Statistica Sinica **33** (2023), 475-498
doi:https://doi.org/10.5705/ss.202020.0405

NEW TESTS FOR HIGH-DIMENSIONAL LINEAR REGRESSION BASED ON RANDOM PROJECTION

Changyu Liu, Xingqiu Zhao and Jian Huang

The Hong Kong Polytechnic University

Abstract: We consider the problem of detecting significance in high-dimensional linear models, in which the dimension of the regression coefficient is greater than the sample size. We propose novel test statistics for hypothesis tests of the global significance of the linear model, as well as for the significance of part of the regression coefficients. The new tests are based on randomly projecting the high-dimensional data onto a low-dimensional space, and then working with the classical F-test using the projected data. An appealing feature of the proposed tests is that they have a simple form and are computationally easy to implement. We derive the asymptotic local power functions of the proposed tests and compare them with the existing methods for hypothesis testing in high-dimensional linear models. We also provide a sufficient condition under which our proposed tests have higher asymptotic relative efficiency. Simulation studies evaluate the finite-sample performance of the proposed tests and demonstrate that it outperforms existing tests in the models considered. Lastly, we illustrate the proposed tests by applying them to real high-dimensional gene expression data.

Key words and phrases: High-dimensional inference, hypothesis testing, linear model, random projection, relative efficiency.

1. Introduction

High-dimensional data are now routinely encountered in many fields of scientific research. For example, in genomic studies, the dimension of data such as gene expression and genetic marker data is typically far greater than the sample size. A common feature of high-dimensional data is that the data dimension p can be greater than the sample size n. This phenomenon brings challenges to classical statistical analysis, even in many basic settings. For example, the Hotelling T^2 statistic for the two-sample testing problem is not well defined when p is larger than n, because the sample covariance matrix is no longer invertible in this setting. In high-dimensional linear regression models, existing methods for statistical inference about regression coefficients are no longer applicable. There-

Corresponding author: Xingqiu Zhao, Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China. E-mail: xingqiu.zhao@polyu.edu.hk.

fore, it is important to develop new approaches for statistical inference in such models.

Consider the linear regression model

$$y_i = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where y_i is a response variable, $\mathbf{x}_i$ is a $p \times 1$ covariate vector, α is an intercept term, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, and ϵ_i is a random error term with mean zero and variance σ^2. We focus on the high-dimensional setting, in which p can exceed the sample size n. We are interested in testing

$$\mathbf{H}_0 : \boldsymbol{\beta} = \mathbf{0} \quad \text{versus} \quad \mathbf{H}_1 : \boldsymbol{\beta} \neq \mathbf{0}. \tag{1.2}$$

In low-dimensional settings, a basic test statistic for this problem is the F-test (Searle and Gruber (2017)). The idea behind this test is the least squares method, which is based on projecting the vector of response variables onto the space generated by the covariates. Under conditions $p < n$ and $y_i|\mathbf{x}_i \sim \mathcal{N}(\alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta}, \sigma^2)$, the exact distribution of the F-test is known and has certain optimal properties, because it can be considered a likelihood ratio statistic. Without the normality assumption, Wang and Cui (2013) proposed a generalized F-test statistic and showed that it is asymptotically normal when $p/n \to \gamma$, with $\gamma \in (0, 1)$. However, neither the F-test nor the generalized F-test is well defined when $p \geq n$. Even when $p < n$, Zhong and Chen (2011) showed that the F-test is adversely affected by the increasing dimension of the covariates and exhibits a poor performance. Recently, there has been much effort devoted to developing new test statistics for (1.2) that are applicable under the $p > n$ setting. Zhong and Chen (2011) proposed a test based on a U-statistic, and extended it to accommodate factorial designs. This approach was further considered in Cui, Guo and Zhong (2018), who implemented a new variance estimation method based on that of Fan, Guo and Hao (2012). Lan, Wang and Tsai (2014) proposed a test for general random design, and Lan et al. (2016) focused on situations with highly correlated predictors. Based on the low-dimensional projection (LDP) method, many statistical tests have been proposed under the sparsity condition. For example, statistical tests for single or low-dimensional components in high-dimensional models are proposed in Zhang and Zhang (2014), van de Geer et al. (2014), Javanmard and Montanari (2014), and Ning and Liu (2017). For global testing problems, Zhang and Cheng (2017) and Ma, Cai and Li (2021) constructed maximal-type statistics based on the LDP method. For linear hypothesis testing problems, Zhu and Bradic (2018) proposed a test applicable in nonsparse linear models, and Shi et al.

(2019) constructed tests based on the constrained partial regularization method. However, these test statistics are relatively complicated in form and tend to be computationally expensive.

In this paper, we propose a new statistical test for hypothesis (1.2) in high-dimensional settings. Using the technique of random projection to reduce the data dimension, we construct F-statistics based on the projected data in the lower-dimensional space. The F-test has a simple form and is easy to compute. Random projection has been applied to several high-dimensional statistical inference problems, including independence testing (Huang and Huo (2022)), two-sample testing (Lopes, Jacob and Wainwright (2011)), and nonparametric testing (Liu, Shang and Cheng (2018)). An important advantage of the random projection-based approach stems from its ability to reduce the dimension, while simultaneously preserving the significant information in the data. The proposed test is shown to be applicable in a general situation under some mild conditions. The use of random projection results in extra randomness in the test statistic, which requires further investigation of the relationship between the response and the projected data, as well as the performance of the new hat matrix. Our analysis is inspired by the results of Diaconis and Freedman (1984) that almost all low-dimensional projection data are close to normal. Under the null hypothesis, we show that the proposed test statistic is asymptotically normal as $(n, p) \to \infty$. We also derive the asymptotic local power functions of the proposed tests. The results show that the asymptotic performance of the test statistics is similar to that in the setting when the data are normal, and demonstrate the benefit of using random projection to reduce the dimension of the data. Finally, we extend the proposed random projection-based (RP) test procedure for the global hypothesis (1.2) to the problem of testing partial regression coefficients, and derive its asymptotic null distribution and local power function.

The rest of this paper is organized as follows. In Section 2, we propose our test statistic and discuss the reasons for its design. In Section 3, we establish the asymptotic null distribution of the proposed test statistic and derive its asymptotic local power function. We also derive the asymptotic relative efficiency of the proposed test and compare it with that of other recent tests. In Section 4, we extend the proposed test to the problem of testing partial regression coefficients and establish its asymptotic theoretical results. In Section 5.1, we conduct simulation studies to evaluate the finite-sample behavior of the proposed test in terms of the type-I error and power, and compare it with that of competing tests. We apply the proposed test to high-dimensional gene expression data sets in 5.2. Section 6 concludes the paper. The proofs of the lemmas and theorems and additional numerical studies are given in the Supplementary Material.

2. Test Statistic

Let $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ be the ith row of the design matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top$ and $\mathbf{y} = (y_1, \ldots, y_n)^\top$. The linear model (1.1) can be written as

$$\mathbf{y} = \alpha\mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.1}$$

with the error vector $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ and $\mathbf{1} = (1, \ldots, 1)^\top$.

To motivate the proposed test, we first recall the classical F-test of overall significance for regressions in $n > p$ settings. For simplicity, we consider the model without an intercept,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{2.2}$$

We assume $\mathbf{X}$ is full column rank. Let $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$ be the projection matrix (or hat matrix) for the regression. The F-statistic for testing $H_0 : \boldsymbol{\beta} = 0$ is

$$F_n = \frac{\mathbf{y}^\top\mathbf{H}\mathbf{y}/p}{\mathbf{y}^\top(\mathbf{I} - \mathbf{H})\mathbf{y}/(n-p)}. \tag{2.3}$$

Under the normality assumption $\mathbf{y}|\mathbf{X} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I})$, F_n has a noncentral F-distribution with degrees of freedom $(p, n-p)$. The F-test can be derived in different ways. For example, it can be derived based on the distribution of the least squares estimator of $\boldsymbol{\beta}$, and it can also be derived as a likelihood ratio test. Indeed, the F-test is the most widely used method for testing hypotheses about regression coefficients in linear models, and enjoys certain optimality properties. In addition, it has a known finite-sample distribution and is uniformly most powerful invariant (Lehmann (1959)). Clearly, the F-test in (2.3) is not applicable to high-dimensional data with $n < p$.

To overcome this difficulty, we first project the high-dimensional predictors onto a lower-dimensional space, and then apply the F-test to the projected data. Specifically, for an integer $1 \le k < \min\{n, p\}$, let $\mathbf{P}_k \in \mathbb{R}^{p\times k}$ denote a random projection matrix with random entries, drawn independently from the data. Define $\mathbf{u}_{ki} = \mathbf{P}_k^\top\mathbf{x}_i$. Let $\mathbf{U}_k = (\mathbf{u}_{k1}, \ldots, \mathbf{u}_{kn})^\top = \mathbf{X}\mathbf{P}_k$. We consider a *working model*

$$\mathbf{y} = \mathbf{U}_k\boldsymbol{\eta} + \boldsymbol{\epsilon}. \tag{2.4}$$

We use this model to motivate the proposed test statistic. Of course, model (2.4) is generally different from model (2.2). However, for the purpose of constructing a valid test, it suffices that the null hypothesis $\mathbf{H}_0 : \boldsymbol{\beta} = \mathbf{0}$ under model (2.2) is equivalent to the null hypothesis $\mathbf{H}_0 : \boldsymbol{\eta} = \mathbf{0}$ under (2.4). To see this, we focus on a random projection $\mathbf{P}_k$ with independent and identically distributed (i.i.d.) $\mathcal{N}(0,1)$ entries. First, for $\boldsymbol{\eta} = \mathbf{0}$, model (2.4) can be written as $\mathbf{y} = \boldsymbol{\epsilon} = \mathbf{X}\mathbf{0} + \boldsymbol{\epsilon}$. Therefore, $\mathbf{y}$ has the same distribution in model (2.2) for $\boldsymbol{\beta} = \mathbf{0}$. Second, for $\boldsymbol{\eta} \neq \mathbf{0}$, $\mathbf{P}_k\boldsymbol{\eta} \neq \mathbf{0}$ holds with probability one, because $\mathbf{P}_k\boldsymbol{\eta}$ is distributed as $\mathcal{N}(\mathbf{0}, ||\boldsymbol{\eta}||_2^2\mathbf{I})$. Consequently, $\boldsymbol{\beta} = \mathbf{0}$ in model (2.2) implies $\boldsymbol{\eta} = \mathbf{0}$ in model (2.4); otherwise, a contradiction will be caused by $\mathbf{P}_k\boldsymbol{\eta} \neq \mathbf{0}$. Now, suppose $\mathbf{U}_k$ is full column rank (this can be guaranteed if $k < n$ and $\mathbf{X}$ is full row rank). The projection matrix for (2.4) is

$$\mathbf{H}_k = \mathbf{U}_k(\mathbf{U}_k^\top\mathbf{U}_k)^{-1}\mathbf{U}_k^\top.$$

The F-statistic based on (2.4) is

$$T_n = \frac{\mathbf{y}^\top\mathbf{H}_k\mathbf{y}/k}{\mathbf{y}^\top(\mathbf{I} - \mathbf{H}_k)\mathbf{y}/(n-k)}. \tag{2.5}$$

For the model with an intercept, $\mathbf{y} = \alpha\mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, we can simply center the design matrix and modify the test statistic as

$$T_n = \frac{\mathbf{y}^\top\mathbf{H}_k\mathbf{y}/k}{\mathbf{y}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_k)\mathbf{y}/(n-k-1)}, \tag{2.6}$$

where $\mathbf{P}_1 = (1/n)\mathbf{1}\mathbf{1}^\top$ and $\mathbf{H}_k = \mathbf{U}_k(\mathbf{U}_k^\top\mathbf{U}_k)^{-1}\mathbf{U}_k^\top$ is a new hat matrix with $\mathbf{U}_k = (\mathbf{I} - \mathbf{P}_1)\mathbf{X}\mathbf{P}_k$.

Note that the matrix $\mathbf{U}_k^\top\mathbf{U}_k$ is of full rank with probability one when $\mathbf{P}_k$ has i.i.d. $\mathcal{N}(0,1)$ entries, which ensures the new hat matrix is well defined, even when $p > n$, as shown in the proof of Theorem 1.

From the definition, the new test is based on a projection of the response vector $\mathbf{y}$ onto the space spanned by the columns of $\mathbf{U}_k$, which is a linear subspace of the space spanned by the columns of the centered $\mathbf{X}$.

A convenient way to construct $\mathbf{P}_k$ is to generate its entries as i.i.d. random variables from the standard normal distribution $N(0,1)$. Li, Hastie and Church (2006) suggested that one can also generate other types of random projections $\mathbf{P}_k$, for example, sparse random projections, to achieve asymptotically the same performance as the normal random projection at a fast convergence rate. A sparse random projection consists of entries p_{ij} that are i.i.d. from distributions satisfying

$$P(p_{ij} = \sqrt{l}) = P(p_{ij} = -\sqrt{l}) = \frac{1}{2l},\ P(p_{ij} = 0) = 1 - \frac{1}{l}, \tag{2.7}$$

where the choice of l is recommended to be $\sqrt{p}$. Under this case, Li, Hastie and Church (2006) showed that the projected data converge to normal at a rate of $O(p^{-1/4})$.

In our theoretical analysis, we focus on random projections of i.i.d. normal random entries. The results can be applied to some non-normal projections. We use the above sparse random projection and evaluate the performance of non-normal projections in the simulation studies.

3. Main Results

This section contains our main theoretical results and related discussions. Specifically, we derive the asymptotic normality and asymptotic power function for the new RP test. We also compare the proposed test with one of the latest tests in terms of their asymptotic relative efficiency.

3.1. Asymptotic normality

Our first main result demonstrates the asymptotic normality of the standardized T_n under the null hypothesis. We work under the following assumptions.

Assumption 1. *$\boldsymbol{x}_i = \boldsymbol{\mu} + \boldsymbol{\Gamma} \boldsymbol{z}_i$, where $\boldsymbol{\Gamma}$ is a $p \times m$ matrix with $m \geq p$, $\boldsymbol{\mu}$ is a p-dimensional vector, and $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{im})^\top$ is an m-variate random vector with $E(\boldsymbol{z}_i) = \boldsymbol{0}, Var(\boldsymbol{z}_i) = \boldsymbol{I}_m$, and $Var(\boldsymbol{z}_i^\top \boldsymbol{z}_i/m) = O(m^{-1})$. For any nonnegative integers $q_1, \ldots, q_m$, with $\sum_{j=1}^m q_j = 4$, the mixed moments $E\big(\Pi_{j=1}^m z_{ij}^{q_j}\big)$ are bounded, and equal to zero when at least one of the q_j is odd.*

Assumption 2. *$\mu_4 = E(\epsilon_1^4) < \infty$.*

Assumption 3. *$p \gg n$ and there is a constant $\rho \in (0, 1)$ such that $k/n \to \rho$.*

As stated in Assumptions 1 and 3, we do not place any concrete relationships between n and p, allowing the dimension p, mean vector $\boldsymbol{\mu}$, and covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top$ to vary implicitly as n goes to infinity. This makes our test accommodate extremely high-dimensional problems. Taking a closer look at Assumption 1, we find it resembles a factor model structure with a linear relationship between $\mathbf{x}_i$ and $\mathbf{z}_i$. It can be proved that the following two assumptions are both included in Assumption 1.

D1 (Pseudo-independence assumption.) Suppose the p-variate random vector $\mathbf{x}_i$ follows the general multivariate model $\mathbf{x}_i = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{z}_i$, where $\boldsymbol{\mu}$ is a p-dimensional real vector, $\boldsymbol{\Gamma}$ is a $p \times m$ matrix, and $\mathbf{z}_i = (z_{i1}, \ldots, z_{im})^\top$ is an m-variate random vector with $E(\mathbf{z}_i) = \mathbf{0}$ and $Var(\mathbf{z}_i) = \mathbf{I}_m$. Furthermore, each

z_{ij} satisfies $E(z_{ij}^4) = 3+\Delta < \infty$, for some constant Δ, and $E(z_{ij_1}^{l_1} \cdots z_{ij_d}^{l_d}) = E(z_{ij_1}^{l_1}) \cdots E(z_{ij_d}^{l_d})$ for any $\sum_{v=1}^{d} l_v \leq 4$ and $j_1 \neq \cdots \neq j_d$, where d is a positive integer. The integers m and p satisfy $m \geq p$.

D2 (Elliptical distribution assumption.) Suppose the p-variate random vector $\mathbf{x}_i$ satisfies the stochastic representation $\mathbf{x}_i = \boldsymbol{\mu} + \boldsymbol{\Gamma} r_i \mathbf{u}_i$, where $\boldsymbol{\mu}$ is a p-dimensional real vector, $\boldsymbol{\Gamma}$ 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 $\mathbf{u}_i$ satisfying $E(r_i^2) = p$ and $Var(r_i^2) = O(p)$.

The pseudo-independence assumption and similar versions are used in Bai and Saranadasa (1996), Zhong and Chen (2011), and Cui, Guo and Zhong (2018). Such assumptions are similar to Assumption 1, but impose stricter conditions on each element of $\mathbf{z}_i$. This is because $\mathbf{z}_i$ in D1 satisfies $Var(\mathbf{z}_i^\top \mathbf{z}_i/m) = (2+\Delta)/m$. In a multivariate statistical analysis, an elliptical distribution is often assumed. It includes a flexible family of distributions, including the multivariate normal distribution, multivariate t-distribution, and multivariate logistic distribution. Let $\mathbf{z}_i = r_i \mathbf{u}_i$ and $m = p$; then, D2 and Assumption 1 enjoy a similar form. Furthermore, Lemma 1, together with Lemma S1 in the Supplementary Material, indicates that the distributions satisfying D2 are included in Assumption 1.

Because T_n is invariant to the location shift of $\mathbf{y}$ and $\mathbf{X}$, we assume that $\alpha = 0$ and $\boldsymbol{\mu} = \mathbf{0}$ in the rest of the paper.

Lemma 1. *Suppose $\boldsymbol{u}_1$ is a random vector uniformly distributed on the unit sphere in $\mathbb{R}^p$ and r_1 is a nonnegative random variable independent of $\boldsymbol{u}_1$ satisfying $E(r_1^2) = p$ and $Var(r_1^2) = O(p)$. Let $\boldsymbol{z}_1 = r_1 \boldsymbol{u}_1$. Then,*

$$E(\boldsymbol{z}_1) = \boldsymbol{0}, \quad Var(\boldsymbol{z}_1) = \boldsymbol{I}_p, \quad Var\left(\frac{\boldsymbol{z}_1^\top \boldsymbol{z}_1}{p}\right) = O(p^{-1}).$$

Therefore, our assumption for the distribution of $\mathbf{x}_i$ is relatively flexible. For example, there is no specific condition on the covariance matrix $\boldsymbol{\Sigma}$. For the error term, we only assume that ϵ_i is generated from a distribution having a finite fourth moment. The projection dimension k is assumed to be asymptotically proportional to n with a coefficient $\rho \in (0,1)$. The choice of ρ is discussed in Section 3.3.

Clearly, to derive the asymptotic distribution of T_n, we need to study the properties of the hat matrix $\mathbf{H}_k$. Because $\mathbf{H}_k = \mathbf{U}_k(\mathbf{U}_k^\top \mathbf{U}_k)^{-1}\mathbf{U}_k^\top$, the properties of $\mathbf{H}_k$ can be established when $\mathbf{U}_k$ is generated from Gaussian variables. Diaconis and Freedman (1984) showed that the empirical distribution of randomly projected data tends to be approximately Gaussian. Inspired by this result, we show in Lemmas S7 and S8 in the Supplementary Material that $\mathbf{U}_k$ is asymptotically close to Gaussian, which demonstrates the advantage of the random projection method. We state the asymptotic distribution of the standardized T_n under the null hypothesis.

Theorem 1. *Suppose the random projection matrix $\boldsymbol{P}_k$ consists of i.i.d. standard normal random variables. Under Assumptions* 1–3 *and $\boldsymbol{H}_0$, as $n \to \infty$, we have*

$$\frac{T_n - 1}{\sqrt{2/n\rho(1-\rho)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

This asymptotic normality result justifies the following test procedure. Given an α-level of significance, the proposed test rejects $\mathbf{H}_0$ if

$$\frac{T_n - 1}{\sqrt{2/n\rho(1-\rho)}} > z_\alpha,$$

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

3.2. Asymptotic power function

We now investigate the asymptotic power function of the proposed test. Additional assumptions are needed to facilitate our analysis.

Assumption 4. $\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta} = o(1)$.

Assumption 4 is known as a local alternative, and is commonly used to study the asymptotic properties of a statistical test. Detailed discussions can be found in van der Vaart (1998, Sec. 14.1).

In the classical F-test in (2.3), the hat matrix $\mathbf{H}$ enjoys the properties $\mathbf{X}^\top \mathbf{H} = \mathbf{X}^\top$ and $\mathbf{H}\mathbf{X} = \mathbf{X}$. Hence,

$$\mathbf{y}^\top \mathbf{H} \mathbf{y} = \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta} + \mathbf{2} \boldsymbol{\beta}^\top \mathbf{X}^\top \boldsymbol{\epsilon} + \boldsymbol{\epsilon}^\top \mathbf{H} \boldsymbol{\epsilon},$$

where $\boldsymbol{\epsilon}^\top \mathbf{H} \boldsymbol{\epsilon}$ does not involve the parameter value. This indicates that the power of the F-test relies on $\boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\beta}$ and $\boldsymbol{\beta}^\top \mathbf{X}^\top \boldsymbol{\epsilon}$. Thus, we can use the properties of $\mathbf{H}$ in the power analysis of the F-test without needing to consider the inverse of $\mathbf{X}^\top \mathbf{X}$.

However, the properties of $\mathbf{H}$ do not hold for the hat matrix $\mathbf{H}_k$. Fortunately, we can get around this problem based on the properties of a random projection. Specifically, the fact that a randomly projected variable is asymptotically normal yields a new representation for model (2.1) by $\mathbf{y} = \mathbf{X}\mathbf{P}_k \boldsymbol{\xi} + \mathbf{e}$, where

$\boldsymbol{\xi} = (\mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k)^{-1} \mathbf{P}_k^\top \boldsymbol{\Sigma} \boldsymbol{\beta}$ and $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{P}_k \boldsymbol{\xi}$. Note that α is assumed to be zero here. It can be shown that the new error term $\mathbf{e}$ is asymptotically conditional independent of $\mathbf{X}\mathbf{P}_k$, making the conventional analysis for the F-test applicable here. To show this rigorously, we need an additional requirement for $\mathbf{z}_i$.

Assumption 5. *The m-variate random vector $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{im})^\top$ has a Lebesgue density $f_{\boldsymbol{z}}$ and satisfies $E(\boldsymbol{z}_i) = \boldsymbol{0}$ and $Var(\boldsymbol{z}_i) = \boldsymbol{I}_m$. For $j = 1, \ldots, m$, the components z_{ij} are assumed to be independent, satisfy $E(z_{ij}^{20}) \leq C$ for a constant C, and have a marginal density bounded by a constant $D \geq 1$.*

Define $\delta_k^2 = \sigma^2 + \boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta} - \boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k \boldsymbol{\xi}$ as the variance of the new error. We derive the asymptotic power function of the proposed test.

Theorem 2. *Suppose Assumptions 1–5 hold. Let $\Psi_n^{RP}(\boldsymbol{\beta}; \boldsymbol{P}_k)$ denote the power function of the proposed RP test T_n. Then,*

$$\Psi_n^{RP}(\boldsymbol{\beta}; \boldsymbol{P}_k) - \Phi\left(- z_\alpha + \sqrt{\frac{n(1-\rho)}{2\rho} \frac{\boldsymbol{\xi}^\top \boldsymbol{P}_k^\top \boldsymbol{\Sigma} \boldsymbol{P}_k \boldsymbol{\xi}}{\delta_k^2}} \right) \to 0,$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and z_α is the upper α-quantile of Φ.

Note that there is no extra assumption made for $\boldsymbol{\Sigma}$, showing that the power property of the proposed test holds over a wide range of alternatives. The asymptotic power function relies on $\mathbf{P}_k$ and is an increasing function of the product $\boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k \boldsymbol{\xi}$. We find that the product is upper bounded by $\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}$, which can be reached when the vector $\boldsymbol{\Gamma}^\top \boldsymbol{\beta}$ is in the space generated by $\boldsymbol{\Gamma}^\top \mathbf{P}_k$. To make the bound achieved asymptotically, we give a sufficient condition.

Assumption 6. *(Tail eigenvalue condition.) There exists an integer s and a real number $\gamma > 0$ such that $s < k$ and $||\boldsymbol{\beta}||_2^2 \sum_{i=s+1}^{p} d_i = o(pn^{-0.5-\gamma})$, where d_i are the eigenvalues of $\boldsymbol{\Sigma}$ satisfying $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$.*

We call Assumption 6 a tail eigenvalue condition, because it requires the product of $||\boldsymbol{\beta}||_2^2$ and the sum of the tail eigenvalues of $\boldsymbol{\Sigma}$ to be of order less than $p/\sqrt{n}$.

Lemma 2. *Let $\boldsymbol{P}_k \in \mathbb{R}^{p \times k}$ consist of i.i.d. $\mathcal{N}(0,1)$ entries. Assume that Assumption 6 holds. Then, we have*

$$\sqrt{n} ||\boldsymbol{\Gamma}^\top \boldsymbol{\beta} - \boldsymbol{\Gamma}^\top \boldsymbol{P}_k \boldsymbol{\eta}||_2^2 = o(1),$$

for some $\boldsymbol{\eta} \in \mathbb{R}^k$ with probability tending to one.

This lemma indicates that we can approximate $\boldsymbol{\Gamma}^\top \boldsymbol{\beta}$ by $\boldsymbol{\Gamma}^\top \mathbf{P}_k \boldsymbol{\eta}$ with a negligible approximation error. In this case, we denote the asymptotic power function as $\Psi_n^{RP}(\boldsymbol{\beta})$, because it is not related to $\mathbf{P}_k$. A formal result is given in the following corollary.

Corollary 1. *Suppose Assumptions* 1–6 *hold. Then,*

$$\Psi_n^{RP}(\boldsymbol{\beta}) - \Phi\left(- z_\alpha + \sqrt{\frac{n(1-\rho)}{2\rho}} \frac{\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}}{\sigma^2} \right) \to 0,$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and z_α is the upper α-quantile of Φ.

3.3. Choice of ρ

The proposed test can be applied with any dimension of the projected space k that satisfies Assumption 3. However, the power of the test depends on ρ. In this subsection, we give a detailed discussion on the choice of ρ.

From Theorem 2, the asymptotic local power function satisfies

$$\Psi_n^{RP}(\boldsymbol{\beta}; \mathbf{P}_k) = \Phi\left(- z_\alpha + \sqrt{\frac{n(1-\rho)}{2\rho}} \frac{\boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k \boldsymbol{\xi}}{\delta_k^2} \right) + o(1). \tag{3.1}$$

Let $\Delta_k^2 = \boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k \boldsymbol{\xi}$. Then, Δ_k^2 can be derived by projecting the vector $\boldsymbol{\Gamma}^\top \boldsymbol{\beta}$ onto the space generated by $\boldsymbol{\Gamma}^\top \mathbf{P}_k$. Intuitively, a larger ρ would lead to larger Δ_k^2, because the dimension of the projection space $k = \rho n$ becomes larger. However, with an increase of ρ, the function $\sqrt{(1-\rho)/\rho}$ becomes smaller. This indicates that the choice of ρ is a compromise between these two values.

First, we consider the situation in which the condition given in Corollary 1 is satisfied. In this case, Δ_k^2 becomes a deterministic value, even with the randomly generated projection matrix $\mathbf{P}_k$. The asymptotic local power function is a decreasing function of ρ, confirmed in the simulation studies. Furthermore, ρ can be arbitrarily small, as long as the tail eigenvalue condition is satisfied.

Then, we consider the other situation, where $\boldsymbol{\Sigma} = \mathbf{I}$. In this case, the eigenvalues of $\boldsymbol{\Sigma}$ are equally significant, with $\Delta_k^2 = \boldsymbol{\beta}^\top \mathbf{P}_k (\mathbf{P}_k^\top \mathbf{P}_k)^{-1} \mathbf{P}_k^\top \boldsymbol{\beta}$. Suppose that the direction of $\boldsymbol{\beta}$ is uniformly generated on the unit sphere. From Proposition 1 in Lopes, Jacob and Wainwright (2011), quantity Δ_k^2 satisfies

$$P\left(\frac{\Delta_k^2}{||\boldsymbol{\beta}||_2^2} \geq \frac{ck}{p} \right) \to 1 \text{ and } P\left(\frac{\Delta_k^2}{||\boldsymbol{\beta}||_2^2} \leq \frac{Ck}{p} \right) \to 1,$$

for some constants c and C. This indicates that Δ_k^2 scales linearly in k up to random fluctuations. Combining this result with (3.1), the influence of ρ on the testing power is achieved mainly based on the function $g(\rho) = \sqrt{(1-\rho)/\rho} \cdot \rho$, which is maximized when $\rho = 0.5$. Therefore, a choice of $k = [0.5n]$ may be asymptotically optimal, in a general sense.

In many applications, no prior information on $\mathbf{\Sigma}$ is available. In such cases, the above discussion suggests that ρ around 0.5 is an applicable choice, because the above setting yields reasonable test performance, even in extreme cases. When estimation methods of $\mathbf{\Sigma}$ or a related function of $\mathbf{\Sigma}$ are available, ρ can be selected based on the estimators. For example, the ratio $tr(\mathbf{\Sigma})^2/tr(\mathbf{\Sigma}^2)$, which lies between one and p, can be viewed as measuring the decay rate of the spectrum of $\mathbf{\Sigma}$ (Lopes, Jacob and Wainwright (2011)). In addition, the tail eigenvalue condition can be satisfied when $tr(\mathbf{\Sigma})^2/tr(\mathbf{\Sigma}^2) \ll p$. Consequently, we can determine ρ from the estimation of the ratio, which is available based on the estimators of $tr(\mathbf{\Sigma})$ and $tr(\mathbf{\Sigma}^2)$ proposed in Chen, Zhang and Zhong (2010).

3.4. Asymptotic relative efficiency

The asymptotic power function of the proposed RP test in Corollary 1 has the same form as the F-test studied in Zhong and Chen (2011). However, our test accommodates high-dimensional settings and has milder assumptions on $\mathbf{X}$ and $\boldsymbol{\epsilon}$. Because it is well known that the F-test has good performance in low dimensions, the new test, as an extension of the F-test to high dimensions, is expected to perform well under certain conditions. To confirm this, we compare the performance of our test with the test proposed by Cui, Guo and Zhong (2018), which is one of the latest tests designed for problem (1.2), and is demonstrated to outperform existing tests for the problem considered. We denote this competing test as the RCV test, and show that our test outperforms it in some situations. In this subsection, we suppose Assumption 6 holds.

With a slight abuse of notation, we also denote the asymptotic power function of our RP test as

$$\Psi_n^{RP}(\boldsymbol{\beta}) = \Phi\left(- z_\alpha + \sqrt{\frac{n(1-\rho)}{2\rho}} \frac{\boldsymbol{\beta}^\top \mathbf{\Sigma} \boldsymbol{\beta}}{\sigma^2} \right).$$

The asymptotic power function of the RCV test proposed by Cui, Guo and Zhong (2018) is given by

$$\Psi_n^{RCV}(\boldsymbol{\beta}) = \Phi\left(- z_\alpha + \frac{n\boldsymbol{\beta}^\top \mathbf{\Sigma}^2 \boldsymbol{\beta}}{\sigma^2 \sqrt{2tr(\mathbf{\Sigma}^2)}} \right).$$

Because the term added to $-z_\alpha$ inside the $\Phi(\cdot)$ function is what controls the power, the ratio of such terms can be defined as the asymptotic relative efficiency (ARE). For comparison, we define the ARE of our test to the RCV test as

$$ARE(\Psi_n^{RP}, \Psi_n^{RCV}) = \left(\sqrt{\frac{n(1-\rho)}{\rho}}\boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta} / \frac{n\boldsymbol{\beta}^\top \boldsymbol{\Sigma}^2 \boldsymbol{\beta}}{\sqrt{tr(\boldsymbol{\Sigma}^2)}}\right)^2. \tag{3.2}$$

Whenever the ARE is larger than one, the proposed test is asymptotically more powerful than the competing test. Therefore, we search for sufficient conditions under which the ARE is greater than one.

Write $\boldsymbol{\beta}$ as $||\boldsymbol{\beta}||_2\boldsymbol{\delta}$, where $\boldsymbol{\delta} = \boldsymbol{\beta}/||\boldsymbol{\beta}||_2$ is the direction of $\boldsymbol{\beta}$. Under Assumptions 4 and 6, we further require the sum of the tail eigenvalues to satisfy $\sum_{i=s+1}^p d_i/\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta} = O(pn^{-0.5-\gamma})$, where γ is a small constant greater than zero. By Jensen's inequality, we have

$$ARE(\Psi_n^{RP}; \Psi_n^{RCV}) \geq \frac{1-\rho}{\rho}\frac{\sum_{i=s+1}^p d_i^2}{n}\left(\frac{\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta}}{\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^2\boldsymbol{\delta}}\right)^2 \geq \frac{1-\rho}{\rho}\frac{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta})^4}{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^2\boldsymbol{\delta})^2} O(pn^{-2-2\gamma}). \tag{3.3}$$

Clearly, if $(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^2\boldsymbol{\delta})^2/(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta})^4 = o(pn^{-2-2\gamma})$, the right side of the inequality goes to infinity as n goes to infinity, which sufficiently demonstrates that the proposed test is more powerful than the RCV test. In addition, this inequality shows that ρ is preferred to be the smallest value such that the tail eigenvalue condition holds.

We give two examples to illustrate situations where $(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^2\boldsymbol{\delta})^2/(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta})^4 = o(pn^{-2-2\gamma})$ is satisfied.

Example 1. Suppose $\boldsymbol{\beta}$ is an eigenvector of $\boldsymbol{\Sigma}$; then, $(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^2\boldsymbol{\delta})^2/(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta})^4 = 1$. Given that $n = o(p^{1/(2+2\gamma)})$ for a constant $\gamma > 0$, which frequently happens when $p \gg n$, we have $(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^2\boldsymbol{\delta})^2/(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta})^4 = o(pn^{-2-2\gamma})$.

Example 2. Suppose the covariance matrix $\boldsymbol{\Sigma}$ has the spectral decomposition

$$\boldsymbol{\Sigma} = \mathbf{O}\boldsymbol{\Lambda}\mathbf{O}^\top = \mathbf{O}\,\mathrm{diag}(d_1, \ldots, d_p)\mathbf{O}^\top,$$

where $\mathbf{O}$ is an orthogonal matrix with the ith column denoted by $\mathbf{O}_i$, and d_i are the eigenvalues of $\boldsymbol{\Sigma}$ satisfying $0 \leq d_1 \leq d_2 \leq \cdots \leq d_p$. We assume there exist integers $1 \leq s_1 \leq s_2 \leq p$ and constants $r_1 \leq r_2$ such that, for $i = s_1, \ldots, s_2$, the order of d_i is between n^{r_1} and n^{r_2} in the sense that $1/d_i = O(n^{-r_1})$ and $d_i = O(n^{r_2})$. Consider $\boldsymbol{\beta} \in Span\{\mathbf{O}_{s_1}, \ldots, \mathbf{O}_{s_2}\}$. Then, we get

$$\frac{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^2\boldsymbol{\delta})^2}{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta})^4} \leq O(n^{4(r_2-r_1)}).$$

When n and p satisfy $n = o(p^{1/2(1+\gamma+2r_2-2r_1)})$ for a constant $\gamma > 0$, we have $(\boldsymbol{\delta}^\top \boldsymbol{\Sigma}^2 \boldsymbol{\delta})^2/(\boldsymbol{\delta}^\top \boldsymbol{\Sigma}\boldsymbol{\delta})^4 = o(pn^{-2-2\gamma})$, and thus our test outperforms the RCV test in these situations.

4. Testing Partial Regression Coefficients

In Section 3, we proposed an RP test for the hypothesis test in (1.2). In many studies, we are also interested in investigating the significance of part of the covariates. In this section, we generalize the test in Section 3 to hypothesis testing of a partial linear regression coefficient, and derive its asymptotic results.

Consider a linear regression model

$$y_i = \alpha + \mathbf{x}_{1i}^\top \boldsymbol{\beta}_1 + \mathbf{x}_{2i}^\top \boldsymbol{\beta}_2 + \epsilon_i, \qquad i = 1, \ldots, n, \tag{4.1}$$

where α is an intercept term, $\mathbf{x}_{1i}$ is a p_1-dimensional covariate and $\mathbf{x}_{2i}$ is a p_2-dimensional covariate, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are vectors of unknown regression coefficients, and ϵ_i is a random variable with mean zero and variance σ^2. We are interested in testing

$$\mathbf{H}_{part,0} : \boldsymbol{\beta}_2 = \mathbf{0} \quad \text{versus} \quad \mathbf{H}_{part,1} : \boldsymbol{\beta}_2 \neq \mathbf{0}. \tag{4.2}$$

Let $\mathbf{y} = (y_1, \ldots, y_n)^\top$ and $\mathbf{x}_{1i} = (x_{i1}^1, \ldots, x_{ip_1}^1)^\top$ be the ith row of the matrix $\mathbf{X}_1 = (\mathbf{x}_{11}, \ldots, \mathbf{x}_{1n})^\top$. Similarly, let $\mathbf{X}_2 = (\mathbf{x}_{21}, \ldots, \mathbf{x}_{2n})^\top$. The linear model (4.1) can be written as

$$\mathbf{y} = \alpha \mathbf{1} + \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \tag{4.3}$$

with the error vector $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ and $\mathbf{1} = (1, \ldots, 1)^\top$.

Following the same idea as in Section 3, we develop a new test for the hypothesis test in (4.2). For an integer $1 \le k_2 < \min\{n - p_1, p_2\}$, let $\mathbf{P}_{k_2} \in \mathbb{R}^{p_2 \times k_2}$ be a matrix with i.i.d. $\mathcal{N}(0, 1)$ entries, drawn independently from the data. We define the following projection matrices.

$$\begin{aligned}
\mathbf{P}_1 &= \frac{1}{n}\mathbf{1}\mathbf{1}^\top, \\
\mathbf{P}_{\mathbf{X}_1} &= (\mathbf{I} - \mathbf{P}_1)\mathbf{X}_1(\mathbf{X}_1^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{X}_1)^{-1}\mathbf{X}_1^T(\mathbf{I} - \mathbf{P}_1), \\
\mathbf{H}_{k_2} &= (\mathbf{I} - \mathbf{P}_1)\mathbf{W}(\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{W})^{-1}\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1),
\end{aligned}$$

where $\mathbf{W} = (\mathbf{X}_1, \mathbf{X}_2\mathbf{P}_{k_2})$. Note that the matrix $\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{W}$ is of full rank with probability one when $\mathbf{P}_{k_2}$ has i.i.d. $\mathcal{N}(0, 1)$ entries and k_2 is selected appropriately. This ensures the projection matrix $\mathbf{H}_{k_2}$ is well defined, even when $p_2 > n$. We propose a new test statistic

$$T_{n,p_2} = \frac{\mathbf{y}^\top(\mathbf{H}_{k_2} - \mathbf{P}_{\mathbf{X}_1})\mathbf{y}/k_2}{\mathbf{y}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_{k_2})\mathbf{y}/(n - 1 - p_1 - k_2)}. \tag{4.4}$$

From the definition, the numerator of T_{n,p_2} represents the part of $\mathbf{y}$ that can only be explained by $\mathbf{X}_2\mathbf{P}_{k_2}$, and the denominator of T_{n,p_2} estimates the variance of the error term.

4.1. Asymptotic null distribution

To study the asymptotic null distribution and the power of the proposed test, we make the following assumptions.

Assumption S1. $\mathbf{x}_i = (\mathbf{x}_{1i}^\top, \mathbf{x}_{2i}^\top)^\top = \boldsymbol{\mu} + \boldsymbol{\Gamma}\mathbf{z}_i$, where $\mathbf{x}_{1i} \in \mathbb{R}^{p_1}$ and $\mathbf{x}_{2i} \in \mathbb{R}^{p_2}$ are covariates, $\boldsymbol{\mu}$ is a p-dimensional mean vector, $\boldsymbol{\Gamma}$ is a $p \times m$ matrix with $m \geq p$, and $\mathbf{z}_i$ is an m-variate random vector with $E(\mathbf{z}_i) = \mathbf{0}, Var(\mathbf{z}_i) = \mathbf{I}_m$, and $Var(\mathbf{z}_i^\top \mathbf{z}_i/m) = O(m^{-1})$. For any nonnegative integers $q_1, \ldots, q_m$, with $\sum_{j=1}^m q_j = 4$, the mixed moments $E\big(\Pi_{j=1}^m z_{ij}^{q_j}\big)$ are bounded, and are equal to zero when at least one q_j is odd.

Assumption S2. $\mu_4 = E(\epsilon_1^4) < \infty$.

Assumption S3. $p = p_1 + p_2 \gg n$, $p_2 \gg p_1$, and there exist constants $\rho_1, \rho_2 \in (0,1)$, with $\rho_1 + \rho_2 < 1$, such that $p_1/n \to \rho_1$ and $k_2/n \to \rho_2$.

Because T_{n,p_2} is invariant to the location shift of $\mathbf{y}$, $\mathbf{X}_1$, and $\mathbf{X}_2$, we assume $\alpha = 0$ and $\boldsymbol{\mu} = \mathbf{0}$ in the following. The dimensions of the covariates are assumed to satisfy $p_2 \gg p_1$, so $\mathbf{X}_2$ is the high-dimensional component. In addition, p_1 is assumed to be less than, but possibly comparable with n. The projection dimension k_2 needs to be asymptotically proportional to n, and the choice of ρ_2 is discussed below.

Theorem 3. *Under Assumptions* S1–S3 *and* $\boldsymbol{H}_{part,0}$*, as* $n \to \infty$*, we have*

$$\frac{T_{n,p_2} - 1}{\sqrt{2(1-\rho_1)/n\rho_2(1-\rho_1-\rho_2)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

The asymptotic normality of the standardized test statistic provides the testing procedure. Given an α-level of significance, $\mathbf{H}_{part,0}$ is rejected when

$$\frac{T_{n,p_2} - 1}{\sqrt{2(1-\rho_1)/n\rho_2(1-\rho_1-\rho_2)}} > z_\alpha,$$

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

4.2. Asymptotic power function

We are now in a position to study the asymptotic power of the test. We first divide $\mathbf{\Gamma} = (\mathbf{\Gamma}_1^\top, \mathbf{\Gamma}_2^\top)^\top$ with $\mathbf{\Gamma}_1 \in \mathbb{R}^{p_1 \times m}$ and $\mathbf{\Gamma}_2 \in \mathbb{R}^{p_2 \times m}$. Define $\mathbf{\Sigma}_{11} = \mathbf{\Gamma}_1\mathbf{\Gamma}_1^\top$, $\mathbf{\Sigma}_{22} = \mathbf{\Gamma}_2\mathbf{\Gamma}_2^\top$, $\mathbf{\Sigma}_{12} = \mathbf{\Gamma}_1\mathbf{\Gamma}_2^\top$, and $\mathbf{\Sigma}_{21} = \mathbf{\Gamma}_2\mathbf{\Gamma}_1^\top$. Following the same idea as in Section 3, we give additional assumptions to facilitate our analysis.

Assumption S4. $\boldsymbol{\beta}_2^\top \mathbf{\Sigma}_{22}\boldsymbol{\beta}_2 = o(1)$ and $\boldsymbol{\beta}_2^\top \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\boldsymbol{\beta}_2 = o(1)$.

Assumption S5. The m-variate random vector $\mathbf{z}_i = (z_{i1}, \ldots, z_{im})^\top$ has a Lebesgue density $f_{\mathbf{z}}$ and satisfies $E(\mathbf{z}_i) = \mathbf{0}$ and $Var(\mathbf{z}_i) = \mathbf{I}_m$. For $j = 1, \ldots, m$, the components z_{ij} are assumed to be independent, satisfy $E(z_{ij}^{20}) \leq C$ for a constant C, and have a marginal density bounded by a constant $D \geq 1$.

Define $\mathbf{V} = \text{diag}(\mathbf{I}_{p_1}, \mathbf{P}_{k_2})$ and $\boldsymbol{\gamma} = (\mathbf{V}^\top\mathbf{\Sigma}\mathbf{V})^{-1}\mathbf{V}^\top\mathbf{\Sigma}\boldsymbol{\beta}$. We write the p-dimensional vector $\mathbf{V}\boldsymbol{\gamma} = (\boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top)^\top$ with $\boldsymbol{\xi}_1 \in \mathbb{R}^{p_1}$ and $\boldsymbol{\xi}_2 \in \mathbb{R}^{p_2}$. Let $\tau_k^2 = \sigma^2 + \boldsymbol{\beta}^\top\mathbf{\Sigma}\boldsymbol{\beta} - \boldsymbol{\gamma}^\top\mathbf{V}^\top\mathbf{\Sigma}\mathbf{V}\boldsymbol{\gamma}$. We derive the asymptotic power function of the proposed test.

Theorem 4. *Under Assumptions* S1–S5*, we have*

$$\Psi_{n,p_2}^{RP}(\boldsymbol{\beta}_2; \boldsymbol{P}_{k_2}) - \Phi\left(-z_\alpha + \sqrt{\frac{n(1-\rho_1-\rho_2)(1-\rho_1)}{2\rho_2}}\frac{\boldsymbol{\xi}_2^\top(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})\boldsymbol{\xi}_2}{\tau_k^2}\right)$$
$$\to 0,$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and z_α is the upper α-quantile of Φ.

Note that no extra assumption is made for $\mathbf{\Sigma}$. From the expression of the asymptotic power function, we can see that the product $\boldsymbol{\xi}_2^\top(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})\boldsymbol{\xi}_2$ is preferred to be larger, is dependent on $\mathbf{P}_{k_2}$, and is upper bounded by $\boldsymbol{\beta}_2^\top(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12})\boldsymbol{\beta}_2$. We give a sufficient condition such that the upper bound can be reached.

Assumption S6. There exist an integer $s_2 < k_2$ and a real number $\gamma_2 > 0$ such that $||\boldsymbol{\beta}_2||_2^2 \sum_{i=s_2+1}^{p_2} d_i = o(p_2 n^{-0.5-\gamma_2})$, where d_i are the eigenvalues of $\mathbf{\Sigma}_{22}$ satisfying $d_1 \geq d_2 \geq \cdots \geq d_{p_2} \geq 0$.

This assumption ensures Lemma 2 is valid for $\boldsymbol{\beta}_2$ and $\mathbf{\Sigma}_{22}$, leading to a negligible distance between the vector $\mathbf{\Gamma}^\top\boldsymbol{\beta}$ and the space generated by $\mathbf{\Gamma}^\top\mathbf{V}$. In this case, we denote the power function of the proposed RP test T_{n,p_2} as $\Psi_{n,p_2}^{RP}(\boldsymbol{\beta}_2)$.

Corollary 2. *Under Assumptions* S1–S6*, we have*

$$\Psi_{n,p_2}^{RP}(\boldsymbol{\beta}_2) - \Phi\left(-z_\alpha + \sqrt{\frac{n(1-\rho_1-\rho_2)(1-\rho_1)}{2\rho_2}}\frac{\boldsymbol{\beta}_2^\top(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})\boldsymbol{\beta}_2}{\sigma^2}\right) \to 0,$$

where $\Phi(\cdot)$ *is the cumulative distribution function of the standard normal distribution, and* z_α *is the upper* α*-quantile of* Φ*.*

5. Numerical Studies

5.1. Simulation studies

We conduct simulations to evaluate the finite-sample performance of the proposed tests and compare it with that of the RCV test.

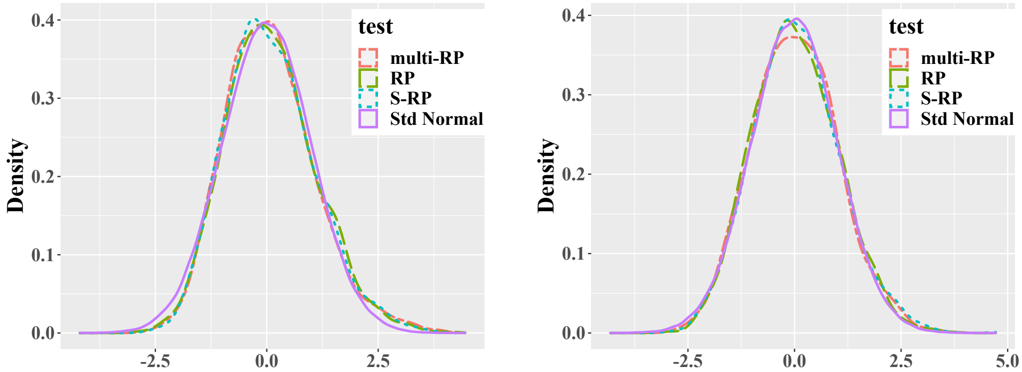
The first simulation study was designed to test: $\mathbf{H}_0 : \boldsymbol{\beta} = \mathbf{0}$ versus $\mathbf{H}_1 : \boldsymbol{\beta} \neq \mathbf{0}$ in the linear regression model

$$y_i = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i.$$

Set $\alpha = 2$. Suppose that ϵ_i is generated from $\mathcal{N}(0,1)$ or $t(5)/\sqrt{5/3}$, and the covariate $\mathbf{x}_i$ is generated from $\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z}_i$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\top$, with μ_i generated from $U(2,3)$ independently, and $\mathbf{z}_i = (z_{i1}, \ldots, z_{ip})^\top$ is generated as (i) $\mathcal{N}(\mathbf{0}, \mathbf{I}_p)$ or (ii) $z_{ij} \overset{i.i.d.}{\sim} U(-\sqrt{3}, \sqrt{3})$. The matrix $\boldsymbol{\Sigma}^{1/2}$ is generated by $\mathbf{U}\sqrt{\mathbf{D}}\mathbf{U}^\top$, where $\mathbf{U}$ is an orthogonal matrix generated from the uniform distribution on the $p \times p$ orthogonal group with the ith column denoted by $\mathbf{u}_i$ and $\sqrt{\mathbf{D}} = \mathrm{diag}(\sqrt{d_1}, \ldots, \sqrt{d_p})$. Let $s = [n^{0.72}]$ and $L = [n^{0.8}]$. The function $[x]$ takes the greatest integer less than or equal to the number x. To achieve the tail eigenvalue condition, we set $d_i = 1$, for $i \leq s$, and $d_i = (L-s)(w_i/W)$, for $i = s+1, \ldots, p$, where $w_i = 1/(i-s)^4$ and $W = \sum_{i=s+1}^p w_i$. Under the alternative hypothesis, the regression coefficient $\boldsymbol{\beta}$ is randomly selected from $Span\{\mathbf{u}_1, \ldots, \mathbf{u}_{s+M}\}$, with $||\boldsymbol{\beta}||_2^2$ taking 0.1, 0.2, and 0.3. In the simulations, we consider $M = 0$ and $M = 50$. Working under high-dimensional settings, we set $(n,p) = (300, 3000), (400, 5000)$, and $(800, 5000)$.

In the simulations, we implemented three types of RP tests according to the choice of random projection: (i) RP test: applying a normal random projection; (ii) multi-RP test: independently generating a normal random projection 10 times and using their mean; (iii) S-RP test: applying the sparse random projection defined in (2.7) with $l = 400$.

We first report the kernel density estimation of the proposed test statistics under $\mathbf{H}_0$ in Figures 1a and 1b, showing that the asymptotic null distribution of



(a) Norm $\mathbf{z}$, norm ϵ and $(n, p) = (300, 3000)$. (b) Norm $\mathbf{z}$, norm ϵ and $(n, p) = (800, 5000)$.

Figure 1. The kernel density estimation of the RP, multi-RP, and S-RP tests under $\mathbf{H}_0$.

the proposed tests can be well approximated by the standard normal distribution. Here, we chose $\rho = 0.4$. The good resemblance to the normal distribution confirms the theoretical result in Theorem 1.

Tables 1 and 2 report the empirical power and sizes of the proposed tests and the RCV test for ϵ distributed from $\mathcal{N}(0, 1)$ and $\sqrt{3/5}t(5)$, based on 2,000 simulations. There are negligible difference between the performance of the three proposed tests, which confirms the discussion in Section 2 and suggests the feasible usage of a different random projection in the test. The empirical sizes of the proposed tests and the RCV test are close to 0.05 under the null hypothesis. The empirical power of the proposed tests are decreasing functions of ρ, which is consistent with the result in Theorem 2. Moreover, we can see that the power of the tests are increasing functions of the norm of $\boldsymbol{\beta}$. Compared with the RCV test, the proposed tests are more powerful.

In the second simulation study, we consider the problem of testing the partial regression coefficients in the linear model. The results are given in Figure 1 and Table 1 in the Supplementary Material. In the third simulation, we conducted a numerical comparison with the LWT test and LDFF test proposed in Lan, Wang and Tsai (2014) and Lan et al. (2016), respectively. The simulation results are given in Table 2 in the Supplementary Material, and show that our proposed test is applicable in highly correlated settings and has higher testing power than that of competing tests.

5.2. Illustrative examples

To illustrate the proposed methods, we consider two examples.

Table 1. Empirical power of the RP, multi-RP, S-RP, and RCV tests at the significance level 0.05 when $\epsilon \sim \mathcal{N}(0,1)$.

M	ρ	$\|\|\boldsymbol{\beta}\|\|_2^2$	$Z \sim U(-\sqrt{3}, \sqrt{3})$				$Z \sim \mathcal{N}(0,1)$			
			RP	multi-RP	S-RP	RCV	RP	multi-RP	S-RP	RCV
$(n,p) = (300, 3000)$										
	0.2	0	0.062	0.066	0.061	0.065	0.062	0.062	0.060	0.062
	0.4	0	0.064	0.065	0.069	0.065	0.069	0.065	0.064	0.062
0	0.2	0.1	0.637	0.623	0.637	0.120	0.647	0.655	0.654	0.120
		0.2	0.956	0.954	0.954	0.188	0.961	0.960	0.959	0.195
		0.3	0.998	0.996	0.998	0.310	0.999	0.998	0.998	0.327
	0.4	0.1	0.437	0.440	0.439	0.120	0.442	0.435	0.441	0.120
		0.2	0.822	0.837	0.834	0.188	0.833	0.836	0.838	0.195
		0.3	0.971	0.968	0.974	0.310	0.976	0.971	0.974	0.327
50	0.2	0.1	0.402	0.389	0.392	0.095	0.382	0.374	0.381	0.095
		0.2	0.762	0.748	0.755	0.135	0.770	0.754	0.755	0.144
		0.3	0.926	0.929	0.933	0.190	0.940	0.942	0.936	0.191
	0.4	0.1	0.276	0.272	0.276	0.095	0.268	0.252	0.247	0.095
		0.2	0.555	0.559	0.546	0.135	0.547	0.544	0.542	0.144
		0.3	0.781	0.780	0.779	0.190	0.783	0.779	0.785	0.191
$(n,p) = (400, 5000)$										
	0.2	0	0.067	0.062	0.066	0.068	0.067	0.065	0.065	0.069
	0.4	0	0.068	0.065	0.064	0.068	0.061	0.065	0.062	0.069
0	0.2	0.1	0.788	0.794	0.784	0.120	0.797	0.794	0.796	0.126
		0.2	0.993	0.992	0.992	0.202	0.992	0.992	0.991	0.204
		0.3	1.000	1.000	1.000	0.333	1.000	1.000	1.000	0.335
	0.4	0.1	0.521	0.519	0.529	0.120	0.527	0.515	0.513	0.126
		0.2	0.906	0.912	0.915	0.202	0.919	0.914	0.910	0.204
		0.3	0.995	0.994	0.996	0.333	0.992	0.994	0.992	0.335
50	0.2	0.1	0.585	0.572	0.587	0.341	0.599	0.593	0.595	0.357
		0.2	0.939	0.941	0.943	0.585	0.942	0.946	0.941	0.593
		0.3	0.993	0.994	0.994	0.758	0.996	0.998	0.997	0.771
	0.4	0.1	0.362	0.364	0.382	0.341	0.366	0.360	0.359	0.357
		0.2	0.742	0.741	0.744	0.585	0.747	0.748	0.743	0.593
		0.3	0.931	0.937	0.939	0.758	0.942	0.941	0.942	0.771
$(n,p) = (800, 5000)$										
	0.2	0	0.057	0.058	0.057	0.068	0.058	0.052	0.056	0.062
	0.4	0	0.058	0.057	0.057	0.068	0.059	0.059	0.059	0.062
0	0.2	0.1	0.959	0.959	0.958	0.145	0.951	0.957	0.954	0.127
		0.2	1.000	1.000	1.000	0.229	1.000	1.000	1.000	0.201
		0.3	1.000	1.000	1.000	0.383	1.000	1.000	1.000	0.345
	0.4	0.1	0.745	0.763	0.747	0.145	0.758	0.763	0.753	0.127
		0.2	0.992	0.994	0.995	0.229	0.993	0.993	0.993	0.201
		0.3	1.000	1.000	1.000	0.383	1.000	1.000	1.000	0.345
50	0.2	0.1	0.849	0.839	0.841	0.325	0.857	0.858	0.866	0.332
		0.2	0.999	0.999	0.999	0.583	1.000	1.000	0.999	0.596
		0.3	1.000	1.000	1.000	0.778	1.000	1.000	1.000	0.792
	0.4	0.1	0.554	0.551	0.551	0.325	0.568	0.559	0.562	0.332
		0.2	0.947	0.951	0.951	0.583	0.955	0.952	0.952	0.596
		0.3	0.997	0.998	0.999	0.778	0.998	0.999	0.997	0.792

Table 2. Empirical power of the RP, multi-RP, S-RP, and RCV tests at the significance level 0.05 when $\epsilon \sim \sqrt{3/5}t(5)$.

M	ρ	$\|\|\boldsymbol{\beta}\|\|_2^2$	$Z \sim U(-\sqrt{3}, \sqrt{3})$				$Z \sim \mathcal{N}(0,1)$			
			RP	multi-RP	S-RP	RCV	RP	multi-RP	S-RP	RCV
$(n,p) = (300, 3000)$										
	0.2	0	0.062	0.059	0.055	0.064	0.052	0.060	0.063	0.066
	0.4	0	0.063	0.062	0.062	0.064	0.066	0.066	0.071	0.066
0	0.2	0.1	0.637	0.646	0.640	0.118	0.639	0.648	0.648	0.117
		0.2	0.947	0.935	0.952	0.120	0.951	0.958	0.956	0.204
		0.3	0.993	0.994	0.995	0.326	0.995	0.996	0.995	0.332
	0.4	0.1	0.452	0.441	0.431	0.118	0.464	0.459	0.463	0.117
		0.2	0.829	0.821	0.834	0.120	0.826	0.829	0.835	0.204
		0.3	0.971	0.967	0.966	0.326	0.968	0.969	0.967	0.332
50	0.2	0.1	0.381	0.381	0.380	0.097	0.390	0.390	0.400	0.095
		0.2	0.757	0.762	0.754	0.142	0.753	0.748	0.735	0.132
		0.3	0.931	0.927	0.926	0.190	0.925	0.925	0.921	0.182
	0.4	0.1	0.271	0.260	0.262	0.097	0.273	0.272	0.273	0.095
		0.2	0.539	0.538	0.548	0.142	0.547	0.552	0.554	0.132
		0.3	0.788	0.780	0.778	0.190	0.778	0.783	0.789	0.182
$(n,p) = (400, 5000)$										
	0.2	0	0.066	0.061	0.060	0.071	0.065	0.067	0.064	0.066
	0.4	0	0.071	0.062	0.064	0.071	0.064	0.064	0.063	0.066
0	0.2	0.1	0.788	0.788	0.798	0.124	0.790	0.785	0.796	0.131
		0.2	0.993	0.990	0.991	0.215	0.991	0.993	0.993	0.208
		0.3	1.000	1.000	1.000	0.351	1.000	1.000	1.000	0.349
	0.4	0.1	0.533	0.535	0.523	0.124	0.533	0.533	0.548	0.131
		0.2	0.914	0.913	0.909	0.215	0.905	0.911	0.909	0.208
		0.3	0.992	0.993	0.992	0.351	0.991	0.993	0.992	0.349
50	0.2	0.1	0.588	0.599	0.592	0.345	0.589	0.596	0.596	0.361
		0.2	0.937	0.939	0.940	0.592	0.942	0.946	0.947	0.608
		0.3	0.995	0.993	0.994	0.757	0.997	0.997	0.998	0.758
	0.4	0.1	0.372	0.388	0.367	0.345	0.367	0.359	0.373	0.361
		0.2	0.757	0.750	0.740	0.592	0.738	0.741	0.754	0.608
		0.3	0.932	0.936	0.936	0.757	0.935	0.939	0.934	0.758
$(n,p) = (800, 5000)$										
	0.2	0	0.060	0.057	0.059	0.061	0.051	0.054	0.051	0.066
	0.4	0	0.058	0.055	0.059	0.061	0.061	0.055	0.051	0.066
0	0.2	0.1	0.957	0.955	0.957	0.128	0.961	0.962	0.960	0.127
		0.2	1.000	0.999	1.000	0.212	1.000	1.000	1.000	0.201
		0.3	1.000	1.000	1.000	0.366	1.000	1.000	1.000	0.349
	0.4	0.1	0.742	0.737	0.744	0.128	0.757	0.755	0.757	0.127
		0.2	0.995	0.992	0.991	0.212	0.994	0.992	0.993	0.201
		0.3	1.000	1.000	1.000	0.366	1.000	1.000	1.000	0.349
50	0.2	0.1	0.837	0.832	0.834	0.338	0.866	0.864	0.868	0.345
		0.2	0.999	0.999	0.998	0.587	0.999	0.999	0.999	0.596
		0.3	1.000	1.000	1.000	0.791	1.000	1.000	1.000	0.784
	0.4	0.1	0.541	0.551	0.553	0.338	0.562	0.573	0.571	0.345
		0.2	0.947	0.942	0.942	0.587	0.948	0.940	0.948	0.596
		0.3	0.999	0.998	0.998	0.791	0.997	0.999	0.998	0.784

Table 3. The p-values of the proposed tests and the RCV test for example 1.

	$\mathbf{H}_0 : \boldsymbol{\beta} = \mathbf{0}$ vs $\mathbf{H}_1 : \boldsymbol{\beta} \neq \mathbf{0}$				$\mathbf{H}_{part,0} : \boldsymbol{\beta}_2 = \mathbf{0}$ vs $\mathbf{H}_{part,1} : \boldsymbol{\beta}_2 \neq \mathbf{0}$		
Tests	RP	multi-RP	S-RP	RCV	RP	multi-RP	S-RP
p-value	0.00	0	0.00	0.00	0.73	0.61	0.65

5.2.1. Example 1

Here, We consider a real data set of riboflavin (vitamin B2) production by bacillus subtilis. The data were analyzed by van de Geer et al. (2014) and are available in the R package "hdi." The real-valued response variable is the logarithm of the riboflavin production rate, and there are $p = 4{,}088$ covariates (genes) measuring the logarithm of the expression levels of 4,088 genes. These measurements are from $n = 71$ samples of genetically engineered mutants of bacillus subtilis. We modeled the data using a high-dimensional linear model. The p-values of the proposed tests and the RCV test are provided in Table 3. All the tests reject the null hypothesis, indicating the considerable significance of gene expressions in predicting the riboflavin production rate.

Then, we were interested in the significance of a partial gene expression. Using the lasso method, we divided the coefficients into two parts, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, where the index of $\boldsymbol{\beta}_2$ corresponds to the index of the zero part in $\hat{\boldsymbol{\beta}}^{Lasso}$. We conducted testing for $\boldsymbol{\beta}_2$ and the results are shown in Table 3. The large p-values indicate that $\mathbf{H}_{part,0}$ is accepted, which is consistent with the lasso result.

5.2.2. Example 2

We applied the proposed tests to a more recent data set, which is available for download under accession number GSE50948 in the Gene Expression Omnibus (GEO). In this data set, gene expression profiling was performed using RNA from $n = 114$ samples of pretreated patients with HER2-positive (HER2+) tumors. Because multiple probes might represent the same gene, the measurement for each gene is from the probe with the highest interquartile range. After a natural logarithm transformation, we obtained expression values of 20,592 genes. Prat et al. (2014) implemented a researched-based prediction analysis of the microarray 50 (PAM50) subtype predictor to the data. They reported that the predominant subtype within HER2+ disease is HER2-enriched (HER2-E) tumors, which have been found to have a high expression of HER2-regulated genes (for example, ERBB2, GRB7, and FGFR4). To gain a better understanding of the HER2-E subtype, we studied the association between HER2-regulated genes and residual genes, with ERBB2 as an example.

Table 4. The p-values of the proposed tests and the RCV test for example 2.

	$\mathbf{H}_0 : \boldsymbol{\beta} = \mathbf{0}$ vs $\mathbf{H}_1 : \boldsymbol{\beta} \neq \mathbf{0}$				$\mathbf{H}_{part,0} : \boldsymbol{\beta}_2 = \mathbf{0}$ vs $\mathbf{H}_{part,1} : \boldsymbol{\beta}_2 \neq \mathbf{0}$		
Tests	RP	multi-RP	S-RP	RCV	RP	multi-RP	S-RP
p-value	0.00	0.00	0.00	0.00	0.42	0.47	0.64

Let the response variable be the gene expression level of ERBB2 and the residual $p = 20{,}591$ gene expression levels be the covariates. Suppose that the data follow a linear model. The RCV test and our proposed tests reported a significant relationship by rejecting the null hypothesis, as shown in Table 4. We moved on to identifying strongly associated genes based on the proposed tests and the lasso estimation. Let the regression coefficient corresponding to the zeros in the lasso estimator be denoted as $\boldsymbol{\beta}_2$. The proposed tests for the testing problem of this partial regression coefficient were conducted. The p-values of the global and partial hypothesis testing in the table suggest that genes with nonzero coefficients, namely, ESR1, MAP4K3, and TLK1, have significant effects on the gene expression of ERBB2, some of which have already been shown to be important to breast cancer. For example, Prat et al. (2014) indicated that a lower expression of the luminal-related gene ESR1 is one of important characteristics of HER2-enriched (HER2-E) tumors. Gamez-Pozo et al. (2014) found the gene expression of MAP4K3 to be related to the PI3K pathway, which is strongly associated with the response to trastuzumab in HER2 breast cancer. Consequently, the new testing procedures can be helpful in confirming existing knowledge and making new discoveries.

6. Conclusion

We have proposed a new testing procedure for hypothesis testing in a high-dimensional linear regression, which involves applying a random projection and then working with the classical F-test. The use of a random projection contributes to the feasible replacement of high-dimensional covariates with their projected versions. Our test is simple, both in form and computation. In addition, we do not assume any explicit relationship between n and p, which indicates that our test accommodates extremely high-dimensional settings. The asymptotic null distribution and power function are derived when $(n, p) \to \infty$. To show the advantage of the new test, we compare it with a powerful high-dimensional test proposed by Cui, Guo and Zhong (2018). We find a sufficient condition that ensures our test outperforms the competing test. Our discussion provides a specific suggestion for the choice of ρ in different situations. Next, we extended the discussion to include testing a partial linear regression coefficient, proposing another test based on the same idea and deriving its asymptotic distribution. Numerical simulations and applications to real data illustrate the finite-sample performance of the proposed tests and demonstrate the feasible use of different random projections.

Supplementary Material

The online Supplementary Material contains proofs of the lemmas and theorems, as well as additional numerical results.

Acknowledgments

The authors are grateful to the editor, associate editor, and two referees for their valuable comments and suggestions. This work was supported in part by the Research Grant Council of Hong Kong (15303319), and The Hong Kong Polytechnic University (P0030124, P0034285).

References

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

Cui, H., Guo, W. and Zhong, W. (2018). Test for high-dimensional regression coefficients using refitted cross-validation variance estimation. *The Annals of Statistics* **46**, 958–988.

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.

Diaconis, P. and Freedman, D. (1984). Asymptotics of graphical projection pursuit. *The Annals of Statistics* **12**, 793–815.

Fan, J., Guo, S. and Hao, N. (2012). Variance estimation using refitted cross-validation in ultrahigh dimensional regression. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)* **74**, 37–65.

Gamez-Pozo, A., Carrión, R. M. P., Manso, L., Crespo, C., Mendiola, C., Lopez- Vacas, R. J. et al. (2014). The Long-HER study: Clinical and molecular analysis of patients with HER2+ advanced breast cancer who become long-term survivors with trastuzumab-based therapy. *PloS one* **9**, e109611.

Huang, C. and Huo, X. (2022). A statistically and numerically efficient independence test based on random projections and distance covariance. *Frontiers in Applied Mathematics and Statistics* **7**, 1–17.

Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research* **15**, 2869–2909.

Lehmann, E. L. (1959). Optimum invariant tests. *The Annals of Mathematical Statistics* **30**, 881–884.

Lan, W., Ding, Y., Fang, Z. and Fang, K. (2016). Testing covariates in high dimension linear regression with latent factors. *Journal of Multivariate Analysis* **144**, 25–37.

Li, P., Hastie, T. J. and Church, K. W. (2006). Very sparse random projections. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 287–296. ACM, New York.

Liu, M., Shang, Z. and Cheng, G. (2018). Nonparametric testing under random projection. *arXiv preprint arXiv:1802.06308.*

Lopes, M., Jacob, L. and Wainwright, M. J. (2011). A more powerful two-sample test in high dimensions using random projection. In *Advances in Neural Information Processing Systems*, 1206–1214.

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

Ma, R., Cai, T. T. and Li, H. (2021). Global and simultaneous hypothesis testing for high-dimensional logistic regression models. *Journal of the American Statistical Association* **116**, 984–998

Ning, Y. and Liu, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *The Annals of Statistics* **45**, 158–195.

Prat, A., Bianchini, G., Thomas, M., Belousov, A., Cheang, M. C., Koehler, A. et al. (2014). Research-based PAM50 subtype predictor identifies higher responses and improved survival outcomes in HER2-positive breast cancer in the NOAH study. *Clinical Cancer Research* **20**, 511–521.

Searle, S. R. and Gruber, M. H. J. (2017). *Linear Models*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken.

Shi, C., Song, R., Chen, Z. and Li, R. (2019). Linear hypothesis testing for high dimensional generalized linear models. *The Annals of Statistics* **47**, 2671–2703.

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

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

Wang, S. and Cui, H. (2013). Generalized F test for high dimensional linear regression coefficients. *Journal of Multivariate Analysis* **117**, 134–149.

Zhu, Y. and Bradic, J. (2018). Linear hypothesis testing in dense high-dimensional linear models. *Journal of the American Statistical Association* **113**, 1583–1600.

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

Zhang, X. Y. and Cheng, G. (2017). Simultaneous inference for high-dimensional linear models. *Journal of the American Statistical Association* **112**, 757–768.

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

Changyu Liu

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China.

E-mail: cy.u.liu@connect.polyu.hk

Xingqiu Zhao

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China.

E-mail: xingqiu.zhao@polyu.edu.hk

Jian Huang

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China.

E-mail: j.huang@polyu.edu.hk

(Received October 2020; accepted June 2021)