}(\Sigma^2)}\big).$$

We can show that the second term of $E(T_{n,p})$ can be negligible when we establish the asymptotical distributions of $T_{n,p}$ under conditions (C1) and (C2), that is,

$$\begin{aligned}\frac{n[E(T_{n,p})-\boldsymbol{\beta}^{\mathrm{T}}\Sigma^2\boldsymbol{\beta}]}{\sqrt{\operatorname{tr}(\Sigma^2)}} &\le O(n^{-1})\frac{\lambda_{\max}(\Sigma)\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta}}{\sqrt{\operatorname{tr}(\Sigma^2)}}\\ &\le O(n^{-1})\frac{[\operatorname{tr}(\Sigma^4)]^{1/4}}{\sqrt{\operatorname{tr}(\Sigma^2)}}\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta}\\ &= o(1).\end{aligned}$$

With the aid of the reconstruction of the test statistic (2.4) in Theorem 3.1 and the martingale central limit theorem, we can establish the asymptotic distribution of $T_{n,p}$ in the following theorem.

THEOREM 3.2. *Assume condition* (C1) *and the pseudo-independence assumption* (C3) *hold. Then under either* H_0 *or the local alternatives in* (C2), *as* $n \to \infty$, *we have*

$$\frac{n(T_{n,p}-\boldsymbol{\beta}^{\mathrm{T}}\Sigma^2\boldsymbol{\beta})}{\sigma^2\sqrt{2\operatorname{tr}(\Sigma^2)}} \xrightarrow{D} N(0,1), \tag{3.1}$$

where $\xrightarrow{D}$ *denotes the convergence in distribution.*

3.2. *Results on elliptical distribution assumption.* In the multivariate statistical analysis, elliptical distributions are often assumed. The family of elliptical distributions includes multivariate normal distribution, multivariate t distribution, multivariate logistic distribution and among others [9]. They extend the multivariate normal distribution to a very flexible family of distributions, which are capable to accommodate tail dependence and considered to be useful for quantitative finance research [16]. In this subsection, we assume the covariates follow the elliptical contoured distributions.

(C4) [The elliptical distribution assumption.] Suppose the covariates satisfy the stochastic representation: $\mathbf{X}_i = \boldsymbol{\mu} + \Gamma R_i \mathbf{U}_i$, where Γ is a $p \times p$ matrix, $\mathbf{U}_i$ is a random vector uniformly distributed on the unit sphere in $\mathbb{R}^p$ and R_i is a nonnegative random variable independent of $\mathbf{U}_i$ and $E(R_i^2) = p$, $\operatorname{Var}(R_i^2) = O(p)$.

We consider the assumption that $E(R_i^2) = p$, $\operatorname{Var}(R_i^2) = O(p)$ as mild. For example, if R_i^2 follows a chi-square distribution with p degrees of freedom, then $E(R_i^2) = p$, $\operatorname{Var}(R_i^2) = 2p$ and $\mathbf{X}_i$ has a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\Sigma = \Gamma\Gamma^{\mathrm{T}}$. This assumption is also satisfied when $R_i^2 = \sum_{j=1}^p \xi_j$, where $\xi_1, \dots, \xi_p$ are independent, identically distributed and nonnegative with $E(\xi_j) = 1$ and $E(\xi_j^2) < +\infty$. Note that there exists an equivalent definition of elliptical contoured distribution, that is, $\mathbf{a}^{\mathrm{T}}(\mathbf{X}_i - \boldsymbol{\mu}) = \sqrt{\mathbf{a}^{\mathrm{T}}\Sigma\mathbf{a}}W_i$ for any p-dimensional vector $\mathbf{a}$, where W_i follows a symmetric distribution. The constraint that $E(R_i^2) = p$ and $\mathrm{Var}(R_i^2) = O(p)$ is equivalent to $E(W_i^2) = 1$ and $E(W_i^4) = 3 + o(1)$, respectively.

Similar to Theorem 3.1 and 3.2, the theoretical properties of the new test statistic (2.4) $T_{n,p}$ are also studied under the elliptical contoured distribution assumption (C4) in the following two theorems.

THEOREM 3.3. *Suppose conditions* (C1) *and* (C4) *hold*, *we have*:

(i) *The expectation of the proposed test statistic* (2.4) *is*

$$E(T_{n,p}) = \boldsymbol{\beta}^{\mathrm{T}}\Sigma^2\boldsymbol{\beta} + \frac{1}{n(n-2)}\left(\frac{ER^4}{p(p+2)} - 1\right)\left(\boldsymbol{\beta}^{\mathrm{T}}\Sigma^2\boldsymbol{\beta} + \frac{\mathrm{tr}(\Sigma)\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta}}{2}\right).$$

(ii) *Under condition* (C2), *the reconstruction of the test statistic* (2.4) *is as follows*:

$$T_{n,p} - \boldsymbol{\beta}^{\mathrm{T}}\Sigma^2\boldsymbol{\beta} = \binom{n}{2}^{-1}\sum_{i=2}^{n}\sum_{j=1}^{i-1}(\mathbf{X}_i - \boldsymbol{\mu})^{\mathrm{T}}(X_j - \boldsymbol{\mu})\varepsilon_i\varepsilon_j + o_P\big(n^{-1}\sqrt{\mathrm{tr}(\Sigma^2)}\big).$$

THEOREM 3.4. *Assume condition* (C1) *and the elliptical contoured distribution assumption* (C4) *hold. Then under either H_0 or the local alternatives in* (C2), *as $n \to \infty$*, *we have*

$$\frac{n(T_{n,p} - \boldsymbol{\beta}^{\mathrm{T}}\Sigma^2\boldsymbol{\beta})}{\sigma^2\sqrt{2\,\mathrm{tr}(\Sigma^2)}} \xrightarrow{D} N(0, 1). \tag{3.2}$$

Under the pseudo-independence assumption (C3) or the elliptical contoured distribution assumption (C4), both Theorems 3.2 and 3.4 state that the asymptotic null distribution of the new statistic $T_{n,p}$ under $H_0 : \boldsymbol{\beta} = \mathbf{0}$ is normal, that is,

$$\frac{nT_{n,p}}{\sigma^2\sqrt{2\,\mathrm{tr}(\Sigma^2)}} \xrightarrow{D} N(0, 1). \tag{3.3}$$

Meanwhile, Theorems 3.2 and 3.4 also imply that the proposed level α test has the following asymptotic local power under the local alternatives in (C2):

$$\Psi_n^{\mathrm{New}} = \Phi\left(-z_\alpha + \frac{n\boldsymbol{\beta}^{\mathrm{T}}\Sigma^2\boldsymbol{\beta}}{\sigma^2\sqrt{2\,\mathrm{tr}(\Sigma^2)}}\right), \tag{3.4}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution and z_α denotes the $1 - \alpha$ quantile of the standard normal distribution. The signal-to-noise ratio term $\eta(\boldsymbol{\beta}, \Sigma) =: \boldsymbol{\beta}^{\mathrm{T}}\Sigma^2\boldsymbol{\beta}/\sigma^2\sqrt{\mathrm{tr}(\Sigma^2)}$ essentially controls the power of the test. When $\eta(\boldsymbol{\beta}, \Sigma)$ is of a smaller order of n^{-1}, the power diminishes to α. In this case, the test can not distinguish the null hypothesis from the local alternatives. If $\eta(\boldsymbol{\beta}, \Sigma)$ has a higher order of n^{-1}, the power converges to 1 which implies that the proposed test is consistent. This asymptotic local power is same as (2.3) in [24] where their asymptotic local power

$$\Psi_n^{ZC} = \Phi\left(-z_\alpha + \frac{n\boldsymbol{\beta}^{\mathrm{T}}\Sigma^2\boldsymbol{\beta}}{\sigma^2\sqrt{2\,\mathrm{tr}(\Sigma^2)}}\right). \tag{3.5}$$

Thus, the new test shares the identical asymptotic local performance with the ZC test in [24]. It will be demonstrated by numerical studies in Section 4.

In practice, we denote $\widehat{\sigma}^2$ and $\widehat{\mathrm{tr}(\Sigma^2)}$ be the estimators of σ^2 and $\mathrm{tr}(\Sigma^2)$, respectively. When the sample size n is large enough, the asymptotic null distribution (3.3) can be used to construct the critical region of the proposed test, that is, we can reject H_0 at the significant level α if

$$nT_{n,p} \ge \widehat{\sigma}^2\sqrt{2\widehat{\mathrm{tr}(\Sigma^2)}}z_\alpha. \tag{3.6}$$

To estimate $\mathrm{tr}(\Sigma^2)$, we choose the unbiased and ratio consistent estimator in [4, 24],

$$\widehat{\mathrm{tr}(\Sigma^2)} = S_{1n} - 2S_{2n} + S_{3n},$$

where $S_{1n} = (n-2)!(n!)^{-1}\sum_{i\neq j}(X_i^{\mathrm{T}}X_j)^2$, $S_{2n} = (n-3)!(n!)^{-1}\sum_{i\neq j\neq k}(X_i^{\mathrm{T}} \times X_jX_j^{\mathrm{T}}X_k)$ and $S_{3n} = (n-4)!(n!)^{-1}\sum_{i\neq j\neq k\neq l}(X_i^{\mathrm{T}}X_jX_k^{\mathrm{T}}X_l)$. To estimate σ^2, [24] simply used the sample variance $\hat{\sigma}_0^2 = (n-1)^{-1}\sum_{i=1}^n(Y_i - \overline{Y})^2$ to estimate σ^2.

3.3. *Refitted cross-validation variance estimation.* It can be shown that $E(\hat{\sigma}_0^2) = \sigma^2 + \boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta}$. When $H_0 : \boldsymbol{\beta} = \mathbf{0}$ holds, $\hat{\sigma}_0^2$ is unbiased and consistent for σ^2. However, when $H_1 : \boldsymbol{\beta} \neq \mathbf{0}$, the sample variance $\hat{\sigma}_0^2$ has a positive bias $\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta}$ which overestimates σ^2 and negatively reduce the power of the test in the finite-sample performance.

In this paper, we suggest the refitted cross-validation (RCV) approach to estimate σ^2 in order to reduce the bias of the sample variance and enhance the power performance of our test. The RCV method was originally proposed by [6] for error variance estimation in ultrahigh-dimensional linear regression. The procedure is implemented in the following way. First, the original data set $\{Y, \mathbf{X}\}$ is partitioned randomly into two even data sets, denoted by $(Y^{(k)}, \mathbf{X}^{(k)})$ with sample size n_k for $k = 1, 2$. Second, a variable selection method is applied for the first data set to select a subset of covariates, denoted by $\hat{M}_1$. In this step, we require that the sure screening property holds, that is, $\hat{M}_1$ contains the true set of covariates in ultrahigh-dimensional space with probability tending to one. In practice, a sure independence screening method can be applied here, such as SIS in [8] for normal linear regression, DC-SIS by [13] in a model-free sense, etc., or a sophisticated penalized regression method can be utilized, such as Lasso [18], SCAD [7], adaptive Lasso [25] and among others. In the third step, we perform an ordinary least squares method on the second data set $(Y^{(2)}, \mathbf{X}^{(2)}_{\hat{M}_1})$ to estimate the variance σ^2, where $\mathbf{X}^{(2)}_{\hat{M}_1}$ denotes the submatrix of $\mathbf{X}^{(2)}$ whose columns are indexed by $\hat{M}_1$, that is,

$$\hat{\sigma}_1^2 = \frac{Y^{(2)T}[I_{n_2} - P_{\hat{M}_1}(\mathbf{X}^{(2)})]Y^{(2)}}{n_2 - |\hat{M}_1|}, \tag{3.7}$$

where $P_{\hat{M}_1}(\mathbf{X}^{(2)}) = \mathbf{X}^{(2)}_{\hat{M}_1}(\mathbf{X}^{(2)T}_{\hat{M}_1}\mathbf{X}^{(2)}_{\hat{M}_1})^{-1}\mathbf{X}^{(2)T}_{\hat{M}_1}$ is the projection operator onto the linear space that is generated by the column vectors of $\mathbf{X}^{(2)}_{\hat{M}_1}$. Then we switch the roles of two data sets and repeat the previous two steps to obtain another estimator of σ^2:

$$\hat{\sigma}_2^2 = \frac{Y^{(1)T}[I_{n_1} - P_{\hat{M}_2}(\mathbf{X}^{(1)})]Y^{(1)}}{n_1 - |\hat{M}_2|}. \tag{3.8}$$

At last, the final Refitted Cross-Validation (RCV) estimator of σ^2 is defined as

$$\hat{\sigma}^2_{\mathrm{RCV}} = (\hat{\sigma}_1^2 + \hat{\sigma}_2^2)/2. \tag{3.9}$$

It is crucial to establish consistency of the RCV estimator $\hat{\sigma}^2_{\mathrm{RCV}}$ under the alternative hypothesis. Under the sparsity assumption, the sure screening property of the variable selection procedure and some regularity conditions, [6] showed that $\hat{\sigma}^2_{\mathrm{RCV}}$ enjoys an oracle property, that is, $\sqrt{n}(\hat{\sigma}^2_{\mathrm{RCV}} - \sigma^2) \xrightarrow{D} N(0, E(\varepsilon^4) - \sigma^4)$, which implies the consistency of $\hat{\sigma}^2_{\mathrm{RCV}}$. If the true coefficient $\boldsymbol{\beta}$ is not sparse but satisfies some decay condition such as $\sum |\beta_j| \le C$ for some positive constant C, the model selection method could select a majority of all variables with large coefficients in the first stage, and the RCV estimator also performs well. More details on estimation and proprieties of $\hat{\sigma}^2_{\mathrm{RCV}}$ can be found in [6].

The following theorem further shows that the RCV estimator $\hat{\sigma}^2_{\mathrm{RCV}}$ is also consistent to σ^2 under the local alternative hypothesis (C2) which may be nonsparse.

THEOREM 3.5. *Assume $|\hat{M}_i|/n \to 0$ for $i = 1, 2$, then under the local alternatives in* (C2), *as $n \to \infty$, we have $\hat{\sigma}^2_{\mathrm{RCV}} \xrightarrow{P} \sigma^2$.*

Thus, we propose to use the RCV estimator $\hat{\sigma}^2_{\mathrm{RCV}}$ to estimate σ^2 for the test statistic (2.4). As a result, the asymptotic normality of the proposed test in Theorems 3.2 and 3.4 is also valid by Slutsky's theorem. Therefore, the proposed test based on $\hat{\sigma}^2_{\mathrm{RCV}}$ rejects H_0 at a significant level α for n sufficiently large if

$$nT_{n,p} \geq \hat{\sigma}^2_{\mathrm{RCV}}\sqrt{2\widehat{\mathrm{tr}(\Sigma^2)}}z_\alpha. \tag{3.10}$$

The numerical studies in the next section will show that the RCV variance estimation can enhance the empirical power performance, especially for the sparse case. For simplicity of presentation, we denote the decision rule (3.6) of our test based on $\hat{\sigma}^2_0$ and $\hat{\sigma}^2_{\mathrm{RCV}}$ by New_0 and $\mathrm{New}_{\mathrm{RCV}}$, respectively.

Note that we can easily obtain $\hat{\sigma}^2_0 = \sigma^2 + \boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta} + O_P(n^{-1/2})$. Thus, the sample variance $\hat{\sigma}^2_0$ is still consistent under the local alternative condition $\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta} = o(1)$. This means that both the ZC test and the New_0 test based on $\hat{\sigma}^2_0$ are also asymptotically valid under the local alternative condition. On the other hand, $\hat{\sigma}^2_{\mathrm{RCV}} = \sigma^2 + O_P(n^{-1/2})$ in the sparse case. Thus, the ratio of the two estimators $\hat{\sigma}^2_0$ and $\hat{\sigma}^2_{\mathrm{RCV}}$ is

$$\begin{aligned}\frac{\hat{\sigma}^2_0}{\hat{\sigma}^2_{\mathrm{RCV}}} &= \frac{\sigma^2 + \boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta} + O_P(n^{-1/2})}{\sigma^2 + O_P(n^{-1/2})} \\ &= 1 + O_P(\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta}),\end{aligned}$$

which tends to one in probability as $n \to \infty$ under the local alternative condition $\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta} = o(1)$. However, it is greater than one in the sample level, which results in that the $\mathrm{New}_{\mathrm{RCV}}$ test obtains substantially higher empirical powers than the ZC test and New_0 especially in the sparse case. Tables 1 and 2 in the simulations will confirm this result.

4. Simulations. In this section, we examine the finite sample performance of the proposed test by Monte Carlo simulations. We compare it with the empirical Bayes (EB) test in [11] and ZC test in [24]. Consider a linear regression model

$$Y_i = \alpha + \mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_i, \tag{4.1}$$

where $\alpha = 2$ and ε_i follows (i) $N(0, 1)$ or (ii) $t(5)/\sqrt{5/3}$. We randomly generate $\mathbf{X}_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$ from three p-dimensional distributions: (1) $N_p(\boldsymbol{\mu}, \Sigma)$, (2) $\sqrt{1 - 2/q}\,t_p(\boldsymbol{\mu}, \Sigma, q)$ with $q = 5$, (3) $X \sim \Sigma^{1/2}Z$ with $Z = (Z_1, Z_2, \ldots, Z_p)$ and $Z_j \sim$Uniform$(-\sqrt{3}, \sqrt{3})$ for $j = 1, 2, \ldots, p$. Here, each element of the mean vector $\boldsymbol{\mu}$ follows Uniform(2, 3) and the covariance matrix $\Sigma = (\sigma_{jl})_{p\times p}$, where $\sigma_{jl} = \sum_{k=1}^{T-|j-l|}\rho_k\rho_{k+|j-l|}I\{|j - l| < T\}$ with $T = 10$ and $\{\rho_k\}_{k=1}^{T} \sim$Uniform(0, 1) independently. Two high-dimensional settings of (n, p) are considered, (20, 178) and (40, 230), which are implied by an exponential form $p = \exp(n^{0.4}) + 150$. We consider two cases for $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$: (a) nonsparse case, $\beta_j = \|\boldsymbol{\beta}\|/\sqrt{p/2}$ for $j = 1, \ldots, p/2$ and $\beta_j = 0$, otherwise; (b) sparse case, $\beta_j = \|\boldsymbol{\beta}\|/\sqrt{5}$ for $j = 1, \ldots, 5$ and $\beta_j = 0$, otherwise, where $\|\boldsymbol{\beta}\|$ is the L_2 norm of $\boldsymbol{\beta}$ which will be specified later in the tables.

TABLE 1

Empirical Powers of EB, ZC, New_0 and New_{RCV} tests at the significant level 5% when $\varepsilon \sim N(0, 1)$

		Nonsparse case				Sparse case			
(n, p)	**$\|\beta\|^2$**	**EB**	**ZC**	**New_0**	**New_{RCV}**	**EB**	**ZC**	**New_0**	**New_{RCV}**
(1) $\mathbf{X}_i \sim N_p(\mu, \Sigma)$									
(20, 178)	0.00	0.05	0.06	0.06	0.06	0.05	0.06	0.06	0.06
	0.05	0.25	0.38	0.38	0.44	0.09	0.18	0.18	0.24
	0.10	0.36	0.56	0.56	0.61	0.13	0.26	0.26	0.37
	0.15	0.44	0.67	0.67	0.69	0.14	0.34	0.34	0.46
	0.20	0.49	0.71	0.71	0.75	0.16	0.38	0.38	0.52
(40, 230)	0.00	0.04	0.05	0.05	0.06	0.04	0.05	0.05	0.06
	0.05	0.58	0.73	0.73	0.73	0.18	0.28	0.28	0.32
	0.10	0.74	0.90	0.90	0.90	0.25	0.47	0.47	0.56
	0.15	0.79	0.95	0.95	0.95	0.31	0.56	0.56	0.74
	0.20	0.83	0.97	0.97	0.98	0.35	0.67	0.67	0.81
(2) $\mathbf{X}_i \sim \sqrt{1-2/q}\, t_p(\mu, \Sigma, q)$									
(20, 178)	0.00	0.05	0.06	0.06	0.06	0.05	0.06	0.06	0.06
	0.05	0.27	0.36	0.38	0.42	0.12	0.17	0.18	0.22
	0.10	0.33	0.52	0.53	0.58	0.15	0.26	0.27	0.32
	0.15	0.37	0.59	0.60	0.64	0.18	0.31	0.32	0.43
	0.20	0.41	0.64	0.65	0.69	0.19	0.34	0.36	0.45
(40, 230)	0.00	0.05	0.06	0.06	0.06	0.05	0.06	0.06	0.06
	0.05	0.50	0.63	0.65	0.66	0.23	0.27	0.27	0.31
	0.10	0.67	0.81	0.82	0.84	0.31	0.43	0.44	0.52
	0.15	0.73	0.88	0.90	0.91	0.38	0.52	0.53	0.63
	0.20	0.77	0.92	0.92	0.94	0.43	0.59	0.60	0.71
(3) $\mathbf{X}_i \sim \Sigma^{1/2} Z$									
(20, 178)	0.00	0.05	0.06	0.06	0.06	0.05	0.06	0.06	0.06
	0.05	0.23	0.39	0.39	0.42	0.08	0.15	0.15	0.19
	0.10	0.34	0.56	0.56	0.61	0.12	0.23	0.23	0.32
	0.15	0.42	0.67	0.67	0.69	0.13	0.29	0.29	0.43
	0.20	0.46	0.71	0.71	0.75	0.15	0.34	0.34	0.50
(40, 230)	0.00	0.05	0.06	0.06	0.06	0.05	0.06	0.06	0.06
	0.05	0.55	0.72	0.72	0.72	0.18	0.30	0.30	0.33
	0.10	0.71	0.89	0.89	0.89	0.23	0.47	0.47	0.56
	0.15	0.75	0.95	0.95	0.95	0.29	0.57	0.57	0.73
	0.20	0.79	0.97	0.97	0.97	0.34	0.66	0.66	0.82

We first display the kernel density estimates of the asymptotic null distributions of the standardized test statistic $T_{n,p}$ in Figure 1, which can be well approximated by a standard normal distribution. This confirms the theoretical results in Theorems 3.2 and 3.4. Then we summarize the empirical sizes and powers of EB, ZC, New_0 and New_{RCV} tests for $\varepsilon \sim N(0, 1)$ and $t(5)/\sqrt{5/3}$ based on 1000 simulations in Tables 1 and 2, respectively. It can been observed that the empirical sizes

TABLE 2
Empirical Powers of EB, ZC, New$_0$ and New$_{\mathrm{RCV}}$ tests at the significant level 5% when $\varepsilon \sim t(5)/\sqrt{5/3}$

		Nonsparse case				Sparse case			
$(\boldsymbol{n}, \boldsymbol{p})$	$\\|\boldsymbol{\beta}\\|^2$	**EB**	**ZC**	**New$_0$**	**New$_{\mathrm{RCV}}$**	**EB**	**ZC**	**New$_0$**	**New$_{\mathrm{RCV}}$**
(1) $\mathbf{X}_i \sim N_p(\mu, \Sigma)$									
(20, 178)	0.00	0.05	0.06	0.06	0.06	0.05	0.06	0.06	0.06
	0.05	0.24	0.38	0.38	0.42	0.09	0.18	0.18	0.22
	0.10	0.35	0.57	0.57	0.58	0.12	0.27	0.27	0.34
	0.15	0.39	0.66	0.67	0.67	0.14	0.34	0.33	0.45
	0.20	0.42	0.73	0.73	0.72	0.16	0.37	0.36	0.53
(40, 230)	0.00	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	0.05	0.59	0.73	0.73	0.71	0.19	0.31	0.30	0.35
	0.10	0.69	0.85	0.85	0.89	0.26	0.50	0.50	0.61
	0.15	0.79	0.97	0.97	0.97	0.32	0.62	0.62	0.75
	0.20	0.85	0.98	0.97	0.98	0.36	0.67	0.67	0.83
(2) $\mathbf{X}_i \sim \sqrt{1-2/q}\, t_p(\mu, \Sigma, q)$									
(20, 178)	0.00	0.05	0.06	0.06	0.06	0.05	0.06	0.06	0.06
	0.05	0.25	0.35	0.37	0.41	0.13	0.17	0.18	0.22
	0.10	0.33	0.48	0.50	0.54	0.16	0.24	0.26	0.36
	0.15	0.37	0.56	0.57	0.63	0.18	0.30	0.32	0.42
	0.20	0.42	0.61	0.63	0.66	0.21	0.35	0.37	0.51
(40, 230)	0.00	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
	0.05	0.52	0.64	0.65	0.65	0.22	0.28	0.29	0.31
	0.10	0.63	0.81	0.81	0.83	0.30	0.44	0.44	0.52
	0.15	0.71	0.87	0.88	0.90	0.36	0.52	0.53	0.66
	0.20	0.75	0.92	0.92	0.93	0.42	0.62	0.63	0.78
(3) $\mathbf{X}_i \sim \Sigma^{1/2} Z$									
(20, 178)	0.00	0.05	0.06	0.06	0.06	0.05	0.06	0.06	0.06
	0.05	0.20	0.42	0.42	0.44	0.10	0.18	0.18	0.21
	0.10	0.33	0.59	0.59	0.60	0.14	0.27	0.27	0.34
	0.15	0.41	0.67	0.68	0.69	0.16	0.32	0.32	0.44
	0.20	0.45	0.73	0.73	0.74	0.19	0.37	0.36	0.51
(40, 230)	0.00	0.04	0.05	0.05	0.06	0.04	0.05	0.05	0.06
	0.05	0.57	0.73	0.73	0.73	0.17	0.29	0.28	0.32
	0.10	0.69	0.89	0.89	0.88	0.26	0.49	0.49	0.61
	0.15	0.77	0.94	0.94	0.94	0.33	0.62	0.62	0.76
	0.20	0.83	0.97	0.96	0.96	0.34	0.70	0.70	0.84

of all tests are quite reasonably around 0.05 when H_0 is true, that is, $\|\boldsymbol{\beta}\| = 0$. The three panels of Tables 1 and 2 show that the New$_0$ test and the ZC test share the very similar power performance. This also confirms the previous theoretical results in Section 3 that the two tests shares the identical asymptotic local power. However, thanks to the efficient RCV estimator of σ^2, the New$_{\mathrm{RCV}}$ test achieves

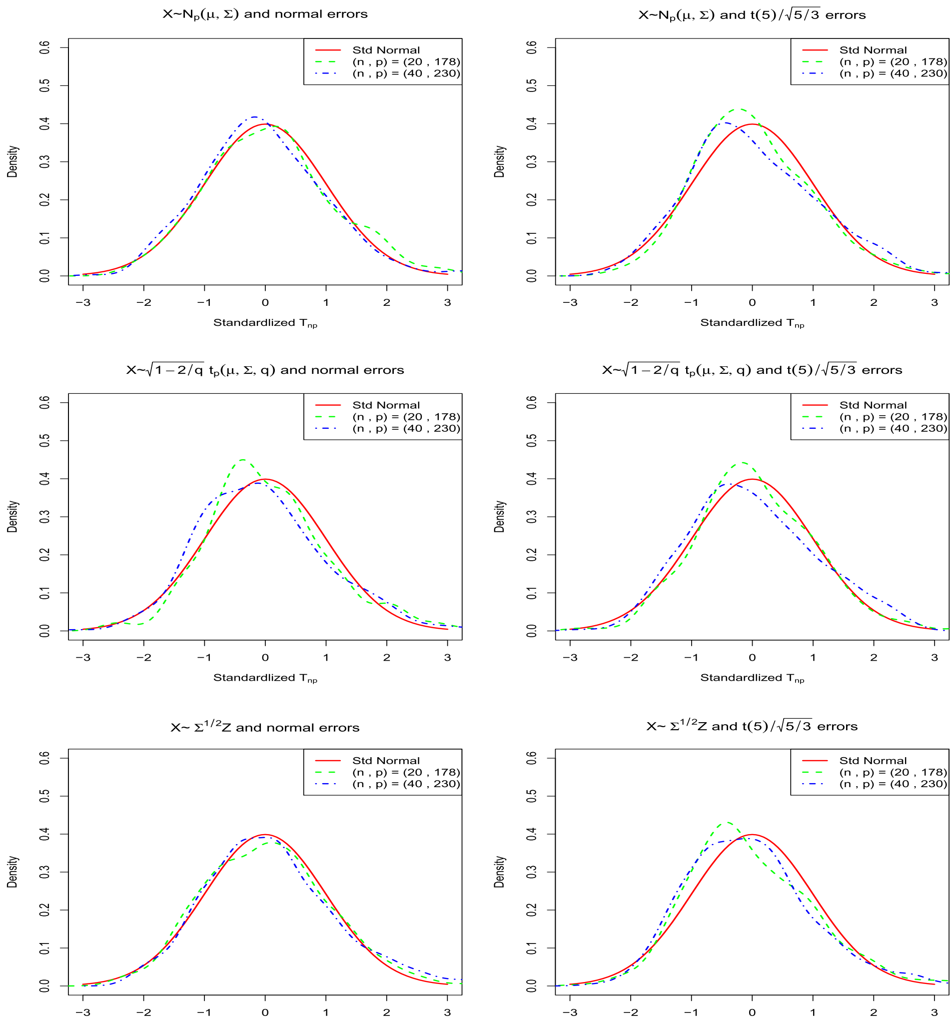

FIG. 1. *The asymptotic null distributions of the standardized UT test statistic* $T_{n,p}$.

higher empirical powers than other tests, especially for the sparse case. For example, when $X \sim \Sigma^{1/2}Z$, $\varepsilon \sim N(0, 1)$, $(n, p) = (40, 230)$ and $\|\boldsymbol{\beta}\|^2 = 0.20$, the empirical power of the New_{RCV} test is 82%, which is much higher than all other tests for the sparse case.

In addition, Table 3 reports the running time of the New_0 test and the ZC test under different scenarios of (n, p). It is shown that the New_0 test can substantially reduce computational complexity compared with the ZC test. This is because the new test statistic is based on an estimated U-statistic of order two while the ZC test statistic is a U-statistic of order four. Note that the running time of the New_0 test

TABLE 3
Computational time comparison between the New_0 *and ZC tests* (*in seconds*)

(n, p)	(20, 178)	(40, 230)	(60, 320)	(60, 470)	(100, 700)	(120, 1036)	(140, 1514)
ZC	3	7	19	55	138	344	822
New_0	1	3	8	23	56	136	313

NOTE: The time is calculated based on 1000 simulations.

and the New_{RCV} test are similar in practice because of computational efficiency of the sure independence screening.

5. Real data analysis. We consider a real data set of 24 six-month-old Yolkshire gilts which was also analyzed by [24] and [14]. The gilts were genetyped by the melanocortin-4 receptor gene, 12 of them with D298 and the rest with N298. Two diet treatments were assigned to the 12 gilts in each genetype randomly. One was feeding without restrictions; the other was fasting. One can refer to [14] for more details about the experiment. The gene expression levels were measured for 24,123 genes in liver tissues, which could be classified into different sets according to their biological functions (The Gene Ontology Consortium 2000). We call them GO terms for short. There are 6176 GO terms whose dimension ranged from 1 to 5158 and many of the GO terms share common genes. The response variable is the triiodothyronine (T_3) measurement that is a vital thyroid hormone to increase the metabolic rate, protein synthesis and stimulates breakdown of cholesterol. The values of T_3 were obtained in the blood of each gilt. The goal is to detect the GO terms, which are significantly correlated with T_3.

We consider the following five models to study the association between each GO term and the response T_3:

$$\begin{aligned}
&\text{Model 0}: \ Y_{..k} = \alpha^g + \mathbf{X}_{..k}^{g\mathrm{T}}\boldsymbol{\beta}^g + \varepsilon_{..k}^g, && k = 1, 2, \ldots, 24;\\
&\text{Model 1}: \ Y_{1.k} = \alpha^g + \mathbf{X}_{1.k}^{g\mathrm{T}}\boldsymbol{\beta}^g + \varepsilon_{1.k}^g, && k = 1, 2, \ldots, 12;\\
&\text{Model 2}: \ Y_{2.k} = \alpha^g + \mathbf{X}_{2.k}^{g\mathrm{T}}\boldsymbol{\beta}^g + \varepsilon_{2.k}^g, && k = 1, 2, \ldots, 12;\\
&\text{Model 3}: \ Y_{.1k} = \alpha^g + \mathbf{X}_{.1k}^{g\mathrm{T}}\boldsymbol{\beta}^g + \varepsilon_{.1k}^g, && k = 1, 2, \ldots, 12;\\
&\text{Model 4}: \ Y_{.2k} = \alpha^g + \mathbf{X}_{.2k}^{g\mathrm{T}}\boldsymbol{\beta}^g + \varepsilon_{.2k}^g, && k = 1, 2, \ldots, 12,
\end{aligned}$$

where Y_{ijk} and $\mathbf{X}_{ijk}^g$ denote the T_3 measurement and the gene expression levels of the gth GO term for the kth gilt in the jth treatment with the ith genotype, $i = 1$ when the genotype is D298 and $i = 2$ for N298, $j = 1$ for the fasting treatment group and otherwise $j = 2$ and $g = 1, \ldots, 6176$. The dot sign here denotes that the corresponding index is ignored. For example, $Y_{1.k}$ denotes the T_3 value for the kth gilt with the first genotype (D298) where the treatment index is ignored.

TABLE 4
The numbers of the significant GO terms

	EB	**ZC**	**New$_0$**	**New$_{RCV}$**
Model 0	110	176	177	279
Model 1	62	137	137	147
Model 2	35	109	109	120
Model 3	83	141	138	162
Model 4	18	92	87	96

For each GO term, we are interested in testing for

$$H_0 : \boldsymbol{\beta}^g = 0 \quad \text{versus} \quad H_1 : \boldsymbol{\beta}^g \neq 0. \tag{5.1}$$

We applied the EB test, the ZC test, the proposed New$_0$ and New$_{RCV}$ tests to test (5.1) for each g. By controlling the false discover rate (FDR) for the p-value of the tests at 1%, Table 4 summaries the numbers of GO terms which are declared statistically significant. The EB test detects less significant GO terms than the other three tests while the numbers of significant GO terms identified by the ZC and New$_0$ tests are quit similar. In general, the New$_{RCV}$ test select more GO terms as significant.

We list in Table 5 the significant GO terms, labeled by ticks "$\surd$," under the five models. GO:0007528 is the only significant gene set under all five models detected by the three tests, ZC, New$_0$ and New$_{RCV}$. There are six GO terms identified as

TABLE 5
The significant GO terms, labeled by "$\surd$," under the five models and the corresponding number of genes

GO term	**Model 0**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**No. of genes**	**Satisfied test(s)**
GO:0007528	$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	8	ZC/New$_0$/New$_{RCV}$
GO:0043204	$\surd$	$\surd$	$\surd$			5	New$_{RCV}$
GO:0032012	$\surd$	$\surd$	$\surd$			12	EB/ZC/New$_0$/New$_{RCV}$
GO:0005840	$\surd$	$\surd$	$\surd$			225	ZC/New$_0$/New$_{RCV}$
GO:0006108	$\surd$	$\surd$	$\surd$			6	ZC/New$_0$/New$_{RCV}$
GO:0009952	$\surd$	$\surd$	$\surd$			47	ZC/New$_0$/New$_{RCV}$
GO:0051287	$\surd$	$\surd$	$\surd$			31	ZC/New$_0$/New$_{RCV}$
GO:0004115	$\surd$			$\surd$	$\surd$	7	ZC/New$_0$/New$_{RCV}$
GO:0005086	$\surd$			$\surd$	$\surd$	14	ZC/New$_0$/New$_{RCV}$
GO:0005677	$\surd$			$\surd$	$\surd$	5	ZC/New$_0$/New$_{RCV}$
GO:0006342	$\surd$			$\surd$	$\surd$	5	ZC/New$_0$/New$_{RCV}$
GO:0009187	$\surd$			$\surd$	$\surd$	12	ZC/New$_0$/New$_{RCV}$
GO:0017136	$\surd$			$\surd$	$\surd$	5	ZC/New$_0$/New$_{RCV}$

significant by at least one test under Models 0, 1 and 2. In particular, GO:0043204 is only detected by our New_{RCV}, and GO:0032012 is selected by all four tests and the other four GO terms are significant under the ZC, New_0 and New_{RCV} tests. Under Models 0, 3 and 4, our tests and the ZC test detect another six significant GO terms. However, the EB test fails to find any of them. In conclusion, the EB test detect quite few GO terms which coincides with the simulation results that it tends to have relatively low powers. The proposed New_{RCV} test has higher powers to detect more significant gene sets in this real data analysis.

6. Discussion. In this paper, we proposed a new test for high-dimensional linear regression coefficients based on the estimated U-statistics of order two and refitted cross-validation (RCV) error variance estimation. The two different distribution assumptions on the covariates, the pseudo-independence assumption and the elliptical distribution assumption were considered to study the theoretical properties. The limiting null distributions of the proposed test statistic are normal under the null hypothesis and the local alternative condition. Moveover, we demonstrated that the RCV variance estimation could substantially enhance empirical powers, especially for the sparse case.

In the construction of the test statistic (2.4), we have used the centralization and the associated bias correction. Another idea is to apply the extended cross-data-matrix (ECDM) methodology in [22]. The ECDM methodology considers the combination of cross data matrices to construct an unbiased estimator efficiently. This idea as well as the associated theoretical properties are worth a future study.

Although we are interested in testing the whole regression parameter $\boldsymbol{\beta}$, similar to [21], our approach can be also extended to the case of testing a part of the regression parameter $\boldsymbol{\beta}_1$, in which the $\boldsymbol{\beta}_2$ could be replaced by a suitable estimator, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^{\mathrm{T}}, \boldsymbol{\beta}_2^{\mathrm{T}})^{\mathrm{T}}$. On the other hand, by following the idea of the nonparametric test based on the spatial sign transformation of the data in [19], it is also of interest to develop a nonparametric test for testing the regression coefficients in high-dimensional linear models.

APPENDIX: TECHNICAL PROOFS

For simplicity of presentation, we introduce some notation: $A_0 = \Gamma^T\Gamma$, $A_1 = \Gamma^T\boldsymbol{\beta}\boldsymbol{\beta}^T\Gamma$, $A_2 = \Gamma^T\Sigma\boldsymbol{\beta}\boldsymbol{\beta}^T\Sigma\Gamma$, $A_3 = \Gamma^T\Sigma\Gamma$, $B_i = \boldsymbol{\beta}^T\Sigma^i\boldsymbol{\beta}$, for $i = 1, 2, 3$. We assume, without loss of generality, that $\alpha = 0$ and $\mu = 0$ in the rest of the article. First, we list some important lemmas in order to simplify the calculation.

LEMMA 1. *Suppose conditions* (C2) *and* (C3) *hold*; *it can be shown that*

$$E(\boldsymbol{Z}_1\boldsymbol{Z}_1^T M\boldsymbol{Z}_1\boldsymbol{Z}_1^T) = 2M + \triangle\,\mathrm{diag}(M) + \mathrm{tr}(M)I_m, \tag{A.1}$$

$$E[(\boldsymbol{\beta}^T\boldsymbol{X}_1\boldsymbol{X}_1^T\boldsymbol{X}_2\boldsymbol{X}_2^T\boldsymbol{\beta})^2] = o(\mathrm{tr}(\Sigma^2)), \tag{A.2}$$

where M is a $m \times m$ symmetric matrix.

PROOF. Denote $M = (m_{ij})$. The $(k,l)'$th element of $E(\mathbf{Z_1}\mathbf{Z_1}^T M\mathbf{Z_1}\mathbf{Z_1}^T)$ is

$$E(\mathbf{Z_1}\mathbf{Z_1}^T M\mathbf{Z_1}\mathbf{Z_1}^T)_{(k,l)} = \sum_{i=1}^m\sum_{j=1}^m m_{ij}\mathbf{Z_1}^{(i)}\mathbf{Z_1}^{(j)}\mathbf{Z_1}^{(k)}\mathbf{Z_1}^{(l)}$$
$$= \begin{cases} \sum_{i=1}^n m_{ii} + (2+\Delta)m_{kk} & \text{if } k=l,\\ m_{kl}+m_{lk} & \text{if } k\neq l.\end{cases}$$

It means that $E(\mathbf{Z_1}\mathbf{Z_1}^T M\mathbf{Z_1}\mathbf{Z_1}^T) = 2M + \Delta\,\mathrm{diag}(M) + \mathrm{tr}(M)I_m$. By (A.1), we have that

$$E[(\boldsymbol{\beta}^T\boldsymbol{X_1}\boldsymbol{X_1}^T\boldsymbol{X_2}\boldsymbol{X_2}^T\boldsymbol{\beta})^2]$$
$$= 4B_1B_3 + 4B_2^2 + B_1^2\,\mathrm{tr}(\Sigma^2)$$
$$+ 4\Delta\,\mathrm{tr}\{A_1\circ A_2\} + 2\Delta B_1\,\mathrm{tr}(A_1\circ A_3) + \Delta^2\,\mathrm{tr}\{(A_0\,\mathrm{diag}(A_1))^2\}.$$

Under condition (C2), we have $E[(\boldsymbol{\beta}^T\boldsymbol{X_1}\boldsymbol{X_1}^T\boldsymbol{X_2}\boldsymbol{X_2}^T\boldsymbol{\beta})^2] = o(\mathrm{tr}(\Sigma^2))$. □

LEMMA 2. *Let* $\mathbf{U} = (U_1,\dots,U_p)^T$ *be a random vector uniformly distributed on the unit sphere in* $\mathbb{R}^p$. *Then* $E(\mathbf{U}) = 0$, $\mathrm{Var}(\mathbf{U}) = p^{-1}I_p$, $E(U_j^4) = \frac{3}{p(p+2)}$, $\forall j$, *and* $E(U_j^2U_k^2) = \frac{1}{p(p+2)}$ *for* $j\neq k$.

PROOF. See Section 3.1 of Fang, Kotz and Ng [9]. □

LEMMA 3. *Suppose condition* (C4) *holds*; *it follows that*

$$E(\boldsymbol{U_1}\boldsymbol{U_1}^T M\boldsymbol{U_1}\boldsymbol{U_1}^T) = \frac{1}{p(p+2)}(2M + \mathrm{tr}(M)I_p), \tag{A.3}$$

$$E(\boldsymbol{X_1}^T\boldsymbol{X_1}Y_1^2) = \frac{2E(R^4)}{p(p+2)}\boldsymbol{\beta}^T\Sigma^2\boldsymbol{\beta} + \frac{E(R^4)}{p(p+2)}\mathrm{tr}(\Sigma)\boldsymbol{\beta}^T\Sigma\boldsymbol{\beta} + \sigma^2\,\mathrm{tr}(\Sigma), \tag{A.4}$$

where M *is a* $m\times m$ *symmetric matrix.*

PROOF. We here only prove equation (A.3). Let $M = (m_{ij})_{p\times p}$, U_{1i} is the ith element of $\boldsymbol{U_1}$, b_{ij} denotes the (i,j)th element of $E(\boldsymbol{U_1}\boldsymbol{U_1}^T M\boldsymbol{U_1}\boldsymbol{U_1}^T)$, then we have

$$b_{ij} = E\sum_{k=1}^p\sum_{l=1}^p m_{kl}U_{1i}U_{1j}U_{1k}U_{1l} = \begin{cases}\dfrac{2m_{ii}+\sum_{k=1}^p m_{kk}}{p(p+2)} & \text{if } i=j,\\[2mm] \dfrac{m_{ij}+m_{ji}}{p(p+2)} & \text{if } i\neq j,\end{cases}$$

by applying Lemma 2. Thus, we have $E(\boldsymbol{U_1}\boldsymbol{U_1}^T M\boldsymbol{U_1}\boldsymbol{U_1}^T) = \frac{1}{p(p+2)}(2M + \operatorname{tr}(M)I_p)$, which completes the proof. $\square$

LEMMA 4. *If* (C1) *and* (C3) *hold, or* (C1) *and* (C4) *hold, then it can be shown that*

$$E\big[(\boldsymbol{X_1}^T\boldsymbol{X_2})^4\big] = o\big(n\operatorname{tr}^2(\Sigma^2)\big), \tag{A.5}$$

$$E\big[(\boldsymbol{X_1}^T\Sigma\boldsymbol{X_1})^2\big] = o\big(n\operatorname{tr}^2(\Sigma^2)\big). \tag{A.6}$$

PROOF. Write $\mathbf{X} = \Gamma\mathbf{Z}$ and $\Gamma^T\Gamma = (\nu_{ij})_{p\times p}$. Then we have

$$\begin{aligned}
&E\{(\boldsymbol{Z_1}^T\Gamma^T\Gamma\boldsymbol{Z_2})^4\}\\
&\quad= E\Bigg[\Bigg(\sum_{i=1}^p\sum_{j=1}^p \nu_{ij}Z_{1i}Z_{2j}\Bigg)^4\Bigg]\\
&\quad= \sum_{i=1}^p\sum_{j=1}^p \nu_{ij}^4 E^2(Z_{1i}^4)\\
&\qquad+ 3\sum_{1\le i\le p, 1\le j_1\ne j_2\le p} \nu_{ij_1}^2\nu_{ij_2}^2 E(Z_{1i}^4)E(Z_{2j_1}^2 Z_{2j_2}^2)\\
&\qquad+ 3\sum_{1\le j\le p, 1\le i_1\ne i_2\le p} \nu_{i_1j}^2\nu_{i_2j}^2 E(Z_{2j}^4)E(Z_{1i_1}^2 Z_{1i_2}^2)\\
&\qquad+ O(1)\sum_{1\le i_1\ne i_2\le p}\sum_{1\le j_1\ne j_2\le p} (\nu_{i_1j_1}^2\nu_{i_2j_2}^2\\
&\qquad+ \nu_{i_1j_1}\nu_{i_1j_2}\nu_{i_2j_1}\nu_{i_2j_2})E(Z_{1i_1}^2 Z_{1i_2}^2)E(Z_{2j_1}^2 Z_{2j_2}^2).
\end{aligned}$$

Under (C3) we obtain

$$\begin{aligned}
&E\big[(\boldsymbol{Z_1}^T\Gamma^T\Gamma\boldsymbol{Z_2})^4\big]\\
&\quad= (3+\Delta)^2\sum_{i=1}^p\sum_{j=1}^p \nu_{ij}^4 + 3(3+\Delta)\sum_{1\le i\le p, 1\le j_1\ne j_2\le p} \nu_{ij_1}^2\nu_{ij_2}^2\\
&\qquad+ 3(3+\Delta)\sum_{1\le j\le p, 1\le i_1\ne i_2\le p} \nu_{i_1j}^2\nu_{i_2j}^2\\
&\qquad+ O(1)\sum_{1\le i_1\ne i_2\le p}\sum_{1\le j_1\ne j_2\le p} (\nu_{i_1j_1}^2\nu_{i_2j_2}^2 + \nu_{i_1j_1}\nu_{i_1j_2}\nu_{i_2j_1}\nu_{i_2j_2})\\
&\quad\le O(1)\operatorname{tr}^2(\Gamma^T\Gamma\Gamma^T\Gamma) + O(1)\operatorname{tr}\big((\Gamma^T\Gamma)^4\big)\\
&\quad\le O(1)\operatorname{tr}^2(\Sigma^2),
\end{aligned}$$

by condition (C1) and $\mathrm{tr}((\Gamma^T\Gamma)^4) \le \mathrm{tr}^2((\Gamma^T\Gamma)^2)$, and noticing that

$$\max\left\{\sum_{i=1}^p\sum_{j=1}^p \nu_{ij}^4, \sum_{1\le i\le p, 1\le j_1\ne j_2\le p} \nu_{ij_1}^2\nu_{ij_2}^2, \sum_{1\le i_1\ne i_2\le p}\sum_{1\le j_1\ne j_2\le p} \nu_{i_1j_1}^2\nu_{i_2j_2}^2\right\}$$
$$\le \left(\sum_{i=1}^p\sum_{j=1}^p \nu_{ij}^2\right)^2$$
$$= \mathrm{tr}^2(\Sigma^2)$$

and

$$\sum_{1\le i_1\ne i_2\le p}\sum_{1\le j_1\ne j_2\le p} \nu_{i_1j_1}\nu_{i_1j_2}\nu_{i_2j_1}\nu_{i_2j_2} \le \sum_{1\le i_1\ne i_2\le p} \nu_{i_1i_2}^{(2)}\nu_{i_1i_2}^{(2)}$$
$$\le \sum_{i=1}^p \nu_{i_1i_1}^{(4)}$$
$$= \mathrm{tr}(\Sigma^4),$$

where $(\Gamma^T\Gamma)^2 = (\nu_{ij}^{(2)})_{p\times p}$ and $(\Gamma^T\Gamma)^4 = (\nu_{ij}^{(4)})_{p\times p}$. Then (A.5) is proved. Under (C4),

$$E[(\boldsymbol{Z}_1^T\Gamma^T\Gamma\boldsymbol{Z}_2)^4]$$
$$= 9\left(\frac{E(R^4)}{p(p+2)}\right)^2\sum_{i=1}^p\sum_{j=1}^p \nu_{ij}^4 + 9\left(\frac{E(R^4)}{p(p+2)}\right)^2\sum_{1\le i\le p, 1\le j_1\ne j_2\le p} \nu_{ij_1}^2\nu_{ij_2}^2$$
$$+ 9\left(\frac{E(R^4)}{p(p+2)}\right)^2\sum_{1\le j\le p, 1\le i_1\ne i_2\le p} \nu_{i_1j}^2\nu_{i_2j}^2$$
$$+ O(1)\left(\frac{E(R^4)}{p(p+2)}\right)^2$$
$$\times \sum_{1\le i_1\ne i_2\le p}\sum_{1\le j_1\ne j_2\le p}(\nu_{i_1j_1}^2\nu_{i_2j_2}^2 + \nu_{i_1j_1}\nu_{i_1j_2}\nu_{i_2j_1}\nu_{i_2j_2})$$
$$\le O(1)\,\mathrm{tr}^2(\Gamma^T\Gamma\Gamma^T\Gamma) + O(1)\,\mathrm{tr}((\Gamma^T\Gamma)^4)$$
$$\le O(1)\,\mathrm{tr}^2(\Sigma^2).$$

Under (C1), we obtain that $E[(\boldsymbol{Z}_1^T\Gamma^T\Gamma\boldsymbol{Z}_2)^4] = o(n\,\mathrm{tr}(\Sigma^2))$. For equation (A.6),

$$E[(\boldsymbol{X}_1^T\Sigma\boldsymbol{X}_1)^2] = E[(\boldsymbol{Z}_1^T\Gamma^T\Gamma E(\boldsymbol{Z}_2\boldsymbol{Z}_2^T)\Gamma^T\Gamma\boldsymbol{Z}_1)^2]$$
$$= E[E^2(\boldsymbol{Z}_1^T\Gamma^T\Gamma\boldsymbol{Z}_2\boldsymbol{Z}_2^T\Gamma^T\Gamma\boldsymbol{Z}_1 \mid \boldsymbol{Z}_1)]$$
$$\le E\{E[(\boldsymbol{Z}_1^T\Gamma^T\Gamma\boldsymbol{Z}_2\boldsymbol{Z}_2^T\Gamma^T\Gamma\boldsymbol{Z}_1)^2 \mid \boldsymbol{Z}_1]\}$$

$$= E\big[\big(\mathbf{Z_1}^T \Gamma^T \Gamma \mathbf{Z_2}\big)^4\big] = E\big[\big(\mathbf{X_1}^T \mathbf{X_2}\big)^4\big].$$

Hence, (A.6) follows from (A.5). This completes the proof. □

In order to obtain the expectation and the asymptotic distribution of the statistic, it is needed to reformulate $\triangle_{i,j}$ as follows:

$$\begin{aligned}
&\frac{n}{n-2}\triangle_{i,j}(\mathbf{X}) \\
&\quad = \left(1-\frac{1}{n}\right)\boldsymbol{X_i}^T\boldsymbol{X_j} - \frac{1}{2n}\big(\boldsymbol{X_i}^T\boldsymbol{X_i} + \boldsymbol{X_j}^T\boldsymbol{X_j} - 2E\big(\boldsymbol{X_1}^T\boldsymbol{X_1}\big)\big) \\
&\qquad - \left(1-\frac{2}{n}\right)\overline{\boldsymbol{X}}_{\boldsymbol{i,j}}^T(\boldsymbol{X_i}+\boldsymbol{X_j}) \qquad \text{(A.7)} \\
&\qquad + \left(1-\frac{2}{n}\right)\left[\overline{\boldsymbol{X}}_{\boldsymbol{i,j}}^T\overline{\boldsymbol{X}}_{\boldsymbol{i,j}} - \frac{E(\boldsymbol{X_1}^T\boldsymbol{X_1})}{n-2}\right] \\
&\quad =: R_{ij1} + R_{ij2} + R_{ij3} + R_{ij4},
\end{aligned}$$

$$\begin{aligned}
&\frac{n}{n-2}\triangle_{i,j}(\mathbf{Y}) \\
&\quad = \left(1-\frac{1}{n}\right)Y_iY_j - \frac{1}{2n}\big(Y_i^2 + Y_j^2 - 2E\big(Y_1^2\big)\big) \qquad \text{(A.8)} \\
&\qquad - \left(1-\frac{2}{n}\right)\overline{Y}_{i,j}(Y_i+Y_j) + \left(1-\frac{2}{n}\right)\left[\overline{Y}_{i,j}\overline{Y}_{i,j} - \frac{E(Y_1^2)}{n-2}\right] \\
&\quad =: S_{ij1} + S_{ij2} + S_{ij3} + S_{ij4},
\end{aligned}$$

where $\overline{\boldsymbol{X}}_{\boldsymbol{i,j}}$ and $\overline{Y}_{i,j}$ are the average of $\boldsymbol{X}'_{\boldsymbol{k}}s$ and $Y'_k s$ with deleting the ith and jth samples respectively, that is, $\overline{\boldsymbol{X}}_{\boldsymbol{i,j}} = \frac{1}{n-2}\sum_{-(i,j)}\boldsymbol{X_k}$ and $\overline{Y}_{i,j} = \frac{1}{n-2}\sum_{-(i,j)}Y_k$. It is obvious that under (C3) or (C4), $E[R_{ijk}S_{ijl}] = 0$, for $k \neq l$, and $k, l = 1, 2, 3, 4$.

PROOF OF THEOREM 3.1. (i) Using Lemma 1, we have under (C3) that

$$\begin{aligned}
E[R_{ij1}S_{ij1}] &= \frac{(n-1)^2}{n^2}\boldsymbol{\beta}^T\Sigma^2\boldsymbol{\beta}, \\
E[R_{ij2}S_{ij2}] &= \frac{1}{n^2}\boldsymbol{\beta}^T\Sigma^2\boldsymbol{\beta} + \frac{1}{2n^2}\Delta\big(\boldsymbol{\beta}^T\Gamma\,\mathrm{diag}\big(\Gamma^T\Gamma\big)\Gamma^T\boldsymbol{\beta}\big), \\
E[R_{ij3}S_{ij3}] &= \frac{2(n-2)}{n^2}\boldsymbol{\beta}^T\Sigma^2\boldsymbol{\beta}, \\
E[R_{ij4}S_{ij4}] &= \frac{2}{n^2}\boldsymbol{\beta}^T\Sigma^2\boldsymbol{\beta} + \frac{1}{n^2(n-2)}\Delta\big(\boldsymbol{\beta}^T\Gamma\,\mathrm{diag}\big(\Gamma^T\Gamma\big)\Gamma^T\boldsymbol{\beta}\big).
\end{aligned}$$

Put summation of the four terms, then we prove (i).

(ii) The reconstruction of the proposed statistic could be obtained by the proof of Theorem 3.2. □

PROOF OF THEOREM 3.3. (i) By (A.7) and (A.8), like the proof of Theorem 3.1, under (C4):

$$\begin{aligned}
E[R_{ij1}S_{ij1}] &= \frac{(n-1)^2}{n^2}\boldsymbol{\beta}^T\Sigma^2\boldsymbol{\beta}, \\
E[R_{ij2}S_{ij2}] &= \frac{E(R^4)}{n^2 p(p+2)}\boldsymbol{\beta}^T\Sigma^2\boldsymbol{\beta} \\
&\quad + \frac{\operatorname{tr}(\Sigma)\boldsymbol{\beta}^T\Sigma\boldsymbol{\beta}}{2n^2}\left(\frac{E(R^4)}{p(p+2)} - 1\right), \\
E[R_{ij3}S_{ij3}] &= \frac{2(n-2)}{n^2}\boldsymbol{\beta}^T\Sigma^2\boldsymbol{\beta}, \\
E[R_{ij4}S_{ij4}] &= \frac{2\boldsymbol{\beta}^T\Sigma^2\boldsymbol{\beta}}{n^2(n-2)}\left(\frac{E(R^4)}{p(p+2)} + (n-3)\right) \\
&\quad + \frac{\operatorname{tr}(\Sigma)\boldsymbol{\beta}^T\Sigma\boldsymbol{\beta}}{n^2(n-2)}\left(\frac{E(R^4)}{p(p+2)} - 1\right).
\end{aligned}$$

Put summation of the four terms, then we prove (i).

(ii) The reconstruction of the proposed test statistic could be obtained by the proof of Theorem 3.4. □

PROOF OF THEOREMS 3.2 AND 3.4. Applying Lemmas 1–4, we prove Theorems 3.2 and 3.4. Because the whole proof of the two theorems are quite similar, we only prove Theorem 3.4 and just list the sketch proof of Theorem 3.2 below in order to save the page capacity of the paper, and the details proof of Theorem 3.2 can be put in the Supplemental Materials [5].

By the fact that

$$\begin{aligned}
&n\bigl(E(T_{n,p}) - \boldsymbol{\beta}^{\mathrm{T}}\Sigma^2\boldsymbol{\beta}\bigr)/\sqrt{\operatorname{tr}(\Sigma^2)} \\
&\quad \le O(n^{-1})\left|\frac{E(R^4) - p^2 - 2p}{p(p+2)}\right|\left(\frac{3}{2}\operatorname{tr}(\Sigma)\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta}\right)\Big/\sqrt{\operatorname{tr}(\Sigma^2)} \\
&\quad \le O(n^{-1})\frac{p^2}{p(p+2)}\lambda_{\max}(\Sigma)\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta}/\sqrt{\operatorname{tr}(\Sigma^2)} \\
&\quad \le O(n^{-1})[\operatorname{tr}(\Sigma^4)]^{1/4}\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta}/\sqrt{\operatorname{tr}(\Sigma^2)} \\
&\quad = o(n^{-1}\boldsymbol{\beta}^{\mathrm{T}}\Sigma\boldsymbol{\beta}) = o(n^{-1}),
\end{aligned}$$

hence we just suffice to prove that

$$\frac{n(T_{n,p} - E(T_{n,p}))}{\sigma^2\sqrt{2\,\mathrm{tr}(\Sigma^2)}} \xrightarrow{D} N(0,1).$$

Write $T_{n,p}^{(k,l)} = n\binom{n}{2}^{-1}\sum_{i>j}(R_{ijk}S_{ijl} - E(R_{ijk}S_{ijl}))$, where $k,l = 1,2,3,4$. By (A.7)–(A.8), we have $n(T_{n,p} - ET_{n,p}) = \sum_k\sum_l T_{n,p}^{(k,l)}$. First, we may rewrite

$$\begin{aligned} T_{n,p}^{(1,1)} &= n\left(1-\frac{1}{n}\right)^2\binom{n}{2}^{-1}\sum_{i=2}^{n}\sum_{j=1}^{i-1}(\boldsymbol{\beta}^T\boldsymbol{X_i}\boldsymbol{X_i}^T\boldsymbol{X_j}\boldsymbol{X_j}^T\boldsymbol{\beta} - \boldsymbol{\beta}^T\Sigma^2\boldsymbol{\beta}) \\ &\quad + (\varepsilon_j\boldsymbol{\beta}^T\boldsymbol{X_i}\boldsymbol{X_i}^T\boldsymbol{X_j} + \varepsilon_i\boldsymbol{\beta}^T\boldsymbol{X_j}\boldsymbol{X_j}^T\boldsymbol{X_i}) + (\varepsilon_i\varepsilon_j\boldsymbol{X_i}^T\boldsymbol{X_j}) \\ &=: T_{n,p}^{(1,1)}(1) + T_{n,p}^{(1,1)}(2) + T_{n,p}^{(1,1)}(3). \end{aligned}$$

We shall prove

$$\frac{T_{n,p}^{(1,1)} - E(T_{n,p}^{(1,1)})}{\sigma^2\sqrt{2\,\mathrm{tr}(\Sigma^2)}} = \frac{T_{n,p}^{(1,1)}(3)}{\sigma^2\sqrt{2\,\mathrm{tr}(\Sigma^2)}} + o_P(1). \tag{A.9}$$

Denote $T_{n,p}^{(1,1)}(k) = n(1-1/n)^2\binom{n}{2}^{-1}\sum_{i=2}^{n}Q_i^{(1,1)}(k)$, $k = 1,2$. Then it follows from Lemma 2 and Lemma 3 that

$$\begin{aligned} &E[Q_i^{(1,1)}(1)Q_i^{(1,1)}(1)] \\ &\quad = (i-1)\left\{\frac{(ER^4)^2}{p^2(p+2)^2}(4B_2^2 + 4B_1B_3 + B_1^2\,\mathrm{tr}(\Sigma^2)) - B_2^2\right\} \\ &\qquad + (i-1)(i-2)\left\{\frac{E(R^4)}{p(p+2)}(2B_2^2 + B_1B_3) - B_2^2\right\}, \\ &E[Q_i^{(1,1)}(1)Q_j^{(1,1)}(1)] \\ &\quad = [(i-1)\wedge(j-1)]\left\{\frac{E(R^4)}{p(p+2)}(2B_2^2 + B_1B_3) - B_2^2\right\}, \end{aligned}$$

for $i \neq j$. Similarly, by simple calculation, we have $E[Q_i^{(1,1)}(2)Q_i^{(1,1)}(2)] = (i-1)\{\frac{E(R^4)}{p(p+2)}(4\sigma^2B_3 + 2\sigma^2B_1\,\mathrm{tr}(\Sigma^2))\} + (i-1)(i-2)\sigma^2B_3$ and $E[Q_i^{(1,1)}(2) \times Q_j^{(1,1)}(2)] = [(i-1)\wedge(j-1)]2\sigma^2B_3$. Hence, under condition (C2), we have

$$\frac{T_{n,p}^{(1,1)}(k)}{\sigma^2\sqrt{2\,\mathrm{tr}(\Sigma^2)}} = o_P(1), \qquad k = 1,2.$$

Thus, (A.9) follows.

Next, we shall prove $T_{n,p}^{(k,l)} =: n\binom{n}{2}^{-1}\sum_{i>j}R_{ijk}(X)S_{ijl}(Y) = o_P(\sqrt{\mathrm{tr}(\Sigma^2)})$, where $k,l = 1,2,3,4$ and $(k,l) \neq (1,1)$. Rewrite $T_{n,p}^{(1,2)} = -\frac{1}{2}(1-\frac{1}{n})\widetilde{T}_{n,p}^{(1,2)}$, where $\widetilde{T}_{n,p}^{(1,2)} = \binom{n}{2}^{-1} \sum_{i>j} \boldsymbol{X_i^T X_j}(Y_i^2 + Y_j^2 - 2E(Y_1^2)) =: \widetilde{T}_{n,p}^{(1,2)}(1) + \widetilde{T}_{n,p}^{(1,2)}(2) + \widetilde{T}_{n,p}^{(1,2)}(3)$, and

$$\begin{aligned}
\widetilde{T}_{n,p}^{(1,2)}(1) &= \binom{n}{2}^{-1} \sum_{i=2}^{n}\sum_{j=1}^{i-1} \boldsymbol{X_i^T X_j}\big(\boldsymbol{\beta}^T \boldsymbol{X_j X_j^T \beta} \\
&\qquad + \boldsymbol{\beta}^T \boldsymbol{X_i X_i^T \beta} - 2E\big(\boldsymbol{\beta}^T \boldsymbol{X_1 X_1^T \beta}\big)\big), \\
\widetilde{T}_{n,p}^{(1,2)}(2) &= \binom{n}{2}^{-1} \sum_{i=2}^{n}\sum_{j=1}^{i-1} 2\boldsymbol{X_i^T X_j}\big(\varepsilon_j \boldsymbol{X_j^T \beta} + \varepsilon_i \boldsymbol{X_i^T \beta}\big) \\
&=: 2\binom{n}{2}^{-1} \sum_{i=2}^{n} \widetilde{Q}_i^{(1,2)}(2), \\
\widetilde{T}_{n,p}^{(1,2)}(3) &= \binom{n}{2}^{-1} \sum_{i=2}^{n}\sum_{j=1}^{i-1} \boldsymbol{X_i^T X_j}\big(\varepsilon_i^2 + \varepsilon_j^2 - 2\sigma^2\big) \\
&=: 2\binom{n}{2}^{-1} \sum_{i=2}^{n} \widetilde{Q}_i^{(1,2)}(3).
\end{aligned}$$

If $\widetilde{T}_{n,p}^{(1,2)} = o_P(\sqrt{\operatorname{tr}(\Sigma^2)})$, then we get $T_{n,p}^{(1,2)} = o_P(\sqrt{\operatorname{tr}(\Sigma^2)})$. In fact, by the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
E\big|\widetilde{T}_{n,p}^{(1,2)}(1)\big| &\le \sqrt{2E\big(\boldsymbol{X_1^T X_2 X_1^T X_2}\big) \cdot \operatorname{Var}\big(\boldsymbol{\beta}^T \boldsymbol{X_1 X_1^T \beta}\big)} \\
&= \sqrt{2\Big(3\frac{E(R^4)}{p(p+2)} - 1\Big) B_1^2 \operatorname{tr}\big(\Sigma^2\big)},
\end{aligned}$$

thus, by condition (C2), $\widetilde{T}_{n,p}^{(1,2)}(1) = o_P(\sqrt{\operatorname{tr}(\Sigma^2)})$ is true. It is easy to check

$$\begin{aligned}
E\big[\widetilde{Q}_i^{(1,2)}(2)\widetilde{Q}_i^{(1,2)}(2)\big] &= 2(i-1)\frac{E(R^4)}{p(p+2)}\sigma^2\big(2B_3 + \operatorname{tr}\big(\Sigma^2\big)B_1\big), \\
E\big[\widetilde{Q}_i^{(1,2)}(3)\widetilde{Q}_i^{(1,2)}(3)\big] &= 2(i-1)\big(E\varepsilon_1^4 - \sigma^4\big)\operatorname{tr}\big(\Sigma^2\big).
\end{aligned}$$

If $i \ne j$, then we have $E[\widetilde{Q}_i^{(1,2)}(2)\widetilde{Q}_j^{(1,2)}(2)] = 0$ and $E[\widetilde{Q}_i^{(1,2)}(3)\widetilde{Q}_j^{(1,2)}(3)] = 0$. Thus, under condition (C2), we have $E[\widetilde{T}_{n,p}^{(1,2)}(2)]^2 = O(n^{-2}\operatorname{tr}(\Sigma^2))$ and $E[\widetilde{T}_{n,p}^{(1,2)}(3)]^2 = O(n^{-2}\operatorname{tr}(\Sigma^2))$. Then $\widetilde{T}_{n,p}^{(1,2)}(2) = o_P(\sqrt{\operatorname{tr}(\Sigma^2)})$ and $\widetilde{T}_{n,p}^{(1,2)}(3) = o_P(\sqrt{\operatorname{tr}(\Sigma^2)})$. Rewrite $T_{n,p}^{(1,3)} = -n(1-\frac{1}{n})(1-\frac{2}{n})\binom{n}{2}^{-1}\sum_{i=2}^{n} Q_i^{(1,3)}$ and $T_{n,p}^{(1,4)} = n(1-\frac{1}{n})(1-\frac{2}{n})\binom{n}{2}^{-1}\sum_{i=2}^{n} Q_i^{(1,4)}$, where $Q_i^{(1,3)} = \sum_{j=1}^{i-1}(\boldsymbol{X_i^T X_j}\overline{Y}_{i,j})(Y_i + Y_j)$

and $Q_i^{(1,4)} = \sum_{j=1}^{i-1} \boldsymbol{X}_{\boldsymbol{i}}^T \boldsymbol{X}_{\boldsymbol{j}} (\overline{Y}_{i,j}^2 - \overline{Y}_{1,2}^2)$. Then, by Lemma 3, we have

$$\begin{aligned}
&E\big[Q_i^{(1,3)} Q_i^{(1,3)}\big] \\
&\quad = 2\frac{(i-1)}{(n-2)}\left\{\left[\frac{E(R^4)}{p(p+2)}\big(2B_3 + \operatorname{tr}(\Sigma^2)B_1\big) + \sigma^2 \operatorname{tr}(\Sigma^2)\right](\sigma^2 + B_1)\right\} \\
&\qquad + \frac{(i-1)(i-2)}{(n-2)^2}\left[\frac{E(R^4)}{p(p+2)}\big(2B_2^2 + B_1 B_3\big) + \sigma^2 B_3\right] \\
&\qquad + (i-1)(i-2)\frac{(n-3)}{(n-2)^2}\big(B_1 B_3 + \sigma^2 B_3\big), \\
&E\big[Q_i^{(1,4)} Q_i^{(1,4)}\big] \\
&\quad = \frac{(i-1)}{(n-2)^3} \operatorname{tr}(\Sigma^2)\big[EY_1^4 + 3(n-3)\big(EY_1^2\big)^2 - (n-2)\big(EY_1^2\big)^2\big] \\
&\qquad + 4\frac{(i-1)(i-2)(n-3)}{(n-2)^4} B_3(\sigma^2 + B_1)
\end{aligned}$$

and

$$\begin{aligned}
&E\big[Q_i^{(1,3)} Q_j^{(1,3)}\big] \\
&\quad = \frac{(i-1)\wedge(j-1)}{(n-2)^2}\bigg[(n-2)\sigma^2 B_3 + (n-3)B_1 B_3 \\
&\qquad + \frac{E(R^4)}{p(p+2)}\big(2B_2^2 + B_1 B_3\big)\bigg] \\
&\qquad + 4\frac{[(i-1)(j-1) - (i-1)\wedge(j-1)]}{(n-2)^2} B_2^2, \\
&E\big[Q_i^{(1,4)} Q_j^{(1,4)}\big] \\
&\quad = 4\frac{(i-1)\wedge(j-1)(n-3)}{(n-2)^4} B_3(\sigma^2 + B_1) \\
&\qquad + 4\frac{[(i-1)(j-1) - (i-1)\wedge(j-1)]}{(n-2)^4} B_2^2,
\end{aligned}$$

for $i \neq j$. Under condition (C2), $T_{n,p}^{(1,3)} = o_P(\sqrt{\operatorname{tr}(\Sigma^2)})$ and $T_{n,p}^{(1,4)} = o_P(\sqrt{\operatorname{tr}(\Sigma^2)})$ are true.

For $T_{n,p}^{(2,1)}$, we have very quite similar to the derivations of $T_{n,p}^{(1,2)}$ that $T_{n,p}^{(2,1)} = o_P(\sqrt{\operatorname{tr}(\Sigma^2)})$ by conditions (C1) and (C2). For $T_{n,p}^{(2,2)}$, we obtain that $E|T_{n,p}^{(2,2)}| \leq$

$\frac{1}{2n}[\mathrm{Var}(\boldsymbol{X}_1^T\boldsymbol{X}_1)\,\mathrm{Var}(Y_1^2)]^{1/2}$, where

$$\begin{aligned}\mathrm{Var}(\boldsymbol{X}_1^T\boldsymbol{X}_1) &= 2\frac{E(R^4)}{p(p+2)}\operatorname{tr}(\Sigma^2)+\left(\frac{E(R^4)}{p(p+2)}-1\right)\operatorname{tr}^2(\Sigma),\\ \mathrm{Var}(Y_1^2) &= \left(3\frac{E(R^4)}{p(p+2)}-1\right)B_1^2+4\sigma^2B_1^2+\mathrm{Var}(\varepsilon_1^2).\end{aligned}\tag{A.10}$$

Under condition (C2) and $\operatorname{tr}^2(\Sigma)\le p\operatorname{tr}(\Sigma^2)$, we have $T_{n,p}^{(2,2)}=o_P(\sqrt{\operatorname{tr}(\Sigma^2)})$. By Lemma 3, we have $\mathrm{Var}(\overline{Y}_{i,j}(Y_i+Y_j))=\frac{2}{n-2}E(Y_1^2)^2=\frac{2}{n-2}(\sigma^2+B_1)^2$. Then, by (A.10), we obtain that $E|T_{n,p}^{(2,3)}|\le[\mathrm{Var}(\boldsymbol{X}_1^T\boldsymbol{X}_1)\,\mathrm{Var}(\overline{Y}_{i,j}(Y_i+Y_j))]^{1/2}=o(\sqrt{\operatorname{tr}(\Sigma^2)}))$. Thus, $T_{n,p}^{(2,3)}=o_P(\sqrt{\operatorname{tr}(\Sigma^2)})$ follows under condition (C2). Since

$$\begin{aligned}\mathrm{Var}(\overline{X}_{1,2}^T\overline{X}_{1,2}) &= 2\frac{(\frac{E(R^4)}{p(p+2)}+(n-3))}{(n-2)^3}\operatorname{tr}(\Sigma^2)+\frac{(\frac{E(R^4)}{p(p+2)}-1)}{(n-2)^3}\operatorname{tr}^2(\Sigma),\\ \mathrm{Var}(\overline{Y}_{1,2}^2) &= \frac{3}{(n-2)^3}\frac{E(R^4)}{p(p+2)}B_1^2+\frac{4}{(n-2)^2}\sigma^2B_1\\ &\quad+\frac{1}{(n-2)^3}\mathrm{Var}(\varepsilon_1^2)+\frac{(2n-7)}{(n-2)^3}B_1^2+\frac{2(n-3)}{(n-2)^3}\sigma^4.\end{aligned}\tag{A.11}$$

Combining (A.10), (A.11) and $\operatorname{tr}^2(\Sigma)\le p\operatorname{tr}(\Sigma^2)$, $T_{n,p}^{(2,4)}=o_P(\sqrt{\operatorname{tr}(\Sigma^2)})$ follows under condition (C2).

The proof of $T_{n,p}^{(k,l)}=o_P(\sqrt{\operatorname{tr}(\Sigma^2)}), k=3,4, l=1,2$ are quite similar to that of $T_{n,p}^{(l,k)}=o_P(\sqrt{\operatorname{tr}(\Sigma^2)})$, so we omitted. We here just need to prove $T_{n,p}^{(k,l)}=o_P(\sqrt{\operatorname{tr}(\Sigma^2)}), (k,l)=(3,3),(3,4),(4,4)$. We denote $T_{n,p}^{(3,3)}=n(1-\frac{2}{n})^2\binom{n}{2}^{-1}\sum_{i>j}\overline{\boldsymbol{X}}_{\boldsymbol{i},\boldsymbol{j}}^T(\boldsymbol{X}_i+\boldsymbol{X}_j)\overline{Y}_{i,j}^T(Y_i+Y_j)=:(1-\frac{2}{n})^2[T_{n,p}^{(3,3)}(1)+T_{n,p}^{(3,3)}(2)+T_{n,p}^{(3,3)}(3)]$, where

$$\begin{aligned}T_{n,p}^{(3,3)}(1) &= \binom{n}{2}^{-1}\sum_{i=2}^{n}\sum_{j=1}^{i-1}n\overline{\boldsymbol{X}}_{\boldsymbol{i},\boldsymbol{j}}^T(\boldsymbol{X}_i+\boldsymbol{X}_j)\overline{\boldsymbol{X}}_{\boldsymbol{i},\boldsymbol{j}}^T\boldsymbol{\beta}(Y_i+Y_j),\\ T_{n,p}^{(3,3)}(2) &= \binom{n}{2}^{-1}\sum_{i=2}^{n}\sum_{j=1}^{i-1}n\overline{\boldsymbol{X}}_{\boldsymbol{i},\boldsymbol{j}}^T(\boldsymbol{X}_i+\boldsymbol{X}_j)\overline{\boldsymbol{\varepsilon}}_{\boldsymbol{i},\boldsymbol{j}}(\boldsymbol{X}_i^T\boldsymbol{\beta}+\boldsymbol{X}_j^T\boldsymbol{\beta}),\\ T_{n,p}^{(3,3)}(3) &= \binom{n}{2}^{-1}\sum_{i=2}^{n}\sum_{j=1}^{i-1}n\overline{\boldsymbol{X}}_{\boldsymbol{i},\boldsymbol{j}}^T(\boldsymbol{X}_i+\boldsymbol{X}_j)\overline{\varepsilon}_{i,j}(\varepsilon_i+\varepsilon_j)\\ &=: \binom{n}{2}^{-1}\sum_{i=2}^{n}nQ_i^{(3,3)}(3).\end{aligned}$$

Then we have

$$E\left|T_{n,p}^{(3,3)}(1)\right| \le \frac{2n}{(n-2)}\sqrt{\operatorname{tr}(\Sigma^2)B_1(\sigma^2+B_1)},$$

$$E\left|T_{n,p}^{(3,3)}(2)\right| \le \frac{2n}{(n-2)}\sqrt{\sigma^2 B_1 \operatorname{tr}(\Sigma^2)},$$

$$E\left[Q_i^{(3,3)}(2)Q_i^{(3,3)}(3)\right] = \frac{(i-1)(i+2)}{(n-2)^2}\sigma^4\operatorname{tr}(\Sigma^2)$$

and

$$\begin{aligned} &E\left[Q_i^{(3,3)}(4)Q_j^{(3,3)}(4)\right] \\ &\quad = \left\{\frac{(i-1)\wedge(j-1)}{(n-2)^2}\right. \\ &\qquad \left. + \frac{4[(i-1)(j-1)-(i-1)\wedge(j-1)]}{(n-2)^4}\right\}\sigma^4\operatorname{tr}(\Sigma^2), \end{aligned}$$

for $i \neq j$. Thus, under condition (C2), we have $T_{n,p}^{(3,3)}(k) = o_P(\sqrt{\operatorname{tr}(\Sigma^2)}), k = 1, 2, 3$. Let $T_{n,p}^{(3,4)} =: -(1-\frac{2}{n})^2\widetilde{T}_{n,p}^{(3,4)}$, where $\widetilde{T}_{n,p}^{(3,4)} = n\binom{n}{2}^{-1}\sum_{i>j}\overline{\boldsymbol{X}}_{\boldsymbol{i,j}}^T(\boldsymbol{X_i}+\boldsymbol{X_j})(\overline{Y}_{i,j}^2 - E(\overline{Y}_{1,2}^2))$. Since

$$\begin{aligned} E\left|\widetilde{T}_{n,p}^{(3,4)}\right| &\le n\sqrt{E\left[\overline{\boldsymbol{X}}_{\boldsymbol{i,j}}^T(\boldsymbol{X_i}+\boldsymbol{X_j})(\boldsymbol{X_i}+\boldsymbol{X_j})^T\overline{\boldsymbol{X}}_{\boldsymbol{i,j}}\right]\cdot E\left(\overline{Y}_{i,j}^2 - E\left(\overline{Y}_{1,2}^2\right)\right)^2} \\ &= n\sqrt{\frac{2}{(n-2)}\operatorname{tr}(\Sigma^2)\cdot\operatorname{Var}\left(\overline{Y}_{1,2}^2\right)}, \end{aligned}$$

then by (A.11), we have $T_{n,p}^{(3,4)} = o_P(\sqrt{\operatorname{tr}(\Sigma^2)})$.

For $T_{n,p}^{(4,4)} = (1-\frac{2}{n})^2\widetilde{T}_{n,p}^{(4,4)}$, where $\widetilde{T}_{n,p}^{(4,4)} = \binom{n}{2}^{-1}\sum_{i>j}(\overline{\boldsymbol{X}}_{\boldsymbol{i,j}}^T\overline{\boldsymbol{X}}_{\boldsymbol{i,j}} - E(\overline{\boldsymbol{X}}_{\boldsymbol{1,2}}^T\overline{\boldsymbol{X}}_{\boldsymbol{1,2}}))(\overline{Y}_{i,j}\overline{Y}_{i,j} - E(\overline{Y}_{1,2}\overline{Y}_{1,2}))$. Because $E|\widetilde{T}_{n,p}^{(4,4)}| \le [\operatorname{Var}(\overline{\boldsymbol{X}}_{\boldsymbol{1,2}}^T\overline{\boldsymbol{X}}_{\boldsymbol{1,2}})\cdot\operatorname{Var}(\overline{Y}_{1,2}^2)]^{1/2}$, then by (A.11), we get $T_{n,p}^{(4,4)} = o_P(\sqrt{\operatorname{tr}(\Sigma^2)})$.

Combining (A.9) and Slutsky's theorem, the proof is complete if we show that

$$\frac{\widetilde{T}_{n,p}}{\sqrt{\operatorname{Var}(\widetilde{T}_{n,p})}} \xrightarrow{D} N(0,1),$$

where $\widetilde{T}_{n,p} =: \sum_{i=2}^n\sum_{j=1}^{i-1}\varepsilon_i\varepsilon_j\boldsymbol{X_i}^T\boldsymbol{X_j}/\sqrt{\binom{n}{2}}$ and $\operatorname{Var}(\widetilde{T}_{n,p}) =: \sigma^2\sqrt{\operatorname{tr}(\Sigma^2)}$. Let $\xi_{ni} = \sum_{j=1}^{i-1}\varepsilon_i\varepsilon_j\boldsymbol{X_i}^T\boldsymbol{X_j}/\sqrt{\binom{n}{2}}$, $\upsilon_{ni} = E(\xi_{ni}^2 \mid \mathcal{F}_{i-1})$, and $V_n = \sum_{i=2}^n \upsilon_{ni}$ for $2 \le i \le n$, where $\mathcal{F}_i = \sigma\{\binom{\boldsymbol{X_1}}{\varepsilon_1},\dots,\binom{\boldsymbol{X_i}}{\varepsilon_i}\}$ denotes the σ-field generated by $\{(\boldsymbol{X_j}^T,\varepsilon_j), j \le i\}$. It is easy to check that $E(\xi_{ni} \mid \mathcal{F}_{i-1}) = 0$ and $\{\sum_{i=2}^k \xi_{ni}, \mathcal{F}_k : 2 \le k \le n\}$ is a zero mean martingale. The martingale central limit theorem in [12] follows if we prove

$$\frac{V_n}{\mathrm{Var}(\widetilde{T}_{n,p})} \xrightarrow{P} 1, \tag{A.12}$$

and for any $\eta > 0$

$$\sum_{i=2}^{n} \sigma^{-4} \mathrm{tr}^{-1}(\Sigma^2) E\{\xi_{ni}^2 I(|\xi_{ni}| > \eta\sigma^2 \sqrt{\mathrm{tr}(\Sigma^2)}) \mid \mathcal{F}_{i-1}\} \xrightarrow{P} 0. \tag{A.13}$$

Note that $v_{ni} = \binom{n}{2}^{-1}\sigma^2\{\sum_{j=1}^{i-1}\varepsilon_j^2 \boldsymbol{X}_{\boldsymbol{j}}^T \Sigma \boldsymbol{X}_{\boldsymbol{j}} + 2\sum_{1\le j<l<i}\varepsilon_j\varepsilon_l \boldsymbol{X}_{\boldsymbol{j}}^T \Sigma \boldsymbol{X}_{\boldsymbol{l}}\}$ and

$$\begin{aligned}\frac{V_n}{\mathrm{Var}(\widetilde{T}_{n,p})} &= \Bigg\{\sum_{j=1}^{n-1}(n-j)\varepsilon_j^2 \boldsymbol{X}_{\boldsymbol{j}}^T \Sigma \boldsymbol{X}_{\boldsymbol{j}} \\ &\quad + 2\sum_{1\le j<l\le n}(n-l)\varepsilon_j\varepsilon_l \boldsymbol{X}_{\boldsymbol{j}}^T \Sigma \boldsymbol{X}_{\boldsymbol{l}}\Bigg\} \Big/ \binom{n}{2} \mathrm{tr}(\Sigma^2)\sigma^2 \\ &=: C_{n1} + C_{n2}.\end{aligned}$$

Observe the fact that $E(C_{n1}) = 1$ and

$$\mathrm{Var}(C_{n1}) = \frac{1}{\binom{n}{2}^2 \mathrm{tr}^2(\Sigma^2)\sigma^4} \sum_{j=1}^{n-1} j^2 E[\varepsilon_j^4 (\boldsymbol{X}_{\boldsymbol{j}}^T \Sigma \boldsymbol{X}_{\boldsymbol{j}})^2 - \mathrm{tr}^2(\Sigma^2)\sigma^4],$$

which combined with the conditions (C1), (C2) and Lemma 4 will imply $\mathrm{Var}(C_{n1}) \longrightarrow 0$. Hence, $C_{n1} \xrightarrow{P} 1$. Similar discussion could be performed on the term C_{n2}. It is elemental to obtain that $E(C_{n2}) = 0$ and

$$\begin{aligned}\mathrm{Var}(C_{n2}) &= \frac{4}{\binom{n}{2}^2 \mathrm{tr}^2(\Sigma^2)\sigma^4} \\ &\quad \times \sum_{j_1<l_1}\sum_{j_2<l_2}(n-l_1)(n-l_2)E[\varepsilon_{j_1}\varepsilon_{l_1}\varepsilon_{j_2}\varepsilon_{l_2} \boldsymbol{X}_{\boldsymbol{j_1}}^T \Sigma \boldsymbol{X}_{\boldsymbol{l_1}} \boldsymbol{X}_{\boldsymbol{j_2}}^T \Sigma \boldsymbol{X}_{\boldsymbol{l_2}}] \\ &= \frac{4}{\binom{n}{2}^2} \sum_{1\le j\le n}(j-1)(n-j)^2 \frac{\mathrm{tr}(\Sigma^4)}{\mathrm{tr}^2(\Sigma^2)}.\end{aligned}$$

Note that $\mathrm{tr}(\Sigma^4) = o(\mathrm{tr}^2(\Sigma^2))$. Markov's inequality yields $C_{n2} \xrightarrow{P} 0$. Thus, (A.12) holds. Hence, it finally remains to show (A.13). Since $E\{\xi_{ni}^2 I(|\xi_{ni}| > \eta\sigma^2\sqrt{\mathrm{tr}(\Sigma^2)})\} \le E(\xi_{ni}^4 \mid \mathcal{F}_{i-1})/(\eta^2\sigma^4 \mathrm{tr}(\Sigma^2))$, then by the law of large numbers, it only needs to prove that

$$\sum_{i=2}^{n} E(\xi_{ni}^4) = o(\mathrm{tr}^2(\Sigma^2)). \tag{A.14}$$

By simple calculation, we have

$$\begin{aligned}\sum_{i=2}^{n} E(\xi_{ni}^4) &= \binom{n}{2}^{-1} [E(\varepsilon^4)]^2 [E(\boldsymbol{X}_1^T \boldsymbol{X}_2)^4] \\ &\quad + 3\binom{n}{2}^{-2} E(\varepsilon^4)\sigma^4 \sum_{i=2}^{n} i(i-1) E[(\boldsymbol{X}_1^T \boldsymbol{X}_2)^2 (\boldsymbol{X}_1^T \boldsymbol{X}_3)^2] \\ &\leq O(n^{-1}) E(\boldsymbol{X}_1^T \boldsymbol{X}_2)^4.\end{aligned}$$

By Lemma 4, (A.14) follows. This completes the proof of Theorem 3.4. □

PROOF OF THEOREM 3.2. Write $T_{n,p}^{(k,l)} = n\binom{n}{2}^{-1} \sum_{i>j}(R_{ijk}S_{ijl} - ER_{ijk}S_{ijl})$, where $k, l = 1, 2, 3, 4$. Then, by (A.7)–(A.8), we have $n(T_{n,p} - E(T_{n,p})) = \sum_k \sum_l T_{n,p}^{(k,l)}$. First, we may rewrite

$$\begin{aligned}T_{n,p}^{(1,1)} &= n\left(1 - \frac{1}{n}\right)^2 \binom{n}{2}^{-1} \sum_{i=2}^{n} \sum_{j=1}^{i-1} (\boldsymbol{\beta}^T \boldsymbol{X_i} \boldsymbol{X_i}^T \boldsymbol{X_j} \boldsymbol{X_j}^T \boldsymbol{\beta} - \boldsymbol{\beta}^T \Sigma^2 \boldsymbol{\beta}) \\ &\qquad + (\varepsilon_j \boldsymbol{\beta}^T \boldsymbol{X_i} \boldsymbol{X_i}^T \boldsymbol{X_j} + \varepsilon_i \boldsymbol{\beta}^T \boldsymbol{X_j} \boldsymbol{X_j}^T \boldsymbol{X_i}) + (\varepsilon_i \varepsilon_j \boldsymbol{X_i}^T \boldsymbol{X_j}) \\ &=: T_{n,p}^{(1,1)}(1) + T_{n,p}^{(1,1)}(2) + T_{n,p}^{(1,1)}(3).\end{aligned}$$

We shall prove

$$\frac{T_{n,p}^{(1,1)} - E(T_{n,p}^{(1,1)})}{\sigma^2\sqrt{2\,\mathrm{tr}(\Sigma^2)}} = \frac{T_{n,p}^{(1,1)}(3)}{\sigma^2\sqrt{2\,\mathrm{tr}(\Sigma^2)}} + o_P(1).$$

Second, under conditions (C1) and (C2), we can prove $T_{n,p}^{(k,l)} =: n\binom{n}{2}^{-1} \times \sum_{i>j} R_{ijk}S_{ijl} = o_P(\sqrt{\mathrm{tr}(\Sigma^2)})$, where $k, l = 1, 2, 3, 4$ and $(k, l) \neq (1, 1)$ by calculating their variance or the expectation of their absolute value. Finally, by the martingale central limit theorem [12] and Slutsky's theorem, we have the asymptotical normality of the statistic. □

PROOF OF THEOREM 3.5. We first deal with $\hat{\sigma}_1^2$ by decomposing $(n_2 - |\hat{M}_1|)(\hat{\sigma}_1^2 - \sigma^2)$ as

$$\begin{aligned}&(n_2 - |\hat{M}_1|)(\hat{\sigma}_1^2 - \sigma^2) \\ &\quad = [\mathbf{X}^{(2)^T} \boldsymbol{\beta} + \varepsilon^{(2)}]^T [I_{n_2} - P_{\widehat{M}_1}(\mathbf{X}^{(2)})][\mathbf{X}^{(2)^T} \boldsymbol{\beta} + \varepsilon^{(2)}] \\ &\qquad - (n_2 - |\hat{M}_1|)\sigma^2 \\ &\quad = [\varepsilon^{(2)^T} \varepsilon^{(2)} - n_2\sigma^2] - [\varepsilon^{(2)^T} P_{\widehat{M}_1}(\mathbf{X}^{(2)})\varepsilon^{(2)} - |\hat{M}_1|\sigma^2]\end{aligned}$$

$$
\begin{aligned}
&+ \boldsymbol{\beta}^T \mathbf{X}^{(2)}[I_{n_2} - P_{\widehat{M}_1}(\mathbf{X}^{(2)})]\mathbf{X}^{(2)^T}\boldsymbol{\beta} \\
&+ 2\varepsilon^{(2)^T}[I_{n_2} - P_{\widehat{M}_1}(\mathbf{X}^{(2)})]\mathbf{X}^{(2)^T}\boldsymbol{\beta} \\
=:& T_1 - T_2 + T_3 + 2T_4.
\end{aligned}
$$

Using central limit theorem we have $T_1 = O_P(\sqrt{n_2}) = O_P(\sqrt{n})$ since $n_2 = [(n+1)/2]$. By following the proof of Theorem 2 in [6], we can obtain that $T_2/\sqrt{|\hat{M}_1|} = O_P(1)$, which implies $T_2 = o_P(\sqrt{n})$, since $|\hat{M}_1| = o(n)$ by the variable screening procedure. To deal with T_3, it is easy to show that

$$
\begin{aligned}
E(|T_3|) &\le E[\boldsymbol{\beta}^T \mathbf{X}^{(2)}\mathbf{X}^{(2)^T}\boldsymbol{\beta}] \\
&= n_2 \boldsymbol{\beta}^T \Sigma \boldsymbol{\beta} \\
&= n_2 B_1 = o(n).
\end{aligned}
$$

Thus, by Markov inequality, $T_3 = o_P(n)$. For the term T_4, we have $E(T_4) = 0$ and $\mathrm{Var}(T_4) = \sigma^2 E(T_3)$. Thus, we have $T_4 = o_P(\sqrt{n})$.

Hence, we obtain that $(n_2 - |\hat{M}_1|)(\hat{\sigma}_1^2 - \sigma^2) = O_P(\sqrt{n}) + o_P(\sqrt{n}) + o_P(n) + o_P(\sqrt{n})$ which implies that $\hat{\sigma}_1^2 - \sigma^2 = o_P(1)$. Similarly, we conclude that $\hat{\sigma}_2^2 - \sigma^2 = o_P(1)$. Thus, we have $\hat{\sigma}_{\mathrm{RCV}}^2 = \sigma^2 + o_P(1)$, which completes the proof of the consistency of $\hat{\sigma}_{\mathrm{RCV}}^2$. □

Acknowledgments. The authors would like to thank the Editor, the Associate Editor and the three referees for their constructive comments which have greatly improved the earlier versions of this paper.

All authors equally contributed to this paper, and the authors are listed in alphabetical order. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NNSFC.

Wenwen Guo is the corresponding author.

SUPPLEMENTARY MATERIAL

Supplement to "Test for high-dimensional regression coefficients using refitted cross-validation variance estimation" (DOI: 10.1214/17-AOS1573SUPP; .pdf). This supplemental article contains the proof of Theorem 3.2 and additional figures of empirical powers of different tests.

REFERENCES

[1] BAI, Z. and SARANADASA, H. (1996). Effect of high dimension: By an example of a two sample problem. *Statist. Sinica* **6** 311–329. MR1399305

[2] CAI, T., LIU, W. and XIA, Y. (2014). Two-sample test of high dimensional means under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 349–372. MR3164870

[3] CHEN, S. X. and QIN, Y. L. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38** 808–835. MR2604697

[4] CHEN, S. X., ZHANG, L. X. and ZHONG, P. S. (2010). Tests for high dimensional covariance matrices. *J. Amer. Statist. Assoc.* **105** 810–819. MR2724863

[5] CUI, H., GUO, W. and ZHONG, W. (2018). Supplement to "Test for high-dimensional regression coefficients using refitted cross-validation variance estimation." DOI:10.1214/17-AOS1573SUPP.

[6] FAN, J., GUO, S. and HAO, N. (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **74** 37–65. MR2885839

[7] FAN, J. and LI, R. (2001). Variable selection via nonconcave penalized likelihood and it oracle properties. *J. Amer. Statist. Assoc.* **96** 1348–1360. MR1946581

[8] FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322

[9] FANG, K. T., KOTZ, S. and NG, K. W. (1990). *Symmetric Multivariate and Related Distributions*. Chapman & Hall, London. MR1071174

[10] GOEMAN, J. J., FINOS, L. and VAN HOUWELINGEN, J. C. (2011). Testing against a high dimensional alternative in the generalized linear model: Asymptotic alpha-level control. *Biometrika* **98** 381–390. MR2806435

[11] GOEMAN, J. J., VAN DE GEER, S. and VAN HOUWELINGEN, J. C. (2006). Testing against a high dimensional alternative. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 477–493. MR2278336

[12] HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York. MR0624435

[13] LI, R., ZHONG, W. and ZHU, L. (2012). Feature screening via distance correlation learning. *J. Amer. Statist. Assoc.* **107** 1129–1139. MR3010900

[14] LKHAGVADORJ, S., QU, L., CAI, W., COUTURE, O. P., BARB, C. R., HAUSMAN, G. J., NETTLETON, D., ANDERSON, L. L., DEKKERS, J. C. M. and TUGGLE, C. K. (2009). Microarray gene expression profiles of fasting induced changes in liver and adipose tissues of pigs expressing the melanocortin-4 receptor D298N variant. *Physiol. Genomics* **38** 98–111.

[15] RAO, C. R., TOUTEBURG, H., SHALABH and HEUMANN, C. (2008). *Linear Models and Generalizations*. Springer, New York. MR2370506

[16] SCHMIDT, R. (2001). Tail dependence for elliptically contoured distributions. *Math. Methods Oper. Res.* **55** 301–327. MR1919580

[17] SRIVASTAVA, M. S. and DU, M. (2008). A test for the mean vector with fewer observations than the dimension. *J. Multivariate Anal.* **99** 386–402. MR2396970

[18] TIBSHIRANI, R. (1996). Regression shrinkage and selection via LASSO. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **58** 267–288. MR1379242

[19] WANG, L., PENG, B. and LI, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *J. Amer. Statist. Assoc.* **110** 1658–1669. MR3449062

[20] WANG, S. and CUI, H. (2013). Generalized F test for high dimensional linear regression coefficients. *J. Multivariate Anal.* **117** 134–149. MR3053539

[21] WANG, S. and CUI, H. (2015). A new test for part of high dimensional regression coefficients. *J. Multivariate Anal.* **137** 187–203. MR3332807

[22] YATA, K. and AOSHIMA, M. (2013). Correlation tests for high-dimensional data using extended cross-data-matrix methodology. *J. Multivariate Anal.* **117** 313–331. MR3053550

[23] ZHANG, C. H. and HUANG, J. (2008). The sparsity and bias of the Lasso selection in high-dimensional linear regression. *Ann. Statist.* **36** 1567–1594. MR2435448

[24] ZHONG, P. S. and CHEN, S. X. (2011). Tests for high-dimensional regression coefficients with factorial designs. *J. Amer. Statist. Assoc.* **106** 260–274. MR2816719

[25] ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469

H. CUI
W. GUO
DEPARTMENT OF STATISTICS
CAPITAL NORMAL UNIVERSITY
BEIJING
CHINA
E-MAIL: hjcui@bnu.edu.cn
wwguo@cnu.edu.cn

W. ZHONG
WANG YANAN INSTITUTE FOR STUDIES IN ECONOMICS
DEPARTMENT OF STATISTICS
SCHOOL OF ECONOMICS
FUJIAN KEY LABORATORY OF STATISTICAL SCIENCE
XIAMEN UNIVERSITY
CHINA
E-MAIL: wzhong@xmu.edu.cn