HIGH-DIMENSIONAL TESTS BASED ON RANDOM PROJECTION APPROACH

LIU CHANGYU

PhD

The Hong Kong Polytechnic University

2021

The Hong Kong Polytechnic University

Department of Applied Mathematics

High-Dimensional Tests Based on Random Projection Approach

LIU Changyu

A thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

July 2021

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

______________________________(Signed)

LIU Changyu (Name of student)

Abstract

In this thesis, we consider hypothesis testing in high-dimensional models, where the dimension of covariates p is greater than the sample size n, which is common in data analysis currently. Novel test statistics are proposed in both linear model and single-index model for high-dimensional settings.

First, we focus on the simple linear model and propose a novel test statistic for the problem of testing global regression coefficients. The proposed test is constructed based on the technique of random projection. Concretely, we first randomly project high-dimensional data into a lower-dimensional space and then apply the projected data to the classical F-test. The proposed test has a simple form and intuitive interpretation. The advantages of this random-projection-based approach are demonstrated both theoretically and numerically. Under mild conditions, we derive the asymptotic normality and the asymptotic local power functions of the proposed test. By comparison with some recent developed methods, our proposed test shows higher asymptotic relative efficiency in a sufficient condition. The proposed method is further extended to the problems of testing partial regression coefficients and we derive its asymptotic properties. Through simulation studies, we evaluate the finite-sample performances of the proposed tests and demonstrate its superior performance than the competing tests. Applications to real high-dimensional gene expression data are also provided for illustration.

Next, we investigate the single-index model, which includes many commonly used models. First, we study the feasibility of applying the classical F-test to a single-index model where $p/n \to \zeta \in (0, 1)$. We derive its asymptotic null distribution and asymptotic local power function. For the ultrahigh-dimensional single-index model where $p \gg n$, we construct F-statistics based on lower-dimensional random projections of the data. For the hypothesis testing of global and partial parameters in the $p > n$ settings, the asymptotic null distribution and the asymptotic local power function of the proposed test statistics are analyzed. The newly proposed test possesses the advantages of intuitive interpretation and simplified computation. We compare the proposed test with other high-dimensional tests and show our test is more efficient in a sufficient condition. We conduct simulation studies to evaluate the finite-sample performances of the proposed tests and demonstrate that it has higher power than some existing methods in the models we consider. The application of real high-dimensional gene expression data is also provided to illustrate the effectiveness of the method.

Overall, we propose new tests applicable to general models in high-dimensional settings. The proposed tests are easy to implement and would present a reasonable performance under mild conditions. For the testing power, it is shown to hold for a wide range of alternatives and possess certain advantages in sparse cases. As a result, our proposed method provides a practicable choice for hypothesis testing in modern data analysis.

Key Words: High-dimensional inference; Hypothesis testing; Linear model; Single-index model; Random projection.

Acknowledgements

First of all, I would like to express my deepest gratitude to Prof. Zhao Xingqiu and Prof. Huang Jian. This work would not has been possible without their invaluable advice and continued support. During the PhD study, they leaded me into the area of statistics and provided me with constructive suggestions and profound insight, from which I benefited so much. Their levels of knowledge, ingenuity, enthusiasm and rigorous attitude are something I will always keep aspiring to. I would also like to show my sincere gratitude to Prof. Li Jialiang, Prof. Ma Shuangge, and Prof. Li Xun for their valuable suggestions on the improvement of my thesis and the organization of my oral examination.

My deep gratitude also goes to Prof. Lin Yuanyuan as well as every teacher and every student within the reading group for their dedication. Their valuable works, eager discussions and wise comments always inspired and motivated me. I am also grateful to Prof. Jiang Binyan and everyone in the summer studying group. I appreciated every moment of sharing and exchanging idea, from which I learned abundantly. I am appreciated to Dr. Liu Kin-yat, Dr. Hu Xiangbin, Su Wen, and Zheng Yangzi for their generous assistance. I am sincerely grateful to all knowledgeable faculty, friendly staff members and intelligent fellow students in the department of applied mathematics. It is my great honor and luck to have a wonderful study period here.

I would like to extend my sincere thanks to Qin Huihui, Yang Yue, Wang Xiaoqiang and He Qi for their support and constant companion. I would also like to express my special thanks to Yang Ruizhang for his unwavering belief and gentleness, which always gives me encouragement and confidence. And I am appreciated for every help and effort he offered.

Lastly, I would like to express my endless gratitude to my parents for giving me utmost care, nurturing my knowledge, encouraging my interests and most of all, for giving me unconditional love and support. They are the most powerful and secure backing for me.

Contents

List of Figures

List of Tables

List of Notations

Abbreviations

ARE	Asymptotic Relative Efficiency.
ESD	Empirical Spectral Distribution.
GLM	Generalized Linear Model.
LASSO	Least Absolute Shrinkage and Selection Operator.
SIM	Single Index Model.
SVD	Singular Value Decomposition.

Scalars, Vectors and Matrices

a	Scalar.
$\mathbf{a}$	Vector.
$\mathbf{a}^\top$	The transpose of vector $\mathbf{a}$.
$(\mathbf{a})_i$	The entry of vector $\mathbf{a}$ in the i-th coordinate.
$\mathbf{e}_i$	The vector with 1 in the i-th coordinate and 0's elsewhere.
$\|\|\mathbf{a}\|\|_2$	The Euclidean norm of $\mathbf{a}$, i.e. $\|\|\mathbf{a}\|\|_2 = \sqrt{\mathbf{a}^\top \mathbf{a}}$.
$\mathbf{A}$	Matrix.
$\mathbf{A}^\top$	The transpose of matrix $\mathbf{A}$.
$(\mathbf{A})_{ij}$	The entry of matrix $\mathbf{A}$ in the i-th row and the j-th column.
$tr(\mathbf{A})$	The trace of matrix $\mathbf{A}$.
$\|\|\mathbf{A}\|\|_F$	The Frobenius norm of matrix $\mathbf{A}$, i.e. $\|\|\mathbf{A}\|\|_F = \sqrt{tr(\mathbf{A}^\top \mathbf{A})}$.
$\|\|\mathbf{A}\|\|_{sp}$	The spectral norm of matrix $\mathbf{A}$, i.e. $\|\|\mathbf{A}\|\|_{sp} = \max_{\|\|\mathbf{a}\|\|_2=1} \|\|\mathbf{A}\mathbf{a}\|\|_2$.
$\mathbf{A} \succeq \mathbf{0}$	The symmetric matrix $\mathbf{A}$ is positive semi-definite.

$\lambda_{\max}(\mathbf{A})$	The largest eigenvalue of symmetric matrix $\mathbf{A}$.
$\lambda_{\min}(\mathbf{A})$	The smallest eigenvalue of symmetric matrix $\mathbf{A}$.
$\mathbf{I}$ or $\mathbf{I}_p$	The identity matrix of size p.
$\text{diag}(\mathbf{a})$	The diagonal matrix whose entries are the elements of vector $\mathbf{a}$.
$\sqrt{\text{diag}(\mathbf{a})}$	The diagonal matrix whose entries are the square roots of the elements of vector $\mathbf{a}$, where each element is non-negative.
$\text{diag}(\mathbf{A}, \mathbf{B})$	The block diagonal matrix whose diagonal blocks are matrix $\mathbf{A}$ and $\mathbf{B}$.

Functions and Set

f'	The first order derivative of function f.
$f^{(r)}$	The r-th order derivative of function f.
$\lceil x \rceil$	The ceiling function, i.e. $\lceil x \rceil$ is the least integer greater than or equal to x.
$\lfloor x \rfloor$	The floor function, i.e. $\lfloor x \rfloor$ is the greatest integer less than or equal to x.
S	Set
S^c	The complement of set S.
$\|S\|$	The cardinality of set S.
$\mathbb{R}$	The set of real numbers.
$\mathbb{R}^m$	The m-dimensional real space.
$\mathbb{R}^{m \times k}$	The set of all $m \times k$ real matrices.
$Span\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$	The set of all linear combinations of a collection of vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$.
$Span\{\mathbf{A}\}$	The set of all linear combinations of the columns of matrix $\mathbf{A}$.
$\mathcal{V}_k(\mathbb{R}^m)$	The Stiefel manifold defined by $\mathcal{V}_k(\mathbb{R}^m) = \{\mathbf{A} \in \mathbb{R}^{m \times k} : \mathbf{A}^\top \mathbf{A} = \mathbf{I}_k\}$.

Moments and Operators

$E(X)$	Expectation of random variable X.
$E(X\|Y)$	Conditional expectation of random variable X given Y.

$Var(X)$	Variance of random variable X.
$Var(X\|Y)$	Conditional variance of random variable X given Y.
$E(X^r)$	The r-th moment of random variable X.
$Cov(X,Y)$	Covariance of random variables X and Y.

Distributions

$\mathcal{N}(\mu,\sigma^2)$	The normal distribution with mean μ and standard deviation σ.
$\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$	The multivariate normal distribution with p-dimensional mean vector $\boldsymbol{\mu}$ and $p \times p$ covariance matrix $\boldsymbol{\Sigma}$.
$t(n)$	The t-distribution with n degrees of freedom.
$F(m,n)$	The F-distribution with m and n degrees of freedom.
$\Phi(\cdot)$	The cumulative distribution function of the standard normal distribution.
$\nu_{m,k}$	The uniform distribution on the Stiefel manifold $\mathcal{V}_k(\mathbb{R}^m)$.

Others

$\xrightarrow{a.s.}$	Almost sure convergence.
$\xrightarrow{p}$	Convergence in probability.
$\xrightarrow{\mathcal{D}}$	Convergence in distribution.
$o(1)$	A sequence of non-random vectors $\mathbf{x}_n$ is said to be $o(1)$ when $\mathbf{x}_n$ converges to zero.
$O(1)$	A sequence of non-random vectors $\mathbf{x}_n$ is said to be $O(1)$ when $\mathbf{x}_n$ is bounded.
$o_p(1)$	A sequence of random variables X_n is said to be $o_p(1)$ when $X_n \xrightarrow{p} 0$.
$O_p(1)$	A sequence of random variables X_n is said to be $O_p(1)$ when X_n is uniformly tight.

Chapter 1

Introduction

1.1 Background

In the modern data analysis, there will always be a typical situation where the dimension of covariates p is much greater than the sample size n. For example, in genomic studies, the dimension of data such as gene expression and genetic marker data are typically far greater than the sample size. This phenomenon brings challenges to the classical statistical testing procedures, even in many basic settings. For example, the Hotelling T^2 statistic for the two-sample testing problem cannot be well-defined when p is greater than n, since the sample covariance matrix is no longer invertible in this setting. Similarly, in high-dimensional regression models, many existing methods for statistical inference about regression coefficients are no longer applicable. Therefore, it is important to develop new testing procedures in high-dimensional models. In this thesis, we mainly focus on linear model and single-index model for their representativeness and ubiquitous application. This section will introduce their basic forms and problems in high-dimensional settings.

1.1.1 Linear Model

We first consider a linear regression model, given as

$$y = \alpha + \mathbf{x}^\top \boldsymbol{\beta} + \epsilon, \tag{1.1}$$

where y is a response variable, $\mathbf{x}$ 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 is a random error term with mean zero and variance σ^2. We are interested in testing the hypothesis

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

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 covariates. Under conditions $p < n$ and $y|\mathbf{x} \sim \mathcal{N}(\alpha + \mathbf{x}^\top \boldsymbol{\beta}, \sigma^2)$, the exact distribution of the F-test is known and has certain optimal properties, since it can be considered as 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 would have a poor performance as the result of increasing covariate dimension.

1.1.2 Single-Index Model

For the single-index model (SIM), it is defined as

$$y = f(\mathbf{x}^\top \boldsymbol{\theta}, \epsilon), \tag{1.3}$$

where y is a response variable, $\mathbf{x}$ is a $p \times 1$ covariate vector, $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown coefficients, ϵ is a random error independent of $\mathbf{x}$, and f is an unspecified link function. Many commonly used parametric and semi-parametric models are included in SIM, such as linear, generalized linear and Cox models. However, the high flexibility of model assumption increases the difficulty to make statistical inference, especially in high-dimensional problems. In the classical settings, Li and Duan

(1989) demonstrated that maximum likelihood-type estimators are consistent for $\boldsymbol{\theta}$ up to a scalar, even though a misspecified link function f might be assumed in model (1.3). In addition, with the assumption that $\mathbf{x}$ had an elliptically symmetric distribution, Li and Duan (1989) showed that the Wald test as well as the likelihood ratio test were workable. Recently, this approach was considered in the usage of LASSO to the nonlinear model (1.3) by Neykov, Liu, and Cai (2016) and Thrampoulidis, Abbasi, and Hassibi (2015), where effective sparse recovery and explicit expressions for the mean-squared-error were obtained, respectively. In addition, Thrampoulidis, Abbasi, and Hassibi (2015) illustrated that the estimation performance of the generalized LASSO in the nonlinear model is asymptotically the same as that of the linear model. However, there is little systematic research on hypothesis testing for high-dimensional SIM at this stage.

1.2 Literature Review

In recent years, many efforts have been devoted to the problems of hypothesis testing in high-dimensional models. For the linear regression model, Zhong and Chen (2011) proposed a test based on a U-statistic of order four and extended it to accommodate factorial designs. This approach was further considered in Cui, Guo, and Zhong (2018) by implementing a new variance estimation method of Fan, Guo, and Hao (2012). Specifically, Cui, Guo, and Zhong (2018) constructed the proposed test from an estimated U-statistics of order two and applied the refitted cross-validation approach for variance estimation to reduce bias. The numerical comparison with the method in Zhong and Chen (2011) indicated the advantages of the proposed method, which were reflected in its less running time and higher empirical powers. From another direction, Lan, Wang, and Tsai (2014) and Lan et al. (2016) proposed novel tests motivated by the limited application range of the method in Zhong and

Chen (2011). Specifically, Lan, Wang, and Tsai (2014) proposed a test for general random design. And Lan et al. (2016) focused on the covariate generated from a latent factor structure, where high correlation among covariates could lead to violation of one critical assumption in Zhong and Chen (2011). The above methods were mainly developed for the testing problems of global or large proportional regression coefficients. In Zhang and Cheng (2017), a more general testing problem was considered. It included the testing problem of any subset of the regression coefficients. For this problem, Zhang and Cheng (2017) introduced a bootstrap-assisted testing procedure, where the test statistic was a maximal-absolute-error type constructed from the debiased estimation and the critical value was built from a multiplier bootstrap method. The proposed method was workable on a sparse setting and was very sensitive in detecting sparse alternatives. In Zhang and Zhang (2014), the testing problems of single or finite number of regression coefficients was studied.

For a more general model, the generalized linear model (GLM), Goeman, van Houwelingen, and Finos (2011) proposed a global test based on empirical Bayes, where p can be greater than n. Later, the feasibility of this test with diverging p was proved in Guo and Chen (2016), and the test was modified to gain more power. Applying the debiased method, Van de Geer et al. (2014) considered the testing problems of single or finite number of regression coefficients. And Ma, Cai, and Li (2020) constructed a maximal-absolute-error-type test statistic for the global testing problem in the high-dimensional logistic regression model.

As mentioned above, testing methods constructed from the debiased method were extensively developed. This kind of estimation method was first proposed in Zhang and Zhang (2014) and was demonstrated to possess various advantages. One of its critical merits was its ability to make direct statistical inference. Compared with other variable-selection-based approaches, the method was applicable in a more general setting. Considering the testing approaches based on the debiased method, statistical tests for single or low-dimensional components of the regression coefficients in the linear regression models (Zhang and Zhang , 2014), GLMs (Van de Geer et al. , 2014) and general models (Ning and Liu , 2017) were developed. For the high-dimensional global testing problem, Zhang and Cheng (2017) and Ma, Cai, and Li (2020) constructed maximal-absolute-error-type test statistics in the linear and logistic regression models, respectively. However, for the need of available estimators, these methods all required certain strong sparsity condition. It leaded to a problem of whether the application setting satisfied the sparsity condition, which became particularly significant in practical data analysis. Recently, related problems have received attention. For example, Zhu and Bradic (2018) and Bradic, Fan, and Zhu (2018) investigated testing problem when sparsity condition might be absent. Carpentier and Verzelen (2021) studied the problem of testing sparsity of the regression coefficients.

Another widely applied technique was the random projection. There were many testing methods developed from it, including independence testing (Huang and Huo , 2017), two-sample testing (Lopes, Jacob, and Wainwright , 2011) and nonparametric testing (Liu, Shang, and Cheng , 2018). An important advantage of the methods based on random projection was its ability to preserve the significant information in data while reducing its dimension. This made the random projection appealing, especially in high dimensions.

At last, we would like to provide a general comparison. Specifically, the test statistics for the testing problem of global regression coefficients were investigated. Generally, there were two common types of testing approaches: one was the sum-of-squares-type test statistics, such as Zhong and Chen (2011) and Cui, Guo, and Zhong (2018), and the other was maximal-absolute-error-type test statistic, such as Zhang and Cheng (2017) and Ma, Cai, and Li (2020). Concerned with the application range, the second type of testing approaches required sparsity restriction on the setting. This condition might be rigorous in certain situations, and the performance of the test statistics could be uncertain when the condition was absent. For the power, these two types of testing approaches showed the advantages in different alternatives. When there was a large proportion of small-to-moderate nonzero regression coefficients, the first type of testing approaches tended to be more powerful. By contrast, when there were only few nonzero regression coefficients, the second type of testing approaches might show stronger testing power. However, when the sparsity condition was unknown in prior, the selection of the type of test statistic become important and it motivated us to consider an applicable choice to provide reasonable performance for both types of alternatives. In addition, test statistics for both types were relatively complicated in form and had certain computation complexity.

1.3 Motivation and Outline

As shown above, there are still many problems to be solved in high-dimensional hypothesis testing problem. Motivated by this, we aim to develop novel testing methods to make some developments for the topic.

For the global hypothesis testing problem in the linear regression model, the existing test statistics are relatively complicated in form and tend to be computationally expensive. This motivates us to propose a novel test statistic that has a simple form and is easy to compute. Specifically, we consider the technique of random projection. This method has been widely implemented in many fields, such as electrical and electronic engineering as well as computer science, and it is recently investigated in statistics. An important advantage of the approach based on random projection comes from its ability to preserve the significant information in data while reducing dimensionality. In addition, according to the results that almost all low-dimensional projection data is close to normal, some theoretical results of the normal distribution is applicable to randomly projected data. According to the above ideas, we apply the technique of random projection to reduce dimension of data and construct F-statistics based on projected data which live in lower-dimensional space. Our proposed test statistic has no explicit restriction on the relationship between n and p, which makes it accommodate extremely high-dimensional settings. And it is shown to be applicable in a general situation under some mild conditions, where no sparsity restriction is required.

Since the linearity assumption of the linear regression model has certain limitation and may be violated in several practical cases, there is a vast literature devoted to developing statistical inference methods that are tailored for nonlinear models. However, according to the results in Li and Duan (1989), as shown in Section 1.1.2, it motivates us to reconsider the possibility of applying the statistical inference methods of linear model to nonlinear models. We study the F-statistic in a relatively high-dimensional SIM, where $p/n \to \zeta$ with $\zeta \in (0, 1)$, and we demonstrate its feasibility by deriving its asymptotic normality and asymptotic local power function. Combining this idea with the technique of random projection, we propose novel test statistics for the testing problems in high-dimensional SIM when $p > n$. The proposed test is simple in form and is easy to compute. And it has a wide application range in terms of general model assumption and mild conditions on the distribution.

The reminder of the thesis is organized as follows.

In Chapter 2, we investigate the testing problems in the high-dimensional linear regression models. Specifically, we propose novel test statistics for the testing problems of global and partial regression coefficients and derive the asymptotic normality and asymptotic local power functions. By comparison with some recent developed methods, our proposed test shows large asymptotic relative efficiency in a sufficient condition. Through simulation studies, we evaluate the finite-sample performances of the proposed tests and demonstrate that it has stronger testing power than the competing tests in considered models. Applications to real high-dimensional gene expression data are also provided for illustration.

In Chapter 3, we study the problems of hypothesis testing in high-dimensional SIMs. First, we demonstrate the feasibility of applying the classical F-test to a SIM in the $p/n \to \zeta \in (0,1)$ regime. For the ultrahigh-dimensional SIM in the $p \gg n$ settings, we construct F-statistics based on lower-dimensional random projections of the data. We derive the asymptotic null distribution and the asymptotic local power function of the proposed test statistics for the hypothesis testing of global and partial regression coefficients in the $p > n$ settings. We compare the proposed tests with other high-dimensional tests and provide sufficient conditions under which the proposed tests are more efficient. According to the simulation studies, the finite-sample performance of the proposed tests is evaluated. The simulation results also indicate higher power of our proposed tests than some existing methods in the models we consider. The application of real high-dimensional gene expression data is also provided to illustrate the effectiveness of the method.

A brief summary of the thesis is provided in Chapter 4, where we also make a short discussion for future research.

Chapter 2

New Tests for High-Dimensional Linear Regression Based on Random Projection

2.1 Introduction

In this chapter, we consider a linear regression model

$$y = \alpha + \mathbf{x}^\top \boldsymbol{\beta} + \epsilon, \tag{2.1}$$

where y is a response variable, $\mathbf{x}$ 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 is a random error term with mean zero and variance σ^2. We focus on the high-dimensional settings when p can exceed the sample size n. We are interested in testing the hypothesis

$$\mathbf{H}_0 : \boldsymbol{\beta} = \mathbf{0} \quad \text{versus} \quad \mathbf{H}_1 : \boldsymbol{\beta} \neq \mathbf{0}. \tag{2.2}$$

We propose a new statistical test for hypothesis (2.2) in high-dimensional settings. Using the technique of random projection to reduce the data dimension, we construct F-statistics based on projected data which live in lower-dimensional space. The F-test based on projected data has a simple form and is easy to compute. An important advantage of random-projection-based approach stems from its ability in dimension reduction while preserving the significant information in data simultaneously. The proposed test is shown to be applicable in a general situation under some mild conditions. The usage of random projection injects extra randomness to 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 that almost all low dimensional projection data is close to normal, according to Diaconis and Freedman (1984). Under the null hypothesis, it is shown 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 is normal, and demonstrate the benefit of the using random projection for reducing the dimension of the data. Finally, we extend the proposed random-projection-based test procedure for the global hypothesis (2.2) to the problem of testing partial regression coefficients and derive its asymptotic null distribution and local power function.

The rest of this chapter is organized as follows. In Section 2.2, we propose our test statistic and give the intuition for its design. In Section 2.3, 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 in comparison with some recent tests. In Section 2.4, we extend the proposed test to the problem of testing partial regression coefficients and establish its asymptotic theoretical results. In Section 2.5.1, we conduct simulation studies to evaluate the finite-sample behavior of the proposed test in terms of type I error and power, and compare it with the competing tests. We also illustrate its applications to high-dimensional gene expression data sets in Section 2.5.2. The proofs of lemmas and theorems are relegated to Section 2.6.

2.2 Test Statistic

Suppose that $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ are i.i.d. copies of $(\mathbf{x}, y)$ from the linear regression model (2.1). Let $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ be the i-th row of the design matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top$ and $\mathbf{y} = (y_1, \ldots, y_n)^\top$. It follows that $\mathbf{y}$ and $\mathbf{X}$ satisfy

$$\mathbf{y} = \alpha\mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2.3}$$

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 regression in the $n > p$ settings. For simplicity, we consider the model without intercept

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{2.4}$$

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 $\mathbf{H}_0 : \boldsymbol{\beta} = \mathbf{0}$ is

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

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, F-test is the most widely used methods for testing hypothesis about regression coefficients in linear models and enjoys certain optimality properties. In addition, it has a known finite-sample distribution and it is uniformly most powerful invariant (Lehmann, 1959). Clearly, the F-test in (2.5) is not applicable to high-dimensional data with $n < p$.

To overcome this difficulty, we first project high-dimensional predictors onto a lower-dimensional space, and then apply the F-test to the projected data. Specifically, for an integer $1 \leq k < \min\{n, p\}$, let $\mathbf{P}_k \in \mathbb{R}^{p \times k}$ denote a random projection matrix with random entries, drawn independently of 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*

$$\mathbf{y} = \mathbf{U}_k \boldsymbol{\eta} + \boldsymbol{\epsilon}. \tag{2.6}$$

We use this model to motivate the proposed test statistic. Of course, model (2.6) is generally different from model (2.4). However, for the purpose of constructing a valid test, it suffices that the null hypothesis $\mathbf{H}_0 : \boldsymbol{\beta} = \mathbf{0}$ under model (2.4) is equivalent to the null hypothesis $\mathbf{H}_0 : \boldsymbol{\eta} = \mathbf{0}$ under (2.6). To see this, we focus on a random projection $\mathbf{P}_k$ with i.i.d. $\mathcal{N}(0, 1)$ entries. First, for $\boldsymbol{\eta} = \mathbf{0}$, model (2.6) can be written as $\mathbf{y} = \boldsymbol{\epsilon} = \mathbf{X}\mathbf{0} + \boldsymbol{\epsilon}$. Therefore, $\mathbf{y}$ has the same distribution in model (2.4) for $\boldsymbol{\beta} = \mathbf{0}$. Second, for $\boldsymbol{\eta} \neq \mathbf{0}$, $\mathbf{P}_k\boldsymbol{\eta} \neq \mathbf{0}$ holds with probability 1, since $\mathbf{P}_k\boldsymbol{\eta}$ is distributed as $\mathcal{N}(\mathbf{0}, ||\boldsymbol{\eta}||_2^2\mathbf{I})$. Consequently, $\boldsymbol{\beta} = \mathbf{0}$ in model (2.4) implies $\boldsymbol{\eta} = \mathbf{0}$ in model (2.6), otherwise, a contradiction will be led 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 (2.6) is

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

The F-statistic based on (2.6) is

$$T_n = \frac{\mathbf{y}^\top \mathbf{H}_k \mathbf{y}/k}{\mathbf{y}^\top(\mathbf{I} - \mathbf{H}_k)\mathbf{y}/(n-k)}. \tag{2.7}$$

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:

$$T_n = \frac{\mathbf{y}^\top \mathbf{H}_k \mathbf{y}/k}{\mathbf{y}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_k)\mathbf{y}/(n-k-1)}, \tag{2.8}$$

where $\mathbf{P}_1 = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ 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 1 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 2.1.

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 $\mathcal{N}(0,1)$. 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

$$P(p_{ij} = \sqrt{l}) = P(p_{ij} = -\sqrt{l}) = \frac{1}{2l},\ P(p_{ij} = 0) = 1 - \frac{1}{l}, \tag{2.9}$$

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

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

2.3 Main Results

This section consists of the statements of our main theoretical results and related discussions. Specifically, we derive the asymptotic normality and the asymptotic power function for the new random-projection-based test. We also conduct a comparison with one of the latest tests in terms of asymptotic relative efficiency.

2.3.1 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 A1. $\boldsymbol{x}_i = \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{z}_i$, *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* $Var(\frac{\boldsymbol{z}_i^\top \boldsymbol{z}_i}{m}) = O(m^{-1})$. *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* 0 *when at least one of the* q_j *is odd.*

Assumption A2. $\mu_4 = E(\epsilon_i^4) < \infty$.

Assumption A3. $p \gg n$ *and there is a constant* $\rho \in (0, 1)$ *such that* $\frac{k}{n} \to \rho$.

As stated in Assumptions A1 and A3, 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 implicitly vary as n goes to infinity. This makes our test accommodate extremely high-dimensional problems. Taking a closer look at Assumption A1, we find it resembles a factor model structure which has a linear relationship between $\mathbf{x}_i$ and $\mathbf{z}_i$. It can be proved that the following two kinds of assumptions are both included in Assumption A1.

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. 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 its similar versions were used in Bai and Saranadasa (1996), Zhong and Chen (2011), and Cui, Guo, and Zhong (2018). Such assumptions are similar to Assumption A1, but imposing stricter conditions on each element of $\mathbf{z}_i$. This is because $\mathbf{z}_i$ in D1 satisfies $Var(\frac{\mathbf{z}_i^\top \mathbf{z}_i}{m}) = \frac{2+\Delta}{m}$. In the multivariate statistical analysis, elliptical distribution is often assumed to facilitate study. It includes a flexible family of distributions, including multivariate normal distribution, multivariate t-distribution and multivariate logistic distribution. Let $\mathbf{z}_i = r_i \mathbf{u}_i$ and $m = p$, D2 and Assumption A1 enjoy a similar form. Furthermore, Lemma 2.1, together with Lemma 2.3, indicates that the distributions satisfying D2 are included in Assumption A1.

Since 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 2.1. *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*

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

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 will be discussed in the subsequent Section 2.3.3.

Clearly, to derive the asymptotic distribution of T_n, we need to study the properties of the hat matrix $\mathbf{H}_k$. Since $\mathbf{H}_k = \mathbf{U}_k(\mathbf{U}_k^\top \mathbf{U}_k)^{-1}\mathbf{U}_k^\top$, 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 will show in Lemmas 2.9 and 2.10 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 2.1. *Suppose the random projection matrix $\boldsymbol{P}_k$ consists of i.i.d. the standard normal random variables. Under Assumptions A1–A3 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 $\mathcal{N}(0,1)$.

2.3.2 Asymptotic Power Function

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

Assumption A4. $\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta} = o(1)$.

Assumption A4 is known as a local alternative, which is commonly used in studying the asymptotic properties of a statistical test. Detailed discussions can be found in Van der Vaart (1998, Section 14.1).

In the classical F-test in (2.5), 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. It 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 the need 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 random projection. Specifically, the fact that randomly projected variable is asymptotically normal yields a new representation for the model (2.3) 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 rigorously show this, additional requirement for $\mathbf{z}_i$ is needed as follows.

Assumption A5. *The m-variate random vector $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{im})^\top$ has a Lebesgue density $f_{\boldsymbol{z}}$ and satisfy $E(\boldsymbol{z}_i) = \boldsymbol{0}$ 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 the marginal density bounded by a constant $D \geq 1$.*

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

Theorem 2.2. *Suppose that Assumptions A1–A5 hold. Let $\Psi_n^{RP}(\boldsymbol{\beta}; \boldsymbol{P}_k)$ denote the* *power function of the proposed random-projection-based test T_n. Then*

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

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

It is remarked 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 a increasing function of the product $\boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k \boldsymbol{\xi}$. It is found 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 A6. *(Tail eigenvalue condition.) There exists an integer s and a real number $\gamma > 0$ such that $s < k$ and $\frac{\sqrt{n}}{p} ||\boldsymbol{\beta}||_2^2 \sum_{i=s+1}^{p} d_i = o(n^{-\gamma})$, 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 A6 a tail eigenvalue condition, as it requires the product of $||\boldsymbol{\beta}||_2^2$ and the sum of tail eigenvalues of $\boldsymbol{\Sigma}$ to be of order less than $p/\sqrt{n}$.

Lemma 2.2. *Let $\boldsymbol{P}_k \in \mathbb{R}^{p \times k}$ consist of i.i.d. $\mathcal{N}(0,1)$ entries. Suppose that Assumption A6 holds, 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 negligible approximation error. In this case, we denote the asymptotic power function as $\Psi_n^{RP}(\boldsymbol{\beta})$ since it is not related to $\mathbf{P}_k$. A formal result is given in the following corollary.

Corollary 2.1. *Suppose that Assumptions A1–A6 hold. Then*

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

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

2.3.3 Choice of ρ

As demonstrated by Theorem 2.1, the proposed test can be practicable with any projection dimension k that satisfies Assumption A3. However, when the asymptotic power is considered, the value of k shows a significant influence through the ratio ρ, as shown in Theorem 2.2. In this subsection, we will give a detailed discussion for the choice of ρ.

From Theorem 2.2, the asymptotic local power function satisfies

$$\Psi_n^{RP}(\boldsymbol{\beta}; \mathbf{P}_k) = \Phi(-z_\alpha + \sqrt{\frac{n(1-\rho)}{2\rho}}\frac{\boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma}\mathbf{P}_k\boldsymbol{\xi}}{\delta_k^2}) + o(1). \tag{2.10}$$

Let $\Delta_k^2 = \boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma}\mathbf{P}_k\boldsymbol{\xi}$. It can be derived by projecting the vector $\boldsymbol{\Gamma}^\top\boldsymbol{\beta}$ onto the space generated by $\boldsymbol{\Gamma}^\top\mathbf{P}_k$, where the dimension is k. Intuitively, larger value of ρ would lead to larger Δ_k^2, since the projection space becomes bigger. However, with the increase of ρ, the value of function $\sqrt{(1-\rho)/\rho}$ would decrease. Therefore, we show that choice of ρ is a compromise between these two values.

First, we consider a situation, where the condition given in Corollary 2.1 is satisfied. In this case, Δ_k^2 becomes a deterministic value, even with randomly generated projection matrix $\mathbf{P}_k$. The asymptotic local power function is a decreasing function of ρ and we confirm this through the simulation studies. Therefore, the value of ρ can be arbitrarily small as long as the tail eigenvalue condition is satisfied.

Then, we consider the other situation, where $\mathbf{\Sigma} = \mathbf{I}$. In this case, the eigenvalues of $\mathbf{\Sigma}$ 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 Δ_k^2 scales linearly in k up to random fluctuations. Combining this result with (2.10), the influence of ρ on the testing power is mainly achieved based on the function $g(\rho) = \sqrt{\frac{1-\rho}{\rho}} \cdot \rho$, which is maximized when $\rho = 0.5$. Therefore, choice of $k = [0.5n]$ may be asymptotically optimal in a general sense.

For most applications, where no prior information of $\mathbf{\Sigma}$ might be available, the above discussion suggests that ρ around 0.5 would be an applicable choice, since the setting above can make the performance of the test reasonable, even in extreme cases. For some situations, where estimation methods of $\mathbf{\Sigma}$ or related function of $\mathbf{\Sigma}$ are available, ρ could be selected based on the estimators. For example, the ratio $tr(\mathbf{\Sigma})^2/tr(\mathbf{\Sigma}^2)$, which lies between 1 and p, can be viewed as measuring the decay rate of the spectrum of $\mathbf{\Sigma}$ (Lopes, Jacob, and Wainwright, 2011). And the tail eigenvalue condition could be satisfied when $tr(\mathbf{\Sigma})^2/tr(\mathbf{\Sigma}^2) \ll p$. Consequently, we could determine ρ according to 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).

2.3.4 Asymptotic Relative Efficiency (ARE)

The asymptotic power function of the proposed random-projection-based test in Corollary 2.1 has the same form as the F-test, which was studied in Zhong and Chen (2011). However, our test accommodates high-dimensional settings and has milder assumptions on $\mathbf{X}$ and $\boldsymbol{\epsilon}$. As 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 the testing problem (2.2) and is demonstrated to have superior performance over the existing tests for the problem considered. We denote this competing test by RCV test and show our test outperforms it in some situations. In this subsection, we suppose Assumption A6 holds.

With a slight abuse of notation, we also denote the asymptotic power function of our random-projection-based (RP) test as

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

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

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

Since the term added to $-z_\alpha$ inside the $\Phi(\cdot)$ function is what controls 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 RCV test as

$$ARE(\Psi_n^{RP}, \Psi_n^{RCV}) = \left(\sqrt{\frac{n(1-\rho)}{\rho}}\boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta} / \frac{n\boldsymbol{\beta}^\top \boldsymbol{\Sigma}^2\boldsymbol{\beta}}{\sqrt{tr(\boldsymbol{\Sigma}^2)}}\right)^2. \tag{2.11}$$

Whenever the ARE is larger than 1, the proposed test is asymptotically more powerful than the competing test. Therefore, we search for sufficient conditions under which the ARE is greater than 1.

Write $\boldsymbol{\beta}$ as $||\boldsymbol{\beta}||_2\boldsymbol{\delta}$, where $\boldsymbol{\delta} = \frac{\boldsymbol{\beta}}{||\boldsymbol{\beta}||_2}$ is the direction of $\boldsymbol{\beta}$. Under Assumptions A4 and A6, we further require the sum of tail eigenvalues satisfying $\sum_{i=s+1}^{p} d_i/\boldsymbol{\delta}^\top \boldsymbol{\Sigma}\boldsymbol{\delta} =$

$O(pn^{-0.5-\gamma})$, where γ is a small constant greater than zero. By Jensen's inequality, we have

$$ARE(\Psi_n^{RP};\Psi_n^{RCV}) \geq \frac{1-\rho}{\rho}\frac{\sum_{i=s+1}^{p} d_i^2}{n}(\frac{\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta}}{\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^2\boldsymbol{\delta}})^2 \geq \frac{1-\rho}{\rho}\frac{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta})^4}{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^2\boldsymbol{\delta})^2}O(pn^{-2-2\gamma}). \tag{2.12}$$

Clearly, if $\frac{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^2\boldsymbol{\delta})^2}{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta})^4} = o(pn^{-2-2\gamma})$, the right side of the inequality goes to infinity as n goes to ∞, which sufficiently demonstrates that the proposed test is more powerful than 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 $\frac{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^2\boldsymbol{\delta})^2}{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta})^4} = o(pn^{-2-2\gamma})$ is satisfied.

Example 2.1. *Suppose $\boldsymbol{\beta}$ is an eigenvector of $\boldsymbol{\Sigma}$, then $\frac{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^2\boldsymbol{\delta})^2}{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta})^4} = 1$. Given that $n = o(p^{1/(2+2\gamma)})$ for a constant $\gamma > 0$, which frequently happens when $p \gg n$, we have $\frac{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^2\boldsymbol{\delta})^2}{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta})^4} = o(pn^{-2-2\gamma})$.*

Example 2.2. *Suppose the covariance matrix $\boldsymbol{\Sigma}$ has the spectral decomposition*

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

where $\boldsymbol{O}$ is an orthogonal matrix with i-th column denoted by $\boldsymbol{o}_i$, 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 $\boldsymbol{\beta} \in Span\{\boldsymbol{o}_{s_1},\ldots,\boldsymbol{o}_{s_2}\}$. 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 $\frac{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}^2\boldsymbol{\delta})^2}{(\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta})^4} = o(pn^{-2-2\gamma})$, and thus our test outperforms RCV test in these situations.

2.4 Testing Partial Regression Coefficients

In Section 2.3, we proposed a random-projection-based test for testing the hypothesis (2.2). In many studies, we are also interested in investigating the significance of partial covariates. In this section, we generalize the test in Section 2.3 to the hypothesis testing of partial linear regression coefficients and derive its asymptotic results.

Consider a linear regression model

$$y = \alpha + \mathbf{x}_1^\top \boldsymbol{\beta}_1 + \mathbf{x}_2^\top \boldsymbol{\beta}_2 + \epsilon, \tag{2.13}$$

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

$$\mathbf{H}_{part,0} : \boldsymbol{\beta}_2 = \mathbf{0} \quad \text{versus} \quad \mathbf{H}_{part,1} : \boldsymbol{\beta}_2 \neq \mathbf{0}. \tag{2.14}$$

Suppose that $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ are i.i.d. copies of $(\mathbf{x}, y)$ from the linear regression model (2.13), where $\mathbf{x}_i = (\mathbf{x}_{1i}^\top, \mathbf{x}_{2i}^\top)^\top$ and $\mathbf{x} = (\mathbf{x}_1^\top, \mathbf{x}_2^\top)^\top$. Let $\mathbf{y} = (y_1, \ldots, y_n)^\top$ and $\mathbf{x}_{1i} = (x_{i1}^1, \ldots, x_{ip_1}^1)^\top$ be the i-th row of the matrix $\mathbf{X}_1 = (\mathbf{x}_{11}, \ldots, \mathbf{x}_{1n})^\top$. Similarly, let $\mathbf{X}_2 = (\mathbf{x}_{21}, \ldots, \mathbf{x}_{2n})^\top$. It follows that $\mathbf{y}$, $\mathbf{X}_1$ and $\mathbf{X}_2$ satisfy

$$\mathbf{y} = \alpha \mathbf{1} + \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2 + \boldsymbol{\epsilon} \tag{2.15}$$

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

Following the same idea in Section 2.3, we develop a new test for testing the hypothesis (2.14). For an integer $1 \leq k_2 < \min\{n - p_1, p_2\}$, 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 of the data. We define the following projection matrices:

$$\mathbf{P}_1 = \frac{1}{n}\mathbf{1}\mathbf{1}^\top,$$

$$\mathbf{P}_{\mathbf{X}_1} = (\mathbf{I} - \mathbf{P}_1)\mathbf{X}_1(\mathbf{X}_1^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{X}_1)^{-1}\mathbf{X}_1^\top(\mathbf{I} - \mathbf{P}_1),$$

$$\mathbf{H}_{k_2} = (\mathbf{I} - \mathbf{P}_1)\mathbf{W}(\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{W})^{-1}\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1),$$

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 1 when $\mathbf{P}_{k_2}$ has i.i.d. $\mathcal{N}(0,1)$ entries and k_2 is appropriately selected. This ensures the projection matrix $\mathbf{H}_{k_2}$ is well-defined, even when $p_2 > n$. We propose a new test statistic

$$T_{n,p_2} = \frac{\mathbf{y}^\top(\mathbf{H}_{k_2} - \mathbf{P}_{\mathbf{X}_1})\mathbf{y}/k_2}{\mathbf{y}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_{k_2})\mathbf{y}/(n-1-p_1-k_2)}. \tag{2.16}$$

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

2.4.1 Asymptotic Null Distribution

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

Assumption S1. $\boldsymbol{x}_i = (\boldsymbol{x}_{1i}^\top, \boldsymbol{x}_{2i}^\top)^\top = \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{z}_i$, *where* $\boldsymbol{x}_{1i} \in \mathbb{R}^{p_1}$ *and* $\boldsymbol{x}_{2i} \in \mathbb{R}^{p_2}$ *are covariates,* $\boldsymbol{\mu}$ *is a p-dimensional mean vector,* $\boldsymbol{\Gamma}$ *is a* $p \times m$ *matrix with* $m \geq p$, *and* $\boldsymbol{z}_i$ *is an m-variate random vector with* $E(\boldsymbol{z}_i) = \boldsymbol{0}, Var(\boldsymbol{z}_i) = \boldsymbol{I}_m$, *and* $Var(\frac{\boldsymbol{z}_i^\top \boldsymbol{z}_i}{m}) = O(m^{-1})$. *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 0 when at least one of the* q_j *is odd.*

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

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* $\frac{p_1}{n} \to \rho_1$ *and* $\frac{k_2}{n} \to \rho_2$.

Since 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 can be comparable with n. The projection dimension k_2 needs to be asymptotically proportional to n and the choice of ρ_2 will be discussed below.

Theorem 2.3. *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 a 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 $\mathcal{N}(0,1)$.

2.4.2 Asymptotic Power Function

We are now in a position to study the asymptotic power of the test. We first divide $\boldsymbol{\Gamma} = (\boldsymbol{\Gamma}_1^\top, \boldsymbol{\Gamma}_2^\top)^\top$ with $\boldsymbol{\Gamma}_1 \in \mathbb{R}^{p_1 \times m}$ and $\boldsymbol{\Gamma}_2 \in \mathbb{R}^{p_2 \times m}$. Define $\boldsymbol{\Sigma}_{11} = \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_1^\top$, $\boldsymbol{\Sigma}_{22} = \boldsymbol{\Gamma}_2\boldsymbol{\Gamma}_2^\top$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Gamma}_1\boldsymbol{\Gamma}_2^\top$ and $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Gamma}_2\boldsymbol{\Gamma}_1^\top$. Following the same idea in Section 2.3, we give additional assumptions to facilitate our analysis.

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

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

Define $\mathbf{V} = \text{diag}(\mathbf{I}_{p_1}, \mathbf{P}_{k_2})$ and $\boldsymbol{\gamma} = (\mathbf{V}^\top \boldsymbol{\Sigma} \mathbf{V})^{-1} \mathbf{V}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}$. 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 $\tau_k^2 = \sigma^2 + \boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta} - \boldsymbol{\gamma}^\top \mathbf{V}^\top \boldsymbol{\Sigma} \mathbf{V} \boldsymbol{\gamma}$. We derive the asymptotic power function of the proposed test.

Theorem 2.4. *Under Assumptions S1–S5, we have*

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

where $\Phi(\cdot)$ is its 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 $\boldsymbol{\Sigma}$. From the expression of the asymptotic power function, we can see that the product $\boldsymbol{\xi}_2^\top (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})\boldsymbol{\xi}_2$ is preferred to be larger, which is dependent on $\mathbf{P}_{k_2}$ and is upper bounded by $\boldsymbol{\beta}_2^\top (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})\boldsymbol{\beta}_2$. 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 $\frac{\sqrt{n}}{p_2}||\boldsymbol{\beta}_2||_2^2 \sum_{i=s_2+1}^{p_2} d_i = o(n^{-\gamma_2})$, where d_i are the eigenvalues of $\boldsymbol{\Sigma}_{22}$ satisfying $d_1 \geq d_2 \geq \cdots \geq d_{p_2} \geq 0$.*

This assumption ensures Lemma 2.2 is valid for $\boldsymbol{\beta}_2$ and $\boldsymbol{\Sigma}_{22}$, leading to a negligible distance between the vector $\boldsymbol{\Gamma}^\top \boldsymbol{\beta}$ and the space generated by $\boldsymbol{\Gamma}^\top \mathbf{V}$. In this case, we denote the power function of the proposed random-projection-based test T_{n,p_2} as $\Psi_{n,p_2}^{RP}(\boldsymbol{\beta}_2)$.

Corollary 2.2. *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 its cumulative distribution function of the standard normal distribution, and z_α is the upper α-quantile of Φ.

2.5 Numerical Studies

2.5.1 Simulation Studies

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

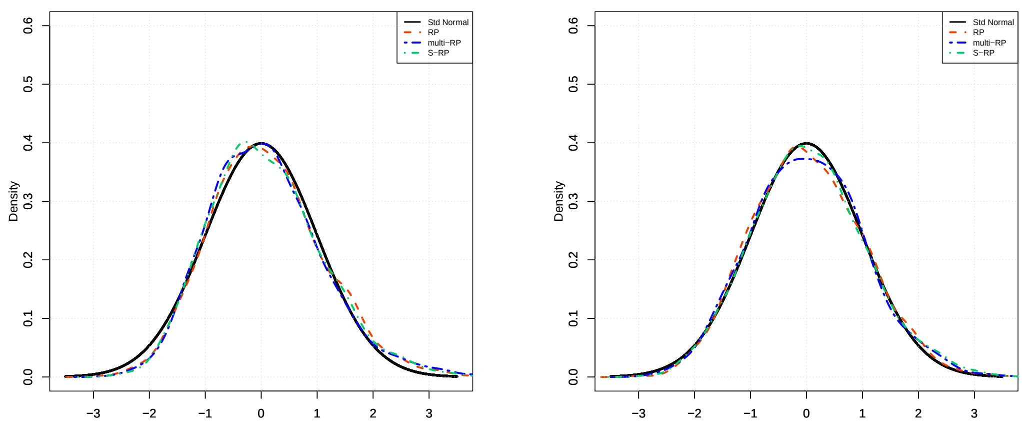
The first simulation study was designed for testing the hypothesis: $\mathbf{H}_0 : \boldsymbol{\beta} = \mathbf{0}$ versus $\mathbf{H}_1 : \boldsymbol{\beta} \neq \mathbf{0}$ in the linear regression model

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

Set $\alpha = 2$. Suppose that ϵ_i was generated from $\mathcal{N}(0,1)$ or $t(5)/\sqrt{5/3}$, and covariate $\mathbf{x}_i$ was generated from $\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z}_i$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^\top$ with μ_i independently generated from $U(2,3)$, and each entry of $\mathbf{z}_i = (z_{i1}, \ldots, z_{ip})^\top$ was i.i.d. from (i) $\mathcal{N}(0,1)$ or (ii) $U(-\sqrt{3}, \sqrt{3})$. The matrix $\boldsymbol{\Sigma}^{1/2}$ was generated by $\mathbf{U}\sqrt{\mathbf{D}}\mathbf{U}^\top$, where $\mathbf{U}$ was an orthogonal matrix generated from the uniform distribution on the $p \times p$ orthogonal group with the i-th column denoted by $\mathbf{u}_i$ and $\sqrt{\mathbf{D}} = \mathrm{diag}(\sqrt{d_1}, \ldots, \sqrt{d_p})$. Let $s = \lfloor n^{0.72} \rfloor$ and $L = \lfloor n^{0.8} \rfloor$. 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 vector of regression coefficients $\boldsymbol{\beta}$ was 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. Different values of M were considered in the simulations: (i) $M = 0$ and (ii) $M = 50$. Working under high-dimensional settings, we set $(n,p) = (300, 3000), (400, 5000), (800, 5000)$.

In the simulations, we implemented three types of random-projection-based tests according to the choice of random projection: (i) RP test: applying the normal random projection; (ii) multi-RP test: independently generating the normal random projection for 10 times and utilizing their mean; (iii) S-RP test: applying the sparse random projection defined in (2.9) with $l = 400$.

We first report the kernel density estimation of the proposed test statistics under $\mathbf{H}_0$ in Figures 2.1(a) and 2.1(b), showing that the asymptotic null distribution of 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 2.1.



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

Figure 2.1: The kernel density estimation of RP,multi-RP and S-RP tests under $\mathbf{H}_0$.

Tables 2.1 and 2.2 report the type I errors and empirical powers of the proposed tests and RCV test for ϵ distributed from $\mathcal{N}(0, 1)$ and $\sqrt{3/5}t(5)$ based on 2000 simulations. It can be observed that the performances of three proposed tests have negligible differences, which confirms the discussion in Section 2.2 and suggests the feasible usage of different random projection in the test. The type I errors of the proposed tests and RCV test are close to 0.05 under the null hypothesis. The empirical powers of the proposed tests are decreasing functions of ρ, which is consistent with the result in Theorem 2.2. Moreover we can see that the power of the tests are increasing functions of the L_2 norm of $\boldsymbol{\beta}$. Compared with the RCV test, the proposed tests are more powerful.

Table 2.1: Type I errors and empirical powers of 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 2.2: Type I errors and empirical powers of RP, multi-RP, S-RP and RCV tests at the significance level 0.05 when $\epsilon \sim \sqrt{3/5}t(5)$.

M	ρ	$\|\|\boldsymbol{\beta}\|\|_2^2$	$Z \sim U(-\sqrt{3}, \sqrt{3})$ RP	multi-RP	S-RP	RCV	$Z \sim \mathcal{N}(0,1)$ 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 2.3: Type I errors and empirical powers of multi-RP, LWT, LDFF and RCV tests at the significance level 0.05.

d	$\boldsymbol{\beta}$	$\|\|\boldsymbol{\beta}\|\|_2^2$	multi-RP	LWT	LDFF	RCV
d=3		0	0.062	0.052	0.050	0.458
	$s=5$	0.04	0.249	0.087	0.086	0.502
		0.08	0.532	0.116	0.118	0.544
	$s=50$	0.04	0.735	0.218	0.216	0.787
		0.08	0.984	0.409	0.388	0.951
d=5		0	0.052	0.071	0.069	0.843
	$s=5$	0.04	0.295	0.183	0.181	0.917
		0.08	0.605	0.312	0.308	0.959
	$s=50$	0.04	0.764	0.387	0.384	0.999
		0.08	0.987	0.698	0.681	1.000

In the second simulation, we conducted numerical comparison with LWT test and LDFF test proposed in Lan, Wang, and Tsai (2014) and Lan et al. (2016), respectively. The data were generated from $y_i = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i$, where $\alpha = 0$ and ϵ_i was generated from $\mathcal{N}(0, 1)$. The covariate $\mathbf{x}_i$ followed a latent factor structure in Lan et al. (2016). Specifically, $\mathbf{x}_i = \boldsymbol{\gamma}\mathbf{z}_i + \sqrt{\mathbf{D}}\widetilde{\mathbf{x}}_i$, where $\mathbf{z}_i$ is a d-dimensional latent factor, $\boldsymbol{\gamma} \in \mathbb{R}^{p\times d}$ is an associated factor loadings, $\widetilde{\mathbf{x}}_i$ is a p-dimensional factor profiled predictor that is independent of $\mathbf{z}_i$, and $\mathbf{D}$ is a diagonal matrix. From Lan et al. (2016), the factor profiled predictor $\widetilde{\mathbf{x}}_i$ represents the information that is contained in $\mathbf{x}_i$ but cannot be fully explained by the low-dimensional latent factor $\mathbf{z}_i$. In the simulation, each element of $\mathbf{z}_i$ and $\widetilde{\mathbf{x}}_i$ was independently generated from $\mathcal{N}(0, 1)$, and each entry of $\boldsymbol{\gamma} \in \mathbb{R}^{p\times d}$ was independently generated from $\mathcal{N}(0, d^{-1})$. The elements of $\sqrt{\mathbf{D}}$ were generated in the same way as that in the first set of simulation, when $s = \lfloor n^{0.5} \rfloor$ and $L = \lfloor n^{1.5} \rfloor$. For the alternative hypothesis, we considered $\boldsymbol{\beta} = ||\boldsymbol{\beta}||_2 \boldsymbol{\delta}$, where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^\top$ with $\delta_j = s^{-1/2}$, for $j \leq s$, and otherwise, $\delta_j = 0$. The integer s took values 5 and 50 to denote different levels of sparsity, and the norm $||\boldsymbol{\beta}||_2^2 = 0.04$ and 0.08. In the simulation, $(n, p) = (300, 3000)$.

As shown in Table 2.3, the type I errors of multi-RP, LWT and LDFF tests are around 0.05, which indicates that the type I error can be well controlled at the nominal level by the tests. But for RCV test, the type I errors are alarmingly larger than the given significance level, which indicates the test might not be applicable in this experimented setting, where the covariates have high correlations based on the latent factor structure. Therefore, the comparison for the empirical powers is only considered among multi-RP test, LWT test and LDFF test. Table 2.3 indicates that empirical powers grow when $||\boldsymbol{\beta}||_2$ increases and the performances of LWT and LDFF tests are similar. The large empirical powers indicate that our proposed test has superior performances in all the experimented alternatives. Therefore, the simulation results demonstrate that our proposed test is applicable in the highly correlated setting and has higher testing power than the competing tests in some cases.

In the third simulation study, we consider the problem of testing partial regression coefficients in the linear regression model

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

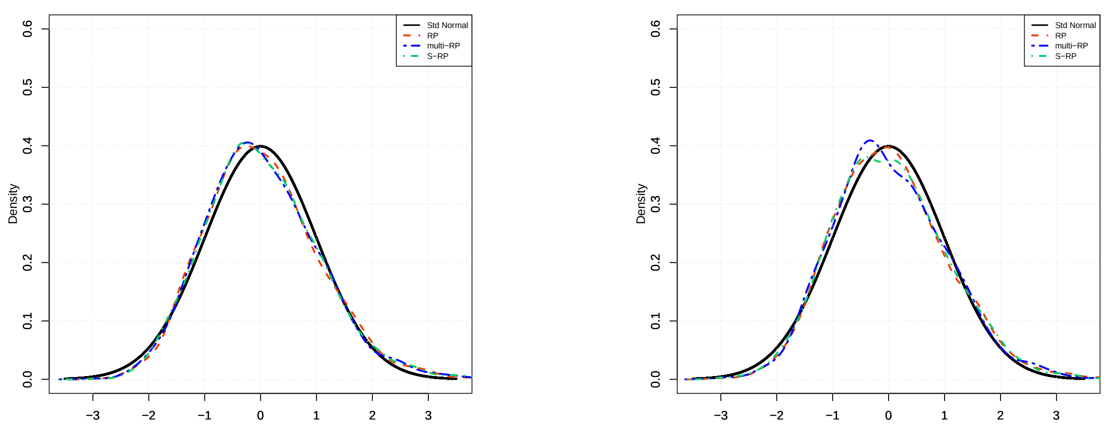
The covariate $(\mathbf{x}_{1i}^\top, \mathbf{x}_{2i}^\top)^\top$ was generated from $\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\mathbf{z}_i$. The setup was almost the same as the first simulation study with differences lying in the design of $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\Sigma}^{1/2}$. Specifically, we generated $\boldsymbol{\Sigma}^{1/2}$ by

$$\begin{pmatrix} c_1 \mathbf{U}_1 \sqrt{\mathbf{D}_1} \mathbf{U}_1^\top & c_2 \mathbf{U}_1 (\sqrt{\mathbf{D}_1}, \mathbf{0}) \mathbf{U}_2^\top \\ \mathbf{0} & \mathbf{U}_2 \sqrt{\mathbf{D}_2} \mathbf{U}_2^\top \end{pmatrix},$$

where $\mathbf{U}_1$ ($\mathbf{U}_2$) was an orthogonal matrix generated from the uniform distribution on the $p_1 \times p_1$ ($p_2 \times p_2$) orthogonal group, the entries of diagonal matrix $\mathbf{D}_1$ were from $\mathcal{N}(\mathbf{0}, \mathbf{I}_{p_1})$ with absolute values taken and the entries of diagonal matrix $\mathbf{D}_2$ were generated in the same way as the first simulation study for the small tail eigenvalue requirement. We used an indicator R for the different cases: (i) uncorrelated case

$(R = 0)$: $c_1 = 1$, $c_2 = 0$; (ii) correlated case $(R = 1)$: $c_1 = c_2 = 1/\sqrt{2}$. Here the values of c_1 and c_2 were selected to ensure the variances of $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$ kept unchanged in the two cases. The vector of regression coefficients $\boldsymbol{\beta}_1$ was generated from $\mathcal{N}(\mathbf{0}, \mathbf{I}_{p_1})$ and $\boldsymbol{\beta}_2$ was randomly selected from the space generated by the first s columns of $\mathbf{U}_2$ with $||\boldsymbol{\beta}_2||_2^2$ taking 0.1, 0.2 and 0.3. This selection was aimed for a better display of the impact from the correlation on the power of the tests. For a high-dimensional design, we chose (n, p_1, p_2) to be $(400, 40, 3960)$.

Figures 2.2(a) and 2.2(b) display the kernel density estimation of the proposed test statistics under $\mathbf{H}_{part,0}$, indicating that the asymptotic null distribution of the proposed tests can be well approximated by the standard normal distribution. Here ρ takes the value 0.2. We show both the correlated and uncorrelated cases. The good resemblance to the normal distribution confirms the theoretical results in Theorem 2.3.



(a) Norm $\mathbf{z}$, norm ϵ and R=0. (b) Norm $\mathbf{z}$, norm ϵ and R=1.

Figure 2.2: The kernel density estimation of RP,multi-RP and S-RP tests under $\mathbf{H}_{part,0}$.

Table 2.4 reports the type I errors and empirical powers of the proposed tests for error term ϵ distributed from $\mathcal{N}(0, 1)$ and $\sqrt{3/5}t(5)$ based on 2000 simulations. It can be observed that the performances of three proposed tests have negligible differences. The type I errors of the proposed tests are close to 0.05 and the power of the tests are increasing functions of the norm $||\boldsymbol{\beta}_2||_2^2$. Compared with the correlated case, the tests show large power when there is no correlation between $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$, which is consistent with the feature of the asymptotic power in Theorem 2.4.2. Moreover, we find the empirical power is close to the asymptotic power, which further confirms the result in Theorem 2.4.2.

Table 2.4: Type I errors and empirical powers of RP, multi-RP, S-RP at the significance level 0.05 when $(n, p_1, p_2) = (400, 40, 3960)$ and $\rho = 0.2$.

Z	R	$\|\|\boldsymbol{\beta}_2\|\|_2^2$	$\epsilon \sim \sqrt{3/5}t(5)$			$\epsilon \sim \mathcal{N}(0,1)$		
			RP	multi-RP	S-RP	RP	multi-RP	S-RP
$\mathcal{N}(0,1)$	0	0	0.056	0.059	0.053	0.063	0.057	0.060
		0.1	0.715	0.729	0.717	0.704	0.707	0.715
		0.2	0.981	0.978	0.979	0.980	0.982	0.980
		0.3	0.999	0.998	0.999	1.000	0.999	0.999
	1	0	0.063	0.060	0.064	0.063	0.060	0.062
		0.1	0.533	0.548	0.532	0.544	0.532	0.545
		0.2	0.903	0.897	0.904	0.898	0.900	0.904
		0.3	0.988	0.983	0.985	0.992	0.991	0.990
$U(-\sqrt{3}, \sqrt{3})$	0	0	0.064	0.058	0.060	0.063	0.066	0.065
		0.1	0.716	0.716	0.720	0.717	0.711	0.722
		0.2	0.983	0.981	0.984	0.981	0.981	0.986
		0.3	1.000	1.000	1.000	1.000	1.000	1.000
	1	0	0.058	0.057	0.056	0.059	0.062	0.060
		0.1	0.533	0.537	0.539	0.533	0.542	0.542
		0.2	0.901	0.895	0.901	0.905	0.911	0.916
		0.3	0.991	0.992	0.991	0.991	0.992	0.993

2.5.2 Illustrative Examples

To illustrate the proposed methods, we consider here two examples.

2.5.2.1 Example 1

We considered a real data set of riboflavin (vitamin B2) production by bacillus subtilis. The data was analyzed by Van de Geer et al. (2014) and is available in R package "hdi". The real-valued response variable is the logarithm of the riboflavin production rate and there are $p = 4088$ covariates (genes) measuring the logarithm of the expression level of 4088 genes. These measurements are from $n = 71$ samples of genetically engineered mutants of bacillus subtilis. We modeled the data with a high-dimensional linear model and obtained the p-values of the proposed tests and RCV test in Table 2.5. It is illustrated that all the tests reject the null hypothesis, indicating a considerable significance of genes expression in predicting riboflavin production rate.

Then we were interested in the significance of partial gene expressions. We randomly divided the data into two subsets. Based on the LASSO for the first subset, we divided coefficients into two parts $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, where the index of $\boldsymbol{\beta}_2$ corresponded to the index of the zero part in $\hat{\boldsymbol{\beta}}^{Lasso}$. We conducted testing for $\boldsymbol{\beta}_2$ on the second subset of data and the results are shown in Table 2.5. These large p-values indicate that $\mathbf{H}_{part,0}$ is accepted and this is consistent with the LASSO result.

2.5.2.2 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 using RNA from $n = 114$ samples of pretreated patients with HER2-positive (HER2+) tumors was performed. As multiple probes might represent the same gene, measurement for each gene was from the probe with the highest interquartile range. After a natural logarithm transformation, we obtained expression values of 20592 genes. In Prat et al. (2014), the implement of researched-based prediction analysis of microarray 50 (PAM50) subtype predictor to the data reported the predominated subtype within HER2+ disease is HER2-enriched (HER2-E) tumors, which has been found the high expression of HER2-regulated genes (for example, ERBB2, GRB7 and FGFR4) is one of the most important characteristics. To have more understanding of HER2-E subtype, we studied the association between HER2-regulated genes and residual genes, with ERBB2 as an example.

Let the response variable be the gene expression level of ERBB2 and the residual $p = 20591$ gene expression levels be the covariates. Suppose that the data follow a linear model, RCV test and our proposed tests reported a significant relationship by rejecting the null hypothesis, which is shown in Table 2.6. We moved on to identifying strongly associated genes based on the cooperation of the proposed tests and the LASSO estimation. Let the regression coefficients corresponding to the zeros in the LASSO estimator denote as $\boldsymbol{\beta}_2$, the proposed tests for the testing problem of this vector of partial regression coefficients were conducted. The p-values of the global and partial hypothesis testing in the table suggest that genes with nonzero coefficient, ESR1, MAP4K3, TLK1 ect., have significant influences 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 the lower expression of luminal-related gene ESR1 is one of important characteristics of HER2-enriched (HER2-E) tumors. Gamez-Pozo et al. (2014) found gene expression of MAP4K3 related to the PI3K pathway, which is strongly associated with response to trastuzumab in HER2 breast cancer. Consequently, the new testing procedures can be helpful in confirming existing knowledge and making new discoveries.

Table 2.5: The p-values of the proposed tests and 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	0.00	0.00	0.00	0.54	0.74	0.58

Table 2.6: The p-values of the proposed tests and 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

2.6 Proofs of the Technical Results

In this section, we provide the proofs of Lemmas 2.1–2.2 and Theorems 2.1–2.4.

2.6.1 Proof of Lemma 2.1

We first state a result from Fang, Kotz, and Ng (1990, Section 3.1), which shows some properties of uniform distribution on the surface of an unit sphere.

Lemma 2.3. *Let $\boldsymbol{u}_1 = (u_{11}, \ldots, u_{1p})^\top$ be a random vector uniformly distributed on the unit sphere in $\mathbb{R}^p$. Then $\boldsymbol{u}_1$ satisfies $E(\boldsymbol{u}_1) = \boldsymbol{0}, Var(\boldsymbol{u}_1) = \frac{1}{p}\boldsymbol{I}_p$. For $\forall j \neq k$, $E(u_{1j}^4) = \frac{3}{p(p+2)}$, $E(u_{1j}^2 u_{1k}^2) = \frac{1}{p(p+2)}$. And for any nonnegative integers $q_1, \ldots, q_p$, with $m = \sum_{j=1}^p q_j$, the mixed moments $E\big(\Pi_{j=1}^p u_{1j}^{q_j}\big) = 0$ if at least one of the q_j is odd.*

Proof of Lemma 2.1. From the definition of r_1, $\mathbf{u}_1$ and Lemma 2.3, we have

$$E(\mathbf{z}_1) = E(r_1\mathbf{u}_1) = E(r_1)E(\mathbf{u}_1) = 0,$$

$$Var(\mathbf{z}_1) = Var(E(\mathbf{z}_1|r_1)) + E(Var(\mathbf{z}_1|r_1)) = E(r_1^2 Var(\mathbf{u}_1)) = \mathbf{I}_p.$$

By definition that $\mathbf{z}_1 = (z_{11}, \ldots, z_{1p})^\top = r_1\mathbf{u}_1$, we have, for $\forall i \neq j$,

$$E(z_{1i}^4) = E(r_1^4 u_{1i}^4) = 3 + O(p^{-1}), \quad E(z_{1i}^2 z_{1j}^2) = E(r_1^4 u_{1i}^2 u_{1j}^2) = 1 + O(p^{-1}).$$

Hence we have

$$Var(\frac{\mathbf{z}_1^\top \mathbf{z}_1}{p}) = \frac{\sum_{i=1}^p E(z_{1i}^4) + \sum_{i\neq j} E(z_{1i}^2 z_{1j}^2)}{p^2} - E(\frac{\mathbf{z}_1^\top \mathbf{z}_1}{p})^2 = O(p^{-1}),$$

and complete the proof. □

2.6.2 Auxiliary Lemmas

We first present a result of asymptotic normality of quadratic form that was discussed by Bhansali, Giraitis, and Kokoszka (2007).

Lemma 2.4. *Consider a general quadratic form*

$$Q_n = \boldsymbol{z}^\top \boldsymbol{A}_n \boldsymbol{z} = \sum_{i,j=1}^{n} z_i a_{ij} z_j,$$

where z_i are i.i.d. variables with $E(z_i) = 0$ and $Var(z_i) = 1$, and a_{ij} are entries of a symmetric matrix $\boldsymbol{A}_n$.

(1) If $E(z_i^4) < \infty$ and $\frac{||\boldsymbol{A}_n||_{sp}}{||\boldsymbol{A}_n||_F} \to 0$, then

$$Var(Q_n)^{-1/2}\big(Q_n - E(Q_n)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

(2) If $\frac{||\boldsymbol{A}_n||_{sp}}{||\boldsymbol{A}_n||_F} \to 0$, $E(z_i^{2+\delta}) < \infty$ (for some $\delta > 0$), and $\sum_{i=1}^{n} a_{ii}^2 = o(||\boldsymbol{A}_n||_F^2)$, then

$$\frac{1}{\sqrt{2}||\boldsymbol{A}_n||_F}\big(Q_n - E(Q_n)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Lemma 2.5 (Woodbury's formula)**.** *Suppose $\boldsymbol{G}$ is an $n \times n$ nonsingular matrix, $\boldsymbol{U}$ and $\boldsymbol{V}$ are $n \times k$ matrices with $n > k$. If the matrix $(\boldsymbol{I}_k + \boldsymbol{V}^\top \boldsymbol{G}^{-1} \boldsymbol{U})$ is invertible, we have*

$$(\boldsymbol{G} + \boldsymbol{U}\boldsymbol{V}^\top)^{-1} = \boldsymbol{G}^{-1} - \boldsymbol{G}^{-1}\boldsymbol{U}(\boldsymbol{I}_k + \boldsymbol{V}^\top \boldsymbol{G}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}^\top \boldsymbol{G}^{-1}.$$

Suppose $\boldsymbol{u}$ and $\boldsymbol{v}$ are vectors. Define $\boldsymbol{H} = \boldsymbol{u}\boldsymbol{v}^\top$ and $g = tr(\boldsymbol{H}\boldsymbol{G}^{-1})$. If $g \neq -1$, we have

$$(\boldsymbol{G} + \boldsymbol{H})^{-1} = \boldsymbol{G}^{-1} - \frac{1}{1+g}\boldsymbol{G}^{-1}\boldsymbol{H}\boldsymbol{G}^{-1}.$$

We then depict some results about sample covariance matrix in high dimensions. The first is the celebrated work of Marčenko and Pastur (1967), which is named the

M-P law by some authors. The second is concerned with the extreme eigenvalues from Bai and Yin (1993, Theorem 2).

Lemma 2.6. *Let* $\boldsymbol{X} = (x_{ij}) \in \mathbb{R}^{k\times n}$ *be a matrix of i.i.d. entries with zero mean and unit variance. Define* $\boldsymbol{S}_n = \frac{1}{n}\boldsymbol{X}\boldsymbol{X}^\top$*. Suppose the eigenvalues of* $\boldsymbol{S}_n$ *are* λ_j*,* $j = 1,\dots,k$*, the empirical spectral distribution (ESD) of the matrix* $\boldsymbol{S}_n$ *is defined as* $F^{\boldsymbol{S}_n} = \frac{1}{k}\sum_{j=1}^{k}\mathbf{1}_{\{\lambda_j\le x\}}$*. If* $E(x_{11}^4) < \infty$*, as* $(n,k)\to\infty$ *with relationship* $k/n \to \rho \in (0,1)$*, we have*

(1) $F^{\boldsymbol{S}_n}$ *tends to the standard M-P law with probability 1, where the standard M-P law* $F_\rho(x)$ *has a density function*

$$p_\rho(x) = \begin{cases} \frac{1}{2\pi x\rho}\sqrt{(b-x)(x-a)}, & \text{if } a \le x \le b, \\ 0, & \text{otherwise,} \end{cases}$$

where $a = (1-\sqrt{\rho})^2$ *and* $b = (1+\sqrt{\rho})^2$.

(2) The extreme eigenvalues of $\boldsymbol{S}_n$ *satisfy*

$$\lambda_{\max}(\boldsymbol{S}_n) \to (1+\sqrt{\rho})^2 \ a.s.,$$

and

$$\lambda_{\min}(\boldsymbol{S}_n) \to (1-\sqrt{\rho})^2 \ a.s..$$

Lemma 2.7. *Let* $\boldsymbol{X} = (\boldsymbol{x}_1,\dots,\boldsymbol{x}_n)$ *be a random matrix with* $\boldsymbol{x}_i$ *i.i.d. from* $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_k)$*. As* $(k,n)\to\infty$ *with relationship* $k/n\to\rho\in(0,1)$*, we have*

(1) $\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{P}_1)\boldsymbol{X}^\top$ *and* $\bar{\boldsymbol{x}}$ *are independent, where* $\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i$.

(2) $E\big((\frac{1}{n-1}\boldsymbol{x}_n^\top\boldsymbol{S}_{n-1}^{-1}\boldsymbol{x}_n - \frac{\rho}{1-\rho})^2\big) = o(1)$*, where* $\boldsymbol{S}_{n-1} = \frac{1}{n-1}\sum_{j=1}^{n-1}\boldsymbol{x}_j\boldsymbol{x}_j^\top$*, and* $E\left((\boldsymbol{x}_i^\top(\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{x}_i - \rho)^2\right) = o(1)$*,* $\boldsymbol{x}_i^\top(\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{x}_i \le \frac{1}{1+(1-\sqrt{\rho})^2}$*, a.s..*

Proof. (1) We first define an orthogonal matrix $\mathbf{O}$ by

$$\mathbf{O} = (\mathbf{o}_1, \ldots, \mathbf{o}_n) = \begin{bmatrix} \frac{1}{\sqrt{n}} & 0 & 0 & \cdots & -\frac{\sqrt{n-1}}{\sqrt{n}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{n}} & 0 & -\frac{\sqrt{2}}{\sqrt{3}} & \cdots & \frac{1}{\sqrt{n(n-1)}} \\ \frac{1}{\sqrt{n}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \cdots & \frac{1}{\sqrt{n(n-1)}} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \cdots & \frac{1}{\sqrt{n(n-1)}} \end{bmatrix}.$$

Let $\mathbf{V} = \mathbf{X}\mathbf{O}$ with the i-th column denoted as $\mathbf{v}_i$. Then the design of orthogonal matrix $\mathbf{O}$ implies $\mathbf{X}\mathbf{X}^\top = \mathbf{X}\mathbf{O}\mathbf{O}^\top\mathbf{X}^\top = \sum_{i=1}^n \mathbf{v}_i\mathbf{v}_i^\top$, $\mathbf{v}_1 = \sqrt{n}\bar{\mathbf{x}}$ and $\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top = \sum_{i=2}^n \mathbf{v}_i\mathbf{v}_i^\top$. To study the properties of $\mathbf{v}_i$, the random matrix $\mathbf{X}$ is divided by rows and denoted as $(\mathbf{r}_1, \ldots, \mathbf{r}_k)^\top$ with k independent variables from $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. It follows that $\mathbf{v}_i = (\mathbf{r}_1, \ldots, \mathbf{r}_k)^\top\mathbf{o}_i$ and is distributed as $\mathcal{N}(\mathbf{0}, \mathbf{I}_k)$. Let $\mathbf{C}^{i,j} = (C^{i,j}_{s,l})^k_{s,l=1} = Cov(\mathbf{v}_i, \mathbf{v}_j)$, for $i \neq j$. Then we have

$$C^{i,j}_{s,l} = E(\mathbf{r}_s^\top\mathbf{o}_i\mathbf{r}_l^\top\mathbf{o}_j) - E(\mathbf{r}_s^\top\mathbf{o}_i)E(\mathbf{r}_l^\top\mathbf{o}_j) = 0,\ s \neq l,$$

$$C^{i,j}_{s,s} = E(\mathbf{r}_s^\top\mathbf{o}_i\mathbf{r}_s^\top\mathbf{o}_j) - E(\mathbf{r}_s^\top\mathbf{o}_i)E(\mathbf{r}_s^\top\mathbf{o}_j) = E(\mathbf{o}_i^\top\mathbf{r}_s\mathbf{r}_s^\top\mathbf{o}_j) = 0,\ s = 1, \ldots, k,$$

which indicates $\mathbf{v}_i$ and $\mathbf{v}_j$ are independent. This is sufficient to show that $\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top$ and $\bar{\mathbf{x}}$ are independent.

(2) From the direct calculation, the standard M-P law $F_\rho(x)$ in Lemma 2.6 satisfies

$$\begin{aligned}
\int \frac{1}{x} dF_\rho(x) &= \int_a^b \frac{1}{2\pi x^2 \rho}\sqrt{(b-x)(x-a)}dx \\
&= \frac{1}{2\pi\rho}\int_{-2\sqrt{\rho}}^{2\sqrt{\rho}} \frac{1}{(1+\rho+z)^2}\sqrt{4\rho - z^2}dz \text{ (with } x = 1+\rho+z) \\
&= \frac{1}{2\pi\rho}\int_{-\pi/2}^{\pi/2} \frac{4\rho\cos^2\theta}{(1+\rho+2\sqrt{\rho}\sin\theta)^2}d\theta \text{ (with } z = 2\sqrt{\rho}\sin\theta) \\
&= \frac{1}{2\pi\rho}\left(\frac{-2\sqrt{\rho}\cos\theta}{1+\rho+2\sqrt{\rho}\sin\theta}\bigg|_{-\pi/2}^{\pi/2} + \int_{-\pi/2}^{\pi/2} \frac{-2\sqrt{\rho}\sin\theta}{1+\rho+2\sqrt{\rho}\sin\theta}d\theta \right)
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{2\rho} + \frac{1}{2\pi\rho}\int_{-\pi/2}^{\pi/2} \frac{1}{1+c\sin\theta} d\theta \ \big(\text{with } c = 2\sqrt{\rho}(1+\rho)^{-1} < 1\big) \\
&= -\frac{1}{2\rho} + \frac{1}{2\pi\rho}\int_{-\pi/2}^{\pi/2} \frac{1}{\cos^2\frac{\theta}{2}(1+\tan^2\frac{\theta}{2} + 2c\tan\frac{\theta}{2})} d\theta \\
&= -\frac{1}{2\rho} + \frac{1}{2\pi\rho}\int_{-1}^{1} \frac{2}{1+t^2+2ct} dt \ (\text{with } t = \tan\frac{\theta}{2}) \\
&= -\frac{1}{2\rho} + \frac{1}{2\pi\rho} \cdot \frac{2}{\sqrt{1-c^2}} \arctan(\frac{t+c}{\sqrt{1-c^2}})\Big|_{-1}^{1} \\
&= \frac{1}{1-\rho}.
\end{aligned}
$$

We first study the asymptotic behavior of $\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n$. From the normality of $\mathbf{x}_i$, Lemma 2.6 and the above calculation, we have

$$
\begin{aligned}
E(\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n|\mathbf{S}_{n-1}) &= \frac{k}{n-1}\frac{tr(\mathbf{S}_{n-1}^{-1})}{k} \\
&= \frac{k}{n-1}\int \frac{1}{x} dF^{\mathbf{S}_{n-1}} \to \frac{\rho}{1-\rho},\ a.s.,
\end{aligned}
$$

$$
\begin{aligned}
Var(\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n|\mathbf{S}_{n-1}) &= \frac{2}{(n-1)^2} tr\left((\mathbf{S}_{n-1}^{-1})^2\right) \\
&\le \frac{2k}{(n-1)^2}(\frac{1}{\lambda_{\min}(\mathbf{S}_{n-1})})^2 \to 0,\ a.s..
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&E(\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n) = E\big(E(\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n|\mathbf{S}_{n-1})\big) \to \frac{\rho}{1-\rho},\ \text{and} \\
&Var(\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n) \to 0.
\end{aligned}
$$

These lead to the first result,

$$
E\left((\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n - \frac{\rho}{1-\rho})^2\right) \to 0.
$$

From Lemma 2.5, we have

$$\mathbf{x}_n^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_n = \frac{\mathbf{x}_n^\top(\sum_{j\neq n}\mathbf{x}_j\mathbf{x}_j^\top)^{-1}\mathbf{x}_n}{1+\mathbf{x}_n^\top(\sum_{j\neq n}\mathbf{x}_j\mathbf{x}_j^\top)^{-1}\mathbf{x}_n}$$

$$= \frac{\frac{1}{n-1}\mathbf{x}_n^\top\mathbf{S}_{n-1}^{-1}\mathbf{x}_n}{1+\frac{1}{n-1}\mathbf{x}_n^\top\mathbf{S}_{n-1}^{-1}\mathbf{x}_n}.$$

Let function $f(x)=\frac{x}{1+x}$. Its derivative satisfies $f'(x)=\frac{1}{(1+x)^2}\leq 1$, for $x\geq 0$. From $\mathbf{x}_n^\top\mathbf{S}_{n-1}^{-1}\mathbf{x}_n\geq 0$ and the mean value theorem, we get

$$\left|\mathbf{x}_n^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_n-\rho\right| \leq \left|\frac{1}{n-1}\mathbf{x}_n^\top\mathbf{S}_{n-1}^{-1}\mathbf{x}_n-\frac{\rho}{1-\rho}\right|,$$

which implies

$$E\left((\mathbf{x}_n^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_n-\rho)^2\right) \leq E\left((\frac{1}{n-1}\mathbf{x}_n^\top\mathbf{S}_{n-1}^{-1}\mathbf{x}_n-\frac{\rho}{1-\rho})^2\right)\to 0.$$

Furthermore, from $\frac{1}{n-1}\mathbf{x}_n^\top\mathbf{S}_{n-1}^{-1}\mathbf{x}_n \leq \lambda_{\min}^{-1}(\mathbf{S}_{n-1})\frac{1}{n-1}\mathbf{x}_n^\top\mathbf{x}_n \to \frac{1}{(1-\sqrt{\rho})^2}$ a.s., we obtain

$$\mathbf{x}_n^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_n \leq \frac{1}{1+(1-\sqrt{\rho})^2}, \quad a.s.,$$

and complete the proof.

□

Lemma 2.8. *Let $\boldsymbol{X}=(\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)$ be a random matrix with $\boldsymbol{x}_i$ i.i.d. from $\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_k)$. The matrix $\boldsymbol{H}$ is defined as $\boldsymbol{H}=(\boldsymbol{I}-\boldsymbol{P}_1)\boldsymbol{X}^\top\left(\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{P}_1)\boldsymbol{X}^\top\right)^{-1}\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{P}_1)$ and has its entries denoted by $\boldsymbol{H}_{ij}$. As $(k,n)\to\infty$ with $k/n\to\rho\in(0,1)$, we have*

$$\max_{i=1,\ldots,n} E\left[\left(\boldsymbol{H}_{ii}-\rho\right)^2\right]\to 0.$$

Proof. From Lemma 2.7, we get

$$E\left((n\bar{\mathbf{x}}^\top(\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top)^{-1}\bar{\mathbf{x}}-\frac{\rho}{1-\rho})^2\right)\to 0, \tag{2.17}$$

$$E\big((n\bar{\mathbf{x}}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\bar{\mathbf{x}}-\rho)^2\big)\to 0, n\bar{\mathbf{x}}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\bar{\mathbf{x}}\le \frac{1}{1+(1-\sqrt{\rho})^2},\ a.s., \tag{2.18}$$

$$E\big((\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_1-\rho)^2\big)\to 0. \tag{2.19}$$

The proof proceeds in two steps. First, we study $\mathbf{x}_1^\top\big(\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top\big)^{-1}\mathbf{x}_1$ and show that it converges to ρ in quadratic mean. Second, we divide $\mathbf{H}_{ii}$ into three parts and investigate them separately. Then we reach the statement in the lemma and complete the proof.

In the first step, we would show $\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_1$ is a well approximation to $\mathbf{x}_1^\top\big(\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top\big)^{-1}\mathbf{x}_1$ and then the convergence is guaranteed by (2.19). Lemma 2.5 and (2.18) imply

$$\big(\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top\big)^{-1}=(\mathbf{X}\mathbf{X}^\top)^{-1}+\frac{1}{1+g}(\mathbf{X}\mathbf{X}^\top)^{-1}n\bar{\mathbf{x}}\bar{\mathbf{x}}^\top(\mathbf{X}\mathbf{X}^\top)^{-1},$$

where $g=-n\bar{\mathbf{x}}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\bar{\mathbf{x}}\ge-\frac{1}{1+(1-\sqrt{\rho})^2}$ $a.s.$ is lower-bounded. Then, we have

$$\begin{aligned}
&\big|\mathbf{x}_1^\top\big(\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top\big)^{-1}\mathbf{x}_1-\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_1\big|\\
=&\frac{1}{1+g}\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}n\bar{\mathbf{x}}\bar{\mathbf{x}}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_1\\
=&\frac{n}{1+g}(\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\bar{\mathbf{x}})^2\\
\le&\frac{2}{1+g}\Big[\frac{1}{n}(\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_1)^2+\frac{1}{n}(\sum_{j\ne1}\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_j)^2\Big].
\end{aligned}$$

Based on (2.19), the expectation of the first part in the sum goes to 0. Then we show the second part $\frac{1}{n}(\sum_{j\ne1}\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_j)^2$ would also converge to 0 in the first mean. Define $\mathbf{A}_{1,j}=\sum_{k\ne1,j}\mathbf{x}_k\mathbf{x}_k^\top$ and $\mathbf{S}_{1,j}=\frac{1}{n-2}\mathbf{A}_{1,j}$. We have $\mathbf{X}\mathbf{X}^\top=\mathbf{A}_{1,j}+\mathbf{x}_1\mathbf{x}_1^\top+\mathbf{x}_j\mathbf{x}_j^\top$. From Lemma 2.5,

$$\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_j=\frac{\mathbf{x}_1^\top\mathbf{A}_{1,j}^{-1}\mathbf{x}_j}{D_{1,j}},$$

where $D_{1,j} = (1 + \mathbf{x}_1^\top \mathbf{A}_{1,j}^{-1}\mathbf{x}_1)(1 + \mathbf{x}_j^\top \mathbf{A}_{1,j}^{-1}\mathbf{x}_j) - (\mathbf{x}_1^\top \mathbf{A}_{1,j}^{-1}\mathbf{x}_j)^2 \geq 1$. Then

$$E\big(\frac{1}{n}(\sum_{j\neq 1} \mathbf{x}_1^\top (\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_j)^2\big) = E\big(\frac{1}{n}(\sum_{j\neq 1} \frac{\mathbf{x}_1^\top \mathbf{A}_{1,j}^{-1}\mathbf{x}_j}{D_{1,j}})^2\big)$$

$$= \sum_{j\neq 1} E(\frac{(\mathbf{x}_1^\top \mathbf{A}_{1,j}^{-1}\mathbf{x}_j)^2}{nD_{1,j}^2}) + \sum_{j\neq \ell\neq 1} E(\frac{\mathbf{x}_j^\top \mathbf{A}_{1,j}^{-1}\mathbf{A}_{1,\ell}^{-1}\mathbf{x}_\ell}{nD_{1,j}D_{1,\ell}}).$$

For any $j \neq \ell \neq 1$, we have

$$E(\frac{(\mathbf{x}_1^\top \mathbf{A}_{1,j}^{-1}\mathbf{x}_j)^2}{D_{1,j}^2}) = E(\frac{(\mathbf{x}_1^\top \mathbf{A}_{1,2}^{-1}\mathbf{x}_2)^2}{D_{1,2}^2}) \text{ and } E(\frac{\mathbf{x}_j^\top \mathbf{A}_{1,j}^{-1}\mathbf{A}_{1,\ell}^{-1}\mathbf{x}_\ell}{D_{1,j}D_{1,\ell}}) = E(\frac{\mathbf{x}_2^\top \mathbf{A}_{1,2}^{-1}\mathbf{A}_{1,3}^{-1}\mathbf{x}_3}{D_{1,2}D_{1,3}}).$$

Therefore,

$$E\big(\frac{1}{n}(\sum_{j\neq 1} \mathbf{x}_1^\top (\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_j)^2\big) = \frac{n-1}{n}E(\frac{(\mathbf{x}_1^\top \mathbf{A}_{1,2}^{-1}\mathbf{x}_2)^2}{D_{1,2}^2}) + \frac{(n-1)(n-2)}{n}E(\frac{\mathbf{x}_2^\top \mathbf{A}_{1,2}^{-1}\mathbf{A}_{1,3}^{-1}\mathbf{x}_3}{D_{1,2}D_{1,3}}). \tag{2.20}$$

Lemma 2.6 asserts the first part in (2.20) converges to 0 by

$$E(\frac{(\mathbf{x}_1^\top \mathbf{A}_{1,2}^{-1}\mathbf{x}_2)^2}{D_{1,2}^2}) \leq E\big((\mathbf{x}_1^\top \mathbf{A}_{1,2}^{-1}\mathbf{x}_2)^2\big) = \frac{k}{(n-2)^2}E(\frac{tr(\mathbf{S}_{1,2}^{-1})^2}{k}) \to 0.$$

Next, we study the second part and show it would also go to 0. Let $\mathbf{A}_{1,2,3} = \sum_{s\neq 1,2,3} \mathbf{x}_s\mathbf{x}_s^\top$, $\mathbf{S}_{1,2,3} = \frac{1}{n-3}\mathbf{A}_{1,2,3}$. Then, $g_3 = \mathbf{x}_3^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3 \geq 0$ and Lemma 2.5 gives the relationship

$$\mathbf{A}_{1,2}^{-1} = \mathbf{A}_{1,2,3}^{-1} - \frac{1}{1+g_3}\mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3\mathbf{x}_3^\top \mathbf{A}_{1,2,3}^{-1}.$$

From calculations and Lemma 2.6, we have

$$E(\mathbf{x}_2^\top (\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_3) = 0,\ E\big((\mathbf{x}_2^\top (\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_3)^2|\mathbf{A}_{1,2,3}\big) = \frac{tr\big((\mathbf{S}_{1,2,3}^{-1})^4\big)}{(n-3)^4} = O(n^{-3}),$$

$$E(\mathbf{x}_2^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3) = 0,\ E\big((\mathbf{x}_2^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3)^2|\mathbf{A}_{1,2,3}\big) = \frac{1}{(n-3)^2}tr\big((\mathbf{S}_{1,2,3}^{-1})^2\big) = O(n^{-1}),$$

$$(\mathbf{x}_2^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3)^2 \leq (\mathbf{x}_2^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_2)(\mathbf{x}_3^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3) \leq \frac{\rho^2}{(1-\sqrt{\rho})^4}\ a.s.,$$

$$(n-2)\mathbf{x}_2^\top(\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_2 \le \frac{k(n-2)}{(n-3)^2}\lambda_{\min}^{-2}(\mathbf{S}_{1,2,3})\frac{\mathbf{x}_2^\top\mathbf{x}_2}{k} \le \frac{\rho}{(1-\sqrt{\rho})^4}\ a.s..$$

These give two upper bounds

$$\frac{1+\frac{(\mathbf{x}_2^\top\mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3)^2}{(1+g_2)(1+g_3)}}{\frac{n}{n-1}D_{1,2}D_{1,3}} \le 1+\frac{\rho^2}{(1-\sqrt{\rho})^4}\ a.s.,$$

$$\frac{(n-2)(\frac{\mathbf{x}_2^\top(\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_2}{1+g_2}+\frac{\mathbf{x}_3^\top(\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_3}{1+g_3})}{\frac{n}{n-1}D_{1,2}D_{1,3}} \le \frac{2\rho}{(1-\sqrt{\rho})^4}\ a.s..$$

Then, we can get

$$(n-2)^2E\Big((\mathbf{x}_2^\top(\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_3\frac{1+\frac{(\mathbf{x}_2^\top\mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3)^2}{(1+g_2)(1+g_3)}}{\frac{n}{n-1}D_{1,2}D_{1,3}})^2\Big)\to 0,$$

$$E\Big((\mathbf{x}_2^\top\mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3\frac{(n-2)(\frac{\mathbf{x}_2^\top(\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_2}{1+g_2}+\frac{\mathbf{x}_3^\top(\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_3}{1+g_3})}{\frac{n}{n-1}D_{1,2}D_{1,3}})^2\Big)\to 0.$$

These together show

$$\begin{aligned}
&E\big[\frac{(n-1)(n-2)}{n}\frac{\mathbf{x}_2^\top\mathbf{A}_{1,2}^{-1}\mathbf{A}_{1,3}^{-1}\mathbf{x}_3}{D_{1,2}D_{1,3}}\big]\\
&=E\big[(n-2)\mathbf{x}_2^\top(\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_3\frac{1+\frac{(\mathbf{x}_2^\top\mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3)^2}{(1+g_2)(1+g_3)}}{\frac{n}{n-1}D_{1,2}D_{1,3}}\big]\\
&-E\big[\mathbf{x}_2^\top\mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3\frac{(n-2)(\frac{\mathbf{x}_2^\top(\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_2}{1+g_2}+\frac{\mathbf{x}_3^\top(\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_3}{1+g_3})}{\frac{n}{n-1}D_{1,2}D_{1,3}}\big]\\
&\to 0.
\end{aligned}$$

Hence, from (2.20), we derive $E(\frac{1}{n}(\sum_{j\neq 1}\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_j)^2)\to 0$. This together with an upper-bound inferred from (2.19) and (2.18) leads to

$$E\big[(\mathbf{x}_1^\top(\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top)^{-1}\mathbf{x}_1-\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_1)^2\big]\to 0.$$

And then (2.19) further shows

$$E\big[(\mathbf{x}_1^\top(\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top)^{-1}\mathbf{x}_1-\rho)^2\big]\to 0. \tag{2.21}$$

For any $i \in \{1, \ldots, n\}$, we divide $\mathbf{H}_{ii}$ into three parts

$$\begin{aligned}\mathbf{H}_{ii} &= (\mathbf{x}_i - \bar{\mathbf{x}})^\top \big(\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top\big)^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) \\ &= \bar{\mathbf{x}}^\top \big(\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top\big)^{-1}\bar{\mathbf{x}} - 2\mathbf{x}_i^\top \big(\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top\big)^{-1}\bar{\mathbf{x}} \\ &+ \mathbf{x}_i^\top \big(\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top\big)^{-1}\mathbf{x}_i.\end{aligned}$$

Based on (2.17) and (2.21), we obtain

$$\begin{aligned}E\big[(\mathbf{H}_{ii} - \rho)^2\big] &= E\Big[\big((\mathbf{x}_i - \bar{\mathbf{x}})^\top(\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top)^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) - \rho\big)^2\Big] \\ &= E\Big[\big((\mathbf{x}_1 - \bar{\mathbf{x}})^\top(\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top)^{-1}(\mathbf{x}_1 - \bar{\mathbf{x}}) - \rho\big)^2\Big] \\ &\le E\Big[3\big(\mathbf{x}_1^\top(\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top)^{-1}\mathbf{x}_1 - \rho\big)^2 + 3\big(\bar{\mathbf{x}}^\top(\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top)^{-1}\bar{\mathbf{x}}\big)^2 \\ &+ 12\big(\mathbf{x}_1^\top(\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top)^{-1}\bar{\mathbf{x}}\big)^2\Big] \\ &= o(1).\end{aligned}$$

Therefore,

$$\max_{i=1,\ldots,n} E\big[(\mathbf{H}_{ii} - \rho)^2\big] \to 0,$$

which completes the proof. □

Lemma 2.9. *Let $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ be i.i.d. m-variate random vectors satisfying $E(\boldsymbol{z}_i) = \boldsymbol{0}, Var(\boldsymbol{z}_i) = \boldsymbol{I}_m$ and $Var(\frac{\boldsymbol{z}_i^\top \boldsymbol{z}_i}{m}) = O(m^{-1})$. Suppose matrix $\boldsymbol{A}$ is uniformly distributed on the Stiefel manifold $\mathcal{V}_k(\mathbb{R}^m) = \{\boldsymbol{A} \in \mathbb{R}^{m\times k} : \boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{I}_k\}$ and is independent of $\boldsymbol{z}_i$. Let $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)^\top$ and*

$$\boldsymbol{H} = (\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Z}\boldsymbol{A}\left(\boldsymbol{A}^\top\boldsymbol{Z}^\top(\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Z}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^\top\boldsymbol{Z}^\top(\boldsymbol{I} - \boldsymbol{P}_1).$$

As $n, k, m \to \infty$, with $k/n \to \rho \in (0, 1)$ and m sufficiently larger than n, we have

$$\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{H}_{ii} - \rho\right)^2 = o_p(1),$$

where $\boldsymbol{H}_{ii}$ denote the i-th diagonal entries of $\boldsymbol{H}$.

Proof. Let $\mathbf{U}\boldsymbol{\Lambda}\mathbf{O}^\top$ be the singular value decomposition (SVD) of $\mathbf{Z}$, where $\mathbf{U}$ is an $n \times n$ orthogonal matrix, $\mathbf{O}$ is an $m \times m$ orthogonal matrix, and $\boldsymbol{\Lambda} = (\mathbf{D}, \mathbf{0})$ with $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$. Let $\mathbf{O}_n$ be the matrix consisting of first n columns of $\mathbf{O}$, then $\mathbf{Z}$ can be denoted as

$$\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{O}_n^\top. \tag{2.22}$$

In the first step, we study the properties of the entries of $\mathbf{D}$. Based on (2.22), we have

$$\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top = \frac{1}{m}\mathbf{U}\mathbf{D}^2\mathbf{U}^\top.$$

This indicates the diagonal entries of $\frac{1}{m}\mathbf{D}^2$ are the eigenvalues of $\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top$, then

$$\max_{i=1,\ldots,n}\left(\frac{d_i^2}{m} - 1\right)^2 = \lambda_{\max}\left\{\left(\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top - \mathbf{I}\right)^2\right\} \leq tr\left\{\left(\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top - \mathbf{I}\right)^2\right\}$$

From the properties of $\mathbf{z}_i$, we have

$$\begin{aligned} E\left\{tr\left[\left(\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top - \mathbf{I}\right)^2\right]\right\} &= \sum_{i=1}^n E\left\{\left(\frac{\mathbf{z}_i^\top\mathbf{z}_i}{m} - 1\right)^2\right\} + \sum_{i\neq j}^n E\left\{\left(\frac{\mathbf{z}_i^\top\mathbf{z}_j}{m}\right)^2\right\} \\ &= nVar\left(\frac{\mathbf{z}_1^\top\mathbf{z}_1}{m}\right) + \frac{n^2 - n}{m} \\ &= O(n^2m^{-1}). \end{aligned}$$

Therefore, from Markov's inequality, for any $t > 0$,

$$P\left\{\max_{i=1,\ldots,n}\left(\frac{d_i}{\sqrt{m}} - 1\right)^2 > t\right\} \leq P\left\{\max_{i=1,\ldots,n}\left(\frac{d_i^2}{m} - 1\right)^2 > t\right\} \leq O(n^2m^{-1}t^{-1}), \tag{2.23}$$

which shows the eigenvalues of $\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top$ are close to 1 when m is sufficiently larger than n.

Let $\mathbf{X} = (\mathbf{I} - \mathbf{P}_1)\,\mathbf{U}\mathbf{O}_n^\top\mathbf{A}$ and $\tilde{\mathbf{Z}} = (\mathbf{I} - \mathbf{P}_1)\,\mathbf{U}\frac{\mathbf{D}}{\sqrt{m}}\mathbf{O}_n^\top\mathbf{A}$. Since the hat matrix for $\tilde{\mathbf{Z}}$ and $(\mathbf{I} - \mathbf{P}_1)\,\mathbf{Z}\mathbf{A}$ are the same, the hat matrix for $\tilde{\mathbf{Z}}$ and $\mathbf{X}$ are denoted as

$$\mathbf{H} = \tilde{\mathbf{Z}}\left(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}}\right)^{-1}\tilde{\mathbf{Z}}^\top, \qquad \mathbf{S} = \mathbf{X}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top,$$

where $\mathbf{H}$ is the target matrix of the lemma. Let $\mathbf{S}_{ii}$ denote the i-th diagonal entry of the matrix $\mathbf{S}$. We will show $\mathbf{H}_{ii}$ and $\mathbf{S}_{ii}$ are close. Let $\mathbf{e}_i$ denote the vector with 1 in the i-th coordinate and 0's elsewhere. Define $\hat{\boldsymbol{\gamma}}_i^{ls} = \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{e}_i$. Based on the least square, then $\hat{\boldsymbol{\gamma}}_i^{ls}$ satisfies

$$\hat{\boldsymbol{\gamma}}_i^{ls} = \underset{\boldsymbol{\gamma}\in\mathbb{R}^k}{\operatorname{argmin}}\big|\big|\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{e}_i - \mathbf{X}\boldsymbol{\gamma}\big|\big|_2^2. \tag{2.24}$$

Similarly, define $\hat{\boldsymbol{\eta}}_i^{ls} = \left(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}}\right)^{-1}\tilde{\mathbf{Z}}^\top\mathbf{e}_i$. Then, it satisfies

$$\hat{\boldsymbol{\eta}}_i^{ls} = \underset{\boldsymbol{\eta}\in\mathbb{R}^k}{\operatorname{argmin}}\big|\big|\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{e}_i - \tilde{\mathbf{Z}}\boldsymbol{\eta}\big|\big|_2^2. \tag{2.25}$$

Based on (2.24) and (2.25), we have

$$\begin{aligned}\big|\big|\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{e}_i - \tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls}\big|\big|_2^2 &\le \big|\big|\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{e}_i - \tilde{\mathbf{Z}}\hat{\boldsymbol{\gamma}}_i^{ls}\big|\big|_2^2 \\ &= \big|\big|\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{e}_i - \mathbf{X}\hat{\boldsymbol{\gamma}}_i^{ls} + \big(\mathbf{X} - \tilde{\mathbf{Z}}\big)\hat{\boldsymbol{\gamma}}_i^{ls}\big|\big|_2^2 \\ &\le \Big(\big|\big|\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{e}_i - \mathbf{X}\hat{\boldsymbol{\gamma}}_i^{ls}\big|\big|_2 + \big|\big|\big(\mathbf{X} - \tilde{\mathbf{Z}}\big)\hat{\boldsymbol{\gamma}}_i^{ls}\big|\big|_2\Big)^2,\end{aligned} \tag{2.26}$$

and

$$\begin{aligned}\big|\big|\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{e}_i - \mathbf{X}\hat{\boldsymbol{\gamma}}_i^{ls}\big|\big|_2^2 &\le \big|\big|\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{e}_i - \mathbf{X}\hat{\boldsymbol{\eta}}_i^{ls}\big|\big|_2^2 \\ &= \big|\big|\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{e}_i - \tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls} + \big(\tilde{\mathbf{Z}} - \mathbf{X}\big)\hat{\boldsymbol{\eta}}_i^{ls}\big|\big|_2^2 \\ &\le \Big(\big|\big|\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{e}_i - \tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls}\big|\big|_2 + \big|\big|\big(\tilde{\mathbf{Z}} - \mathbf{X}\big)\hat{\boldsymbol{\eta}}_i^{ls}\big|\big|_2\Big)^2.\end{aligned} \tag{2.27}$$

To study (2.26) and (2.27), we first investigate the value of $\big|\big|\big(\mathbf{X} - \tilde{\mathbf{Z}}\big)\hat{\boldsymbol{\gamma}}_i^{ls}\big|\big|_2$ and $\big|\big|\big(\tilde{\mathbf{Z}} - \mathbf{X}\big)\hat{\boldsymbol{\eta}}_i^{ls}\big|\big|_2$. From Theorem 2.2.1 in Chikuse (2003), matrix $\mathbf{A}$ can be expressed as $\mathbf{A} = \mathbf{G}\left(\mathbf{G}^\top\mathbf{G}\right)^{-1/2}$, where the elements of $m \times k$ matrix $\mathbf{G}$ are i.i.d. from $\mathcal{N}(0,1)$. Let $\mathbf{E} = \mathbf{O}_n^\top \mathbf{G}$. Then $\mathbf{O}_n^\top \mathbf{A} = \mathbf{E}\left(\mathbf{G}^\top\mathbf{G}\right)^{-1/2}$. From Lemma 2.13, for any $h_1 > 0$ and $h_2 > 0$, the independence between $\mathbf{A}$ and $\mathbf{Z}$ leads to

$$\begin{aligned} P\left[\lambda_{\max}\left(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\right) \geq (1+\sqrt{k/n}+h_1)^2\right] &\leq \exp\left(-nh_1^2/2\right), \\ P\left[\lambda_{\min}\left(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\right) \leq (1-\sqrt{k/n}-h_2)^2\right] &\leq \exp\left(-nh_2^2/2\right). \end{aligned} \tag{2.28}$$

For any matrix $\mathbf{M}$, SVD shows the nonzero eigenvalues of $\mathbf{M}^\top\mathbf{M}$ and $\mathbf{M}\mathbf{M}^\top$ are the same. Therefore, with $k < n$, it indicates $\lambda_{\min}\big(\mathbf{E}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{U}\mathbf{E}\big) = \lambda_{\min}\big(\mathbf{E}^\top\mathbf{E}\big)$ and $\lambda_{\min}\big(\mathbf{E}^\top\frac{\mathbf{D}}{\sqrt{m}}\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{U}\frac{\mathbf{D}}{\sqrt{m}}\mathbf{E}\big) = \lambda_{\min}\big(\mathbf{E}^\top\frac{\mathbf{D}^2}{m}\mathbf{E}\big)$. Based on the property that $\lambda_{\max}(\mathbf{M}^\top\mathbf{M}) = \lambda_{\max}(\mathbf{M}\mathbf{M}^\top)$ and (2.28), we have

$$\begin{aligned} \lambda_{\max}\left(\mathbf{X}\big(\mathbf{X}^\top\mathbf{X}\big)^{-1}\mathbf{A}^\top\mathbf{O}_n\mathbf{O}_n^\top\mathbf{A}\big(\mathbf{X}^\top\mathbf{X}\big)^{-1}\mathbf{X}^\top\right) &= \lambda_{\max}\left(\mathbf{E}\left(\mathbf{E}^\top\mathbf{U}^\top\big(\mathbf{I}-\mathbf{P}_1\big)\mathbf{U}\mathbf{E}\right)^{-1}\mathbf{E}^\top\right) \\ &\leq \lambda_{\max}\big(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\big)\frac{1}{\lambda_{\min}\big(\frac{1}{n}\mathbf{E}^\top\mathbf{U}^\top\big(\mathbf{I}-\mathbf{P}_1\big)\mathbf{U}\mathbf{E}\big)} \\ &\leq \frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2}, \end{aligned} \tag{2.29}$$

and

$$\begin{aligned} \lambda_{\max}\left(\tilde{\mathbf{Z}}(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}})^{-1}\mathbf{A}^\top\mathbf{O}_n\mathbf{O}_n^\top\mathbf{A}\big(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}}\big)^{-1}\tilde{\mathbf{Z}}^\top\right) &= \lambda_{\max}\left(\mathbf{E}\left(\mathbf{E}^\top\frac{\mathbf{D}}{\sqrt{m}}\mathbf{U}^\top\big(\mathbf{I}-\mathbf{P}_1\big)\mathbf{U}\frac{\mathbf{D}}{\sqrt{m}}\mathbf{E}\right)^{-1}\mathbf{E}^\top\right) \\ &\leq \lambda_{\max}\big(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\big)\frac{1}{\lambda_{\min}\big(\frac{1}{n}\mathbf{E}^\top\frac{\mathbf{D}}{\sqrt{m}}\mathbf{U}^\top\big(\mathbf{I}-\mathbf{P}_1\big)\mathbf{U}\frac{\mathbf{D}}{\sqrt{m}}\mathbf{E}\big)} \\ &\leq \frac{1}{\lambda_{\min}\big(\frac{\mathbf{D}^2}{m}\big)}\cdot\frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2} \end{aligned} \tag{2.30}$$

with probability at least $1-\exp\left(-nh_1^2/2\right)-\exp\left(-nh_2^2/2\right)$. Based on (2.23), (2.29) and (2.30), upper bounds can be derived as follows.

$$
\begin{aligned}
\big|\big|(\mathbf{X}-\tilde{\mathbf{Z}})\hat{\boldsymbol{\gamma}}_i^{ls}\big|\big|_2^2 &= \big|\big|(\mathbf{I}-\mathbf{P}_1)\mathbf{U}(\mathbf{I}-\frac{\mathbf{D}}{\sqrt{m}})\mathbf{O}_n^\top\mathbf{A}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{e}_i\big|\big|_2^2 \\
&\leq \max_{i=1,\ldots,n}\big(1-\frac{d_i}{\sqrt{m}}\big)^2\big|\big|\mathbf{O}_n^\top\mathbf{A}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{e}_i\big|\big|_2^2 \\
&\leq t\cdot\frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2}
\end{aligned}
\tag{2.31}
$$

and

$$
\begin{aligned}
\big|\big|(\tilde{\mathbf{Z}}-\mathbf{X})\hat{\boldsymbol{\eta}}_i^{ls}\big|\big|_2^2 &= \big|\big|(\mathbf{I}-\mathbf{P}_1)\mathbf{U}(\mathbf{I}-\frac{\mathbf{D}}{\sqrt{m}})\mathbf{O}_n^\top\mathbf{A}(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}^\top\mathbf{e}_i\big|\big|_2^2 \\
&\leq \max_{i=1,\ldots,n}\big(1-\frac{d_i}{\sqrt{m}}\big)^2\big|\big|\mathbf{O}_n^\top\mathbf{A}(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}^\top\mathbf{e}_i\big|\big|_2^2 \\
&\leq \max_{i=1,\ldots,n}\big(1-\frac{d_i}{\sqrt{m}}\big)^2\cdot\frac{1}{\min_{i=1,\ldots,n}\big(\frac{d_i^2}{m}\big)}\cdot\frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2} \\
&\leq \frac{t}{(1-\sqrt{t})^2}\cdot\frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2},
\end{aligned}
\tag{2.32}
$$

with probability at least $1-O(n^2m^{-1}t^{-1})-\exp(-nh_1^2/2)-\exp(-nh_2^2/2)$. Combining (2.26), (2.27), (2.31) and (2.32), with $h_1=n^{-1/4}$, $h_2=n^{-1/4}$ and $t=n^{-c}$, where c is a positive constant, we have

$$
\big|\big|\left(\mathbf{I}-\mathbf{P}_1\right)\mathbf{e}_i-\tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls}\big|\big|_2^2 \leq \big|\big|\left(\mathbf{I}-\mathbf{P}_1\right)\mathbf{e}_i-\mathbf{X}\hat{\boldsymbol{\gamma}}_i^{ls}\big|\big|_2^2+3n^{-c/2}\cdot\frac{1+\sqrt{k/n}+n^{-1/4}}{1-\sqrt{k/n}-n^{-1/4}},
$$

$$
\big|\big|\left(\mathbf{I}-\mathbf{P}_1\right)\mathbf{e}_i-\mathbf{X}\hat{\boldsymbol{\gamma}}_i^{ls}\big|\big|_2^2 \leq \big|\big|\left(\mathbf{I}-\mathbf{P}_1\right)\mathbf{e}_i-\tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls}\big|\big|_2^2+\frac{3}{n^{c/2}-1}\cdot\frac{1+\sqrt{k/n}+n^{-1/4}}{1-\sqrt{k/n}-n^{-1/4}}
$$

with probability at least $1-O(n^{2+c}m^{-1})-2\exp\left(-n^{1/2}/2\right)$. Since $||\left(\mathbf{I}-\mathbf{P}_1\right)\mathbf{e}_i-\tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls}||_2^2=\mathbf{e}_i^\top\left(\mathbf{I}-\mathbf{P}_1\right)\mathbf{e}_i-\mathbf{H}_{ii}$ and $||\left(\mathbf{I}-\mathbf{P}_1\right)\mathbf{e}_i-\mathbf{X}\hat{\boldsymbol{\gamma}}_i^{ls}||_2^2=\mathbf{e}_i^\top\left(\mathbf{I}-\mathbf{P}_1\right)\mathbf{e}_i-\mathbf{S}_{ii}$, and the above derivation is valid for any $\mathbf{e}_i$, we obtain

$$
|\mathbf{H}_{ii}-\mathbf{S}_{ii}| \leq \frac{3}{n^{c/2}-1}\cdot\frac{1+\sqrt{k/n}+n^{-1/4}}{1-\sqrt{k/n}-n^{-1/4}}, \qquad i=1,\ldots,n,
$$

with probability at least $1 - O(n^{2+c}m^{-1}) - 2\exp\left(-n^{1/2}/2\right)$. When $n \to \infty$ and $n^{2+c}m^{-1} = o(1)$, there is a constant $C \geq \frac{12+12\sqrt{\rho}}{1-\sqrt{\rho}}$ such that

$$P\left[\max_{i=1\ldots,n} |\mathbf{H}_{ii} - \mathbf{S}_{ii}| \geq Cn^{-c/2}\right] = o(1). \tag{2.33}$$

According to the definitions of $\mathbf{X}$ and $\mathbf{A}$, the hat matrix $\mathbf{S}$ can be denoted as

$$\mathbf{S} = \left(\mathbf{I} - \mathbf{P}_1\right)\mathbf{U}\mathbf{O}_n^\top\mathbf{G}\left(\mathbf{G}^\top\mathbf{O}_n\mathbf{U}^\top\left(\mathbf{I} - \mathbf{P}_1\right)\mathbf{U}\mathbf{O}_n^\top\mathbf{G}\right)^{-1}\mathbf{G}^\top\mathbf{O}_n\mathbf{U}^\top\left(\mathbf{I} - \mathbf{P}_1\right),$$

where $\mathbf{U}\mathbf{O}_n^\top$ is independent of $\mathbf{G}$ and satisfies $\mathbf{U}\mathbf{O}_n^\top\mathbf{O}_n\mathbf{U}^\top = \mathbf{I}_n$. From the definition of $\mathbf{G}$, Lemma 2.8 and the dominated convergence theorem, we obtain

$$E\left[\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{S}_{ii} - \rho\right)^2\right] \to 0.$$

Then, $\frac{1}{n}\sum_{i=1}^{n}(\mathbf{S}_{ii} - \rho)^2 = o_p(1)$ can be derived based on Markov's inequality. Combining this with (2.33) and Slutsky's theorem, it shows

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{H}_{ii} - \rho\right)^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{H}_{ii} - \mathbf{S}_{ii} + \mathbf{S}_{ii} - \rho\right)^2 \\
&\leq \frac{2}{n}\sum_{i=1}^{n}\left(\mathbf{H}_{ii} - \mathbf{S}_{ii}\right)^2 + \frac{2}{n}\sum_{i=1}^{n}\left(\mathbf{S}_{ii} - \rho\right)^2 \\
&\leq \max_{i=1\ldots,n} 2\left(\mathbf{H}_{ii} - \mathbf{S}_{ii}\right)^2 + \frac{2}{n}\sum_{i=1}^{n}\left(\mathbf{S}_{ii} - \rho\right)^2 \\
&= o_p(1),
\end{aligned}$$

which completes the proof. □

Conditional on $\mathbf{A}^\top\mathbf{z}$, Theorem 2.1 in Steinberger and Leeb (2018) showed that the mean of $\mathbf{z}$ is approximately linear in $\mathbf{A}^\top\mathbf{z}$ under certain conditions. Based on this result, we derived the following lemma.

Lemma 2.10. *Suppose m-variate random vector $\boldsymbol{z} = (z_1, \ldots, z_m)^\top$ has a Lebesgue density $f_{\boldsymbol{z}}$ and satisfies $E(\boldsymbol{z}) = \boldsymbol{0}$ and $E(\boldsymbol{z}\boldsymbol{z}^\top) = \boldsymbol{I}_m$. For all $i = 1, \ldots, m$, the components z_i are independent and the moments satisfy $E(z_i^{20}) \leq C$ for some constants C. And all the marginal densities of the components of $\boldsymbol{z}$ are bounded by a constant $D \geq 1$. Suppose matrix $\boldsymbol{A}$ is uniformly distributed on the Stiefel manifold $\mathcal{V}_k(\mathbb{R}^m) = \{\boldsymbol{A} \in \mathbb{R}^{m\times k} : \boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{I}_k\}$. Let $\nu_{m,k}$ denote the uniform distribution on $\mathcal{V}_k(\mathbb{R}^m)$. Let $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ be the i.i.d. copies of $\boldsymbol{z}$ and $\boldsymbol{A}$ be independent of $\boldsymbol{z}_i$. For any nonzero vector $\boldsymbol{b} \in \mathbb{R}^m$, as $n \to \infty$, with $k/n \to \rho \in (0,1)$ and m sufficiently larger than n, there is a series of Borel set $F_n \subseteq \mathcal{V}_k(\mathbb{R}^m)$ such that*

$$\sup_{\boldsymbol{A}\in F_n} P\left(\sum_{i=1}^n \left(E(\boldsymbol{b}^\top\boldsymbol{z}_i|\boldsymbol{A}^\top\boldsymbol{z}_i) - \boldsymbol{b}^\top\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{z}_i\right)^2 > ||\boldsymbol{b}||_2^2\right) = o(1),$$

$$\sup_{\boldsymbol{A}\in F_n} P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \left|Var\big(\boldsymbol{b}^\top\boldsymbol{z}_i|\boldsymbol{A}^\top\boldsymbol{z}_i\big) - \boldsymbol{b}^\top\left(\boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{A}^\top\right)\boldsymbol{b}\right| > 5||\boldsymbol{b}||_2^2\right) = o(1),$$

and $\nu_{m,k}(F_n) \to 1$.

Proof. Based on Example 3.1 and Theorem 2.1 given in Steinberger and Leeb (2018), for each $\tau \in (0,1)$, there is a Borel set $F_n \subseteq \mathcal{V}_k(\mathbb{R}^m)$ such that

$$\sup_{\mathbf{A}\in F_n} P\left(\left|\left|E\big(\mathbf{z}|\mathbf{A}^\top\mathbf{z}\big) - \mathbf{A}\mathbf{A}^\top\mathbf{z}\right|\right|_2 > t\right) \leq \frac{m^{-\tau/10}}{t} + \frac{\gamma_2}{1-\tau}\frac{2k}{\log m},$$

$$\sup_{\mathbf{A}\in F_n} P\left(\left|\left|E\big(\mathbf{z}\mathbf{z}^\top|\mathbf{A}^\top\mathbf{z}\big) - \left(\mathbf{I}_m - \mathbf{A}\mathbf{A}^\top + \mathbf{A}\mathbf{A}^\top\mathbf{z}\mathbf{z}^\top\mathbf{A}\mathbf{A}^\top\right)\right|\right|_{sp} > t\right) \leq \frac{m^{-\tau/10}}{t} + \frac{\gamma_2}{1-\tau}\frac{2k}{\log m},$$

for each $t > 0$, and such that $\nu_{m,k}(F_n^{\mathsf{c}}) \leq \kappa_2 m^{-(\tau/10)\cdot(1-\frac{\gamma_2}{\tau}\frac{10k}{\log m})}$, where κ_2 and γ_2 are constants. Therefore, when $t = n^{-1/2}$, we have

$$\begin{aligned}\sup_{\mathbf{A}\in F_n} P\left(\sum_{i=1}^n \left|\left|E\big(\mathbf{z}_i|\mathbf{A}^\top\mathbf{z}_i\big) - \mathbf{A}\mathbf{A}^\top\mathbf{z}_i\right|\right|_2^2 > 1\right) &\leq \sum_{i=1}^n \sup_{\mathbf{A}\in F_n} P\left(\left|\left|E\big(\mathbf{z}_i|\mathbf{A}^\top\mathbf{z}_i\big) - \mathbf{A}\mathbf{A}^\top\mathbf{z}_i\right|\right|_2 > t\right) \\ &\leq n^{3/2}m^{-\tau/10} + \frac{\gamma_2}{1-\tau}\frac{2nk}{\log m},\end{aligned} \tag{2.34}$$

$$\sup_{\mathbf{A}\in F_n} P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|\left|E\left[\mathbf{z}_i\mathbf{z}_i^\top\middle|\mathbf{A}^\top\mathbf{z}_i\right]-\mathbf{A}\mathbf{A}^\top\mathbf{z}_i\mathbf{z}_i^\top\mathbf{A}\mathbf{A}^\top-\left(\mathbf{I}_m-\mathbf{A}\mathbf{A}^\top\right)\right|\right|_{sp}>1\right)$$

$$\leq\sum_{i=1}^{n}\sup_{\mathbf{A}\in F_n} P\left(\left|\left|E\left[\mathbf{z}_i\mathbf{z}_i^\top\middle|\mathbf{A}^\top\mathbf{z}_i\right]-\mathbf{A}\mathbf{A}^\top\mathbf{z}_i\mathbf{z}_i^\top\mathbf{A}\mathbf{A}^\top-\left(\mathbf{I}_m-\mathbf{A}\mathbf{A}^\top\right)\right|\right|_{sp}>t\right)$$

$$\leq n^{3/2}m^{-\tau/10}+\frac{\gamma_2}{1-\tau}\frac{2nk}{\log m}, \tag{2.35}$$

and $\nu_{m,k}(F_n^{\mathsf{c}})\leq\kappa_2 m^{-(\tau/10)\cdot(1-\frac{\gamma_2}{\tau}\frac{10k}{\log m})}$.

For each i, define $r_i=E(\mathbf{b}^\top\mathbf{z}_i|\mathbf{A}^\top\mathbf{z}_i)-\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{z}_i$ and $q_i=\mathbf{b}^\top\mathbf{z}_i-E(\mathbf{b}^\top\mathbf{z}_i|\mathbf{A}^\top\mathbf{z}_i)$. Based on the definition of the conditional variance, we could derive

$$Var\big(q_i|\mathbf{A}^\top\mathbf{z}_i\big)=\mathbf{b}^\top E\left[\mathbf{z}_i\mathbf{z}_i^\top\middle|\mathbf{A}^\top\mathbf{z}_i\right]\mathbf{b}-E\left[\mathbf{b}^\top\mathbf{z}_i|\mathbf{A}^\top\mathbf{z}_i\right]^2,$$

then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|Var\big(q_i|\mathbf{A}^\top\mathbf{z}_i\big)-\mathbf{b}^\top\left(\mathbf{I}_m-\mathbf{A}\mathbf{A}^\top\right)\mathbf{b}\right|$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|\mathbf{b}^\top\left\{E\left[\mathbf{z}_i\mathbf{z}_i^\top\middle|\mathbf{A}^\top\mathbf{z}_i\right]-\mathbf{A}\mathbf{A}^\top\mathbf{z}_i\mathbf{z}_i^\top\mathbf{A}\mathbf{A}^\top-\left(\mathbf{I}_m-\mathbf{A}\mathbf{A}^\top\right)\right\}\mathbf{b}-2\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{z}_i r_i-r_i^2\right|$$

$$\leq||\mathbf{b}||_2^2\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|\left|E\left[\mathbf{z}_i\mathbf{z}_i^\top\middle|\mathbf{A}^\top\mathbf{z}_i\right]-\mathbf{A}\mathbf{A}^\top\mathbf{z}_i\mathbf{z}_i^\top\mathbf{A}\mathbf{A}^\top-\left(\mathbf{I}_m-\mathbf{A}\mathbf{A}^\top\right)\right|\right|_{sp}$$

$$+2\sqrt{\sum_{i=1}^{n}\frac{(\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{z}_i)^2}{n}}\sqrt{\sum_{i=1}^{n}r_i^2}+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}r_i^2. \tag{2.36}$$

From the calculation,

$$Var\left\{\left(\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{z}_i\right)^2\right\}\leq(C^{1/5}+1)\left(\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{b}\right)^2,$$

Markov's inequality leads to

$$P\left(\sum_{i=1}^{n}\frac{(\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{z}_i)^2}{n} > 2\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{b}\right) \leq \frac{C^{1/5}+1}{n}. \tag{2.37}$$

According to Cauchy–Schwarz inequality,

$$r_i^2 = \left\{E(\mathbf{b}^\top\mathbf{z}_i|\mathbf{A}^\top\mathbf{z}_i) - \mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{z}_i\right\}^2 \leq \big|\big|\mathbf{b}\big|\big|_2^2 \cdot \big|\big|E\big(\mathbf{z}_i|\mathbf{A}^\top\mathbf{z}_i\big) - \mathbf{A}\mathbf{A}^\top\mathbf{z}_i\big|\big|_2^2.$$

Therefore, combining (2.34), (2.35), (2.36) and (2.37), we can derive

$$\sup_{\mathbf{A}\in F_n} P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big|Var\big(q_i|\mathbf{A}^\top\mathbf{z}_i\big) - \mathbf{b}^\top\big(\mathbf{I}_m - \mathbf{A}\mathbf{A}^\top\big)\mathbf{b}\big| > 5||\mathbf{b}||_2^2\right)$$

$$\leq \sup_{\mathbf{A}\in F_n} P\big(\sum_{i=1}^{n} r_i^2 > ||\mathbf{b}||_2^2\big) + \sup_{\mathbf{A}\in F_n} P\big(\sum_{i=1}^{n}\frac{(\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{z}_i)^2}{n} > 2\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{b}\big)$$

$$+ \sup_{\mathbf{A}\in F_n} P\big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big|\big|E\big[\mathbf{z}_i\mathbf{z}_i^\top|\mathbf{A}^\top\mathbf{z}_i\big] - \mathbf{A}\mathbf{A}^\top\mathbf{z}_i\mathbf{z}_i^\top\mathbf{A}\mathbf{A}^\top - \big(\mathbf{I}_m - \mathbf{A}\mathbf{A}^\top\big)\big|\big|_{sp} > 1\big)$$

$$\leq 2n^{3/2}m^{-\tau/10} + \frac{\gamma_2}{1-\tau}\frac{4nk}{\log m} + \frac{2C}{n}.$$

When m is sufficiently large such that $n^2 = o(\log m)$, as $n \to \infty$, we have

$$\sup_{\mathbf{A}\in F_n} P\left(\sum_{i=1}^{n} r_i^2 > ||\mathbf{b}||_2^2\right) = o(1).$$

and

$$\sup_{\mathbf{A}\in F_n} P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big|Var\big(q_i|\mathbf{A}^\top\mathbf{z}_i\big) - \mathbf{b}^\top\big(\mathbf{I}_m - \mathbf{A}\mathbf{A}^\top\big)\mathbf{b}\big| > 5||\mathbf{b}||_2^2\right) = o(1),$$

where $\nu_{m,k}(F_n) \to 1$. The proof is completed. □

2.6.3 Proof of Lemma 2.2

First we present a trace inequality (Lopes, Jacob, and Wainwright, 2011, Lemma 2).

Lemma 2.11. *If* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *are square matrices of the same size with* $\boldsymbol{A} \succeq \boldsymbol{0}$ *and* $\boldsymbol{B} = \boldsymbol{B}^\top$*, then*

$$\lambda_{\min}(\boldsymbol{B}) tr(\boldsymbol{A}) \leq tr(\boldsymbol{A}\boldsymbol{B}) \leq \lambda_{\max}(\boldsymbol{B}) tr(\boldsymbol{A}).$$

Some results for Gaussian concentration inequalities will be introduced. The following concentration bounds for Gaussian quadratic forms are given in Bechar (2009).

Lemma 2.12. *Let* $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ *with* $\boldsymbol{A} \succeq \boldsymbol{0}$ *and* $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_p)$*. For any* $t > 0$*, we have*

$$P\left[\boldsymbol{z}^\top \boldsymbol{A}\boldsymbol{z} \geq tr(\boldsymbol{A}) + 2||\boldsymbol{A}||_F \sqrt{t} + 2||\boldsymbol{A}||_{sp} t\right] \leq \exp(-t), \quad \textit{and}$$

$$P\left[\boldsymbol{z}^\top \boldsymbol{A}\boldsymbol{z} \leq tr(\boldsymbol{A}) - 2||\boldsymbol{A}||_F \sqrt{t}\right] \leq \exp(-t).$$

Davidson and Szarek (2001, Theorem 2.13) gave an upper-bound and a lower-bound on the extreme eigenvalues of Wishart matrices.

Lemma 2.13. *For* $k \leq p$*, let* $\boldsymbol{P}_k \in \mathbb{R}^{p\times k}$ *be a random matrix with i.i.d.* $\mathcal{N}(0,1)$ *entries. Then, for all* $t \geq 0$*, we have*

$$P\left[\lambda_{\max}(\frac{1}{p}\boldsymbol{P}_k^\top \boldsymbol{P}_k) \geq (1 + \sqrt{k/p} + t)^2\right] \leq \exp(-pt^2/2), \quad \textit{and}$$

$$P\left[\lambda_{\min}(\frac{1}{p}\boldsymbol{P}_k^\top \boldsymbol{P}_k) \leq (1 - \sqrt{k/p} - t)^2\right] \leq \exp(-pt^2/2).$$

As a restatement of partial proof in Lopes, Jacob, and Wainwright (2011, Lemma 5), we obtain an upper bound for $tr(\mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k)$.

Lemma 2.14. *For* $k \leq p$*, let* $\boldsymbol{P}_k \in \mathbb{R}^{p\times k}$ *be a random matrix with i.i.d.* $\mathcal{N}(0,1)$ *entries. Suppose matrix* $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ *satisfies* $\boldsymbol{\Sigma} \succeq \boldsymbol{0}$*. Then, as* $(k,p) \to \infty$*, for any constant* $C > 1$*, we have*

$$P\left[tr(\boldsymbol{P}_k^\top \boldsymbol{\Sigma} \boldsymbol{P}_k) \leq Cktr(\boldsymbol{\Sigma})\right] \to 1.$$

Proof. Let $\mathbf{U}^\top \mathbf{D} \mathbf{U}$ be a spectral decomposition of $\mathbf{\Sigma}$. Then $\mathbf{P}_k^\top \mathbf{\Sigma} \mathbf{P}_k$ can be written as $(\mathbf{U}\mathbf{P}_k)^\top \mathbf{D}(\mathbf{U}\mathbf{P}_k)$. As $\mathbf{U}\mathbf{P}_k$ has the same distribution as $\mathbf{P}_k$, $\mathbf{P}_k^\top \mathbf{\Sigma} \mathbf{P}_k$ is distributed as $\mathbf{P}_k^\top \mathbf{D} \mathbf{P}_k$. In the following, we work under $\mathbf{P}_k^\top \mathbf{D} \mathbf{P}_k$.

Let $\boldsymbol{\xi}_i$ be the i-th column of $\mathbf{P}_k$ and $\mathbf{Z}^\top = (\boldsymbol{\xi}_1^\top, \dots, \boldsymbol{\xi}_k^\top)$. Then $\mathbf{Z} \in \mathbb{R}^{pk\times 1}$ and is distributed as $\mathcal{N}(\mathbf{0}, \mathbf{I}_{pk})$. Likewise, let $\tilde{\boldsymbol{D}} \in \mathbb{R}^{pk\times pk}$ be a diagonal matrix obtained by arranging k copies of $\mathbf{D}$ along the diagonal, i.e.

$$\tilde{\boldsymbol{D}} := \begin{pmatrix} \mathbf{D} & & \\ & \ddots & \\ & & \mathbf{D} \end{pmatrix}.$$

Consider the diagonal entries of $\mathbf{P}_k^\top \mathbf{D} \mathbf{P}_k$

$$tr(\mathbf{P}_k^\top \mathbf{D} \mathbf{P}_k) = \sum_{i=1}^{k} \boldsymbol{\xi}_i^\top \mathbf{D} \boldsymbol{\xi}_i = \mathbf{Z}^\top \tilde{\boldsymbol{D}} \mathbf{Z}.$$

Applying Lemma 2.12 to the quadratic form $\mathbf{Z}^\top \tilde{\boldsymbol{D}} \mathbf{Z}$, and noting that $\frac{||\mathbf{D}||_F}{tr(\mathbf{D})}$ and $\frac{||\mathbf{D}||_{sp}}{tr(\mathbf{D})}$ are at most 1, we get

$$\begin{aligned} tr(\mathbf{P}_k^\top \mathbf{D} \mathbf{P}_k) &\leq tr(\tilde{\boldsymbol{D}}) + 2||\tilde{\boldsymbol{D}}||_F \sqrt{t_1} + 2||\tilde{\boldsymbol{D}}||_{sp} t_1 \\ &= ktr(\mathbf{D}) + 2||\mathbf{D}||_F \sqrt{t_1 k} + 2||\mathbf{D}||_{sp} t_1 \\ &\leq ktr(\mathbf{\Sigma})(1 + \frac{2\sqrt{t_1}}{\sqrt{k}} + \frac{2t_1}{k}) \end{aligned}$$

with probability at least $1 - \exp(-t_1)$.

Choose $t_1 = \sqrt{k}$. The probability of the event tends to 1 as $(k, p) \to \infty$ with

$$(1 + \frac{2\sqrt{t_1}}{\sqrt{k}} + \frac{2t_1}{k}) \to 1.$$

Hence, for large k and any constant $C > 1$, we can obtain $(1 + \frac{2\sqrt{t_1}}{\sqrt{k}} + \frac{2t_1}{k}) < C$ and complete the proof. $\square$

Proof of Lemma 2.2. Let $\mathbf{U}^\top\mathbf{D}\mathbf{U}$ be a spectral decomposition of $\boldsymbol{\Sigma}$, where $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_p)$ and $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$. From this decomposition,

$$\sqrt{n}||\boldsymbol{\Gamma}^\top\boldsymbol{\beta} - \boldsymbol{\Gamma}^\top\mathbf{P}_k\boldsymbol{\eta}||_2^2 = \sqrt{n}||\sqrt{\mathbf{D}}\mathbf{U}\boldsymbol{\beta} - \sqrt{\mathbf{D}}\mathbf{U}\mathbf{P}_k\boldsymbol{\eta}||_2^2. \tag{2.38}$$

To cover general cases, we assume $\boldsymbol{\beta}/||\boldsymbol{\beta}||_2$ distributed uniformly on the unit sphere. Then, we work under the assumption $\boldsymbol{\beta}/||\boldsymbol{\beta}||_2 = \boldsymbol{\delta}/\sqrt{p}$, where $\boldsymbol{\delta}$ follows $\mathcal{N}(\mathbf{0}, \mathbf{I}_p)$. In light of this, $\mathbf{U}\boldsymbol{\beta}/||\boldsymbol{\beta}||_2$ and $\boldsymbol{\beta}/||\boldsymbol{\beta}||_2$ have the same distributions and then $\mathbf{U}\boldsymbol{\beta}/||\boldsymbol{\beta}||_2$ is denoted by $\boldsymbol{\delta}/\sqrt{p}$ for simplicity. For the same reason, we denote $\mathbf{U}\mathbf{P}_k$ as $\mathbf{P}_k$.

For the s given in Assumption A6, we let $\boldsymbol{\delta} = (\boldsymbol{\delta}_s^\top, \boldsymbol{\delta}_{p-s}^\top)^\top$, where $\boldsymbol{\delta}_s \in \mathbb{R}^s$ and $\boldsymbol{\delta}_{p-s} \in \mathbb{R}^{p-s}$. Correspondingly, $\mathbf{D}$ is divided into $\mathbf{D}_s$ and $\mathbf{D}_{p-s}$, where $\mathbf{D}_s = \mathrm{diag}(d_1, \ldots, d_s)$ and $\mathbf{D}_{p-s} = \mathrm{diag}(d_{s+1}, \ldots, d_p)$. Let $\mathbf{P}_k = (\mathbf{P}_{s,k}^\top, \mathbf{P}_{p-s,k}^\top)^\top$ with $\mathbf{P}_{s,k} \in \mathbb{R}^{s\times k}$ and $\mathbf{P}_{p-s,k} \in \mathbb{R}^{(p-s)\times k}$. We define $\boldsymbol{\eta}_0 \in \mathbb{R}^k$ as

$$\boldsymbol{\eta}_0 = \mathbf{P}_{s,k}^\top(\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\frac{\boldsymbol{\delta}_s}{\sqrt{p}}.$$

Plugging $\boldsymbol{\eta}_0$ into (2.38), we have

$$\begin{aligned}
&\min_{\boldsymbol{\eta}\in\mathbb{R}^k} \frac{\sqrt{n}||\boldsymbol{\Gamma}^\top\boldsymbol{\beta} - \boldsymbol{\Gamma}^\top\mathbf{P}_k\boldsymbol{\eta}||_2^2}{||\boldsymbol{\beta}||_2^2} \\
&= \min_{\boldsymbol{\eta}\in\mathbb{R}^k} \sqrt{n}||\sqrt{\mathbf{D}}\frac{\boldsymbol{\delta}}{\sqrt{p}} - \sqrt{\mathbf{D}}\mathbf{P}_k\boldsymbol{\eta}||_2^2 \\
&= \min_{\boldsymbol{\eta}\in\mathbb{R}^k} \sqrt{n}\left(||\sqrt{\mathbf{D}_s}(\frac{\boldsymbol{\delta}_s}{\sqrt{p}} - \mathbf{P}_{s,k}\boldsymbol{\eta})||_2^2 + ||\sqrt{\mathbf{D}_{p-s}}(\frac{\boldsymbol{\delta}_{p-s}}{\sqrt{p}} - \mathbf{P}_{p-s,k}\boldsymbol{\eta})||_2^2\right) \\
&\leq \sqrt{n}||\sqrt{\mathbf{D}_s}(\frac{\boldsymbol{\delta}_s}{\sqrt{p}} - \mathbf{P}_{s,k}\boldsymbol{\eta}_0)||_2^2 + \sqrt{n}||\sqrt{\mathbf{D}_{p-s}}(\frac{\boldsymbol{\delta}_{p-s}}{\sqrt{p}} - \mathbf{P}_{p-s,k}\boldsymbol{\eta}_0)||_2^2 \\
&= \sqrt{n}||\sqrt{\mathbf{D}_{p-s}}\frac{\boldsymbol{\delta}_{p-s}}{\sqrt{p}} - \sqrt{\mathbf{D}_{p-s}}\mathbf{P}_{p-s,k}\mathbf{P}_{s,k}^\top(\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\frac{\boldsymbol{\delta}_s}{\sqrt{p}}||_2^2 \\
&\leq 2\sqrt{n}||\sqrt{\mathbf{D}_{p-s}}\frac{\boldsymbol{\delta}_{p-s}}{\sqrt{p}}||_2^2 + 2\sqrt{n}||\sqrt{\mathbf{D}_{p-s}}\mathbf{P}_{p-s,k}\mathbf{P}_{s,k}^\top(\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\frac{\boldsymbol{\delta}_s}{\sqrt{p}}||_2^2 \\
&= T_1 + T_2.
\end{aligned} \tag{2.39}$$

Next we show that $||\boldsymbol{\beta}||_2^2 T_1$ and $||\boldsymbol{\beta}||_2^2 T_2$ both converge to 0 with probability tending to 1.

In the first step, the concentration inequality for quadratic forms in Lemma 2.12 gives an upper bound on T_1, that is

$$P\left[T_1 \leq \frac{2\sqrt{n}}{p}\left(tr(\mathbf{D}_{p-s}) + 2\sqrt{h_1}||\mathbf{D}_{p-s}||_F + 2h_1||\mathbf{D}_{p-s}||_{sp}\right)\right] \geq 1 - \exp(-h_1),$$

where h_1 is a positive real number that may vary with n. From Assumption A6 and the properties of $||\cdot||_F$ and $||\cdot||_{sp}$, we select $h_1 = n^\gamma$ and get

$$\begin{aligned}||\boldsymbol{\beta}||_2^2 T_1 &\leq \frac{2\sqrt{n}||\boldsymbol{\beta}||_2^2}{p}\left(tr(\mathbf{D}_{p-s}) + 2\sqrt{h_1}||\mathbf{D}_{p-s}||_F + 2h_1||\mathbf{D}_{p-s}||_{sp}\right)\\ &\leq \frac{2\sqrt{n}||\boldsymbol{\beta}||_2^2}{p} tr(\mathbf{D}_{p-s})\left(1 + 2\sqrt{h_1} + 2h_1\right)\\ &\leq \frac{10n^{0.5+\gamma}||\boldsymbol{\beta}||_2^2 tr(\mathbf{D}_{p-s})}{p} = o(1)\end{aligned} \tag{2.40}$$

with probability at least $1 - \exp(-n^\gamma)$.

In the next step, Lemmas 2.14 and 2.13 give upper bounds by

$$k\lambda_{\max}\left((\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\right) = \frac{1}{\lambda_{\min}(\frac{\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top}{k})} \leq \frac{1}{(1-\sqrt{s/k} - k^{-1/4})^2},$$

$$\frac{tr(\mathbf{P}_{p-s,k}^\top\mathbf{D}_{p-s}\mathbf{P}_{p-s,k})}{k} \leq 2tr(\mathbf{D}_{p-s})$$

with probability converging to 1. These inequalities together with Lemma 2.11 lead to

$$\begin{aligned}&tr\left((\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\mathbf{P}_{s,k}\mathbf{P}_{p-s,k}^\top\mathbf{D}_{p-s}\mathbf{P}_{p-s,k}\mathbf{P}_{s,k}^\top(\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\right)\\ &\leq k\lambda_{\max}\left(\mathbf{P}_{s,k}^\top(\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-2}\mathbf{P}_{s,k}\right)\frac{tr(\mathbf{P}_{p-s,k}^\top\mathbf{D}_{p-s}\mathbf{P}_{p-s,k})}{k}\\ &= k\lambda_{\max}\left((\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\right)\frac{tr(\mathbf{P}_{p-s,k}^\top\mathbf{D}_{p-s}\mathbf{P}_{p-s,k})}{k}\\ &\leq \frac{2tr(\mathbf{D}_{p-s})}{(1-\sqrt{s/k}-k^{-1/4})^2}\end{aligned} \tag{2.41}$$

with probability converging to 1. To study the randomness from $\boldsymbol{\delta}_s$, we apply the same method in the first step of investigating $||\boldsymbol{\beta}||_2^2 T_1$ with the help from upper bound in (2.41) and get

$$||\boldsymbol{\beta}||_2^2 T_2 \le \frac{20n^{0.5+\gamma}||\boldsymbol{\beta}||_2^2 tr(\mathbf{D}_{p-s})}{p(1-\sqrt{s/k}-k^{-1/4})^2} = o(1) \tag{2.42}$$

with probability tending to 1.

Combining (2.39), (2.40) and (2.42), we have

$$\min_{\boldsymbol{\eta}\in\mathbb{R}^k} \sqrt{n}||\boldsymbol{\Gamma}^\top\boldsymbol{\beta} - \boldsymbol{\Gamma}^\top\mathbf{P}_k\boldsymbol{\eta}||_2^2 = o(1)$$

with probability tending to 1 and complete the proof. □

2.6.4 Proof of Theorem 2.1

Proof. Under $\mathbf{H}_0$, we have

$$T_n - 1 = \frac{\boldsymbol{\epsilon}^\top\mathbf{M}\boldsymbol{\epsilon}}{\boldsymbol{\epsilon}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\boldsymbol{\epsilon}/(n-1-k)},$$

where $\mathbf{M} = (m_{ij}) = \frac{\mathbf{H}_k}{k} - \frac{\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k}{n-k-1}$. The property that $\mathbf{H}_k$ is idempotent with rank k leads to $tr(\mathbf{M}) = 0$ and $\mathbf{M}^\top\mathbf{M} = \frac{\mathbf{H}_k}{k^2} + \frac{\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k}{(n-k-1)^2}$. Therefore,

$$\frac{||\mathbf{M}||_{sp}^2}{||\mathbf{M}||_F^2} = \frac{\lambda_{\max}(\mathbf{M}^\top\mathbf{M})}{tr(\mathbf{M}^\top\mathbf{M})} \le \frac{\lambda_{\max}(\frac{\mathbf{H}_k}{k^2}) + \lambda_{\max}(\frac{\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k}{(n-k-1)^2})}{\frac{1}{k}+\frac{1}{n-k-1}} = O(n^{-1}).$$

And we have

$$E(\boldsymbol{\epsilon}^\top\mathbf{M}\boldsymbol{\epsilon}|\mathbf{M}) = \sigma^2 tr(\mathbf{M}) = 0,$$

$$Var(\boldsymbol{\epsilon}^\top\mathbf{M}\boldsymbol{\epsilon}|\mathbf{M}) = (\mu_4 - 3\sigma^4)\sum_{i=1}^n m_{ii}^2 + 2\sigma^4(\frac{1}{k} + \frac{1}{n-k-1}),$$

where the error term $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top$ has $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2$ and $E(\epsilon_i^4) = \mu_4$. When $\mathbf{M}$ is given, these together with Lemma 2.4 imply

$$\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sqrt{Var\left(\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon} | \mathbf{M}\right)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

The randomness brought from $\mathbf{M}$ in fact does not influence the asymptotic normality. From the law of total expectation, we have, for $\forall \alpha \in \mathbb{R}$,

$$P\big(\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sqrt{Var(\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon} | \mathbf{M})}} \leq \alpha\big) = E\big(P(\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sqrt{Var(\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon} | \mathbf{M})}} \leq \alpha | \mathbf{M})\big).$$

And the aforementioned result shows

$$P(\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sqrt{Var(\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon} | \mathbf{M})}} \leq \alpha | \mathbf{M}) \to \Phi(\alpha).$$

Based on the dominated convergence theorem, we get

$$\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sigma^2 \sqrt{(\frac{\mu_4}{\sigma^4} - 3) \sum_{i=1}^n m_{ii}^2 + 2(\frac{1}{k} + \frac{1}{n-k-1})}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{2.43}$$

Let $G_n = \sum_{i=1}^n m_{ii}^2$. Next we will show $nG_n = op(1)$. From the definition,

$$\begin{aligned} nG_n = n \sum_{i=1}^n m_{ii}^2 &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{(\mathbf{H}_k)_{ii} - \frac{k}{n-1}(1 - \frac{1}{n})}{\frac{k}{n-1}(1 - \frac{k+1}{n})} \right\}^2 \\ &\leq \frac{2}{n} \sum_{i=1}^n \frac{\left\{(\mathbf{H}_k)_{ii} - \rho\right\}^2 + \left\{\rho - \frac{k}{n-1}(1 - \frac{1}{n})\right\}^2}{\left\{\frac{k}{n-1}(1 - \frac{k+1}{n})\right\}^2}. \end{aligned} \tag{2.44}$$

Let $\boldsymbol{\Sigma}_1 = \mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k$. From Lemma 2.6, we find the smallest eigenvalue of $\frac{1}{p}\mathbf{P}_k^\top \mathbf{P}_k$ is bounded away from 0 a.s., showing $\mathbf{P}_k$ is of full column rank with probability 1. Therefore, $\boldsymbol{\Sigma}_1$ is of full rank with probability 1. Define $\widetilde{\mathbf{U}}_k = \mathbf{X}\mathbf{P}_k \boldsymbol{\Sigma}_1^{-1/2}$. Since $\mathbf{H}_k$ is invariant to the full rank linear transform of $\mathbf{U}_k$, the hat matrix can be expressed as

$$\mathbf{H}_k = \mathbf{U}_k(\mathbf{U}_k^\top \mathbf{U}_k)^{-1}\mathbf{U}_k^\top = (\mathbf{I} - \mathbf{P}_1)\widetilde{\mathbf{U}}_k(\widetilde{\mathbf{U}}_k^\top (\mathbf{I} - \mathbf{P}_1)\widetilde{\mathbf{U}}_k)^{-1}\widetilde{\mathbf{U}}_k^\top (\mathbf{I} - \mathbf{P}_1).$$

From Assumption A1, $\widetilde{\mathbf{U}}_k$ can be denoted by $\mathbf{ZA}$, where $\mathbf{A} = \mathbf{\Gamma}^\top \mathbf{P}_k \mathbf{\Sigma}_1^{-1/2}$ is an $m \times k$ matrix. From Section 2.4.2 in Chikuse (2003), matrix $\mathbf{A}$ is on the Stiefel manifold $\mathcal{V}_k(\mathbb{R}^m)$ with probability 1, which demonstrates $\mathbf{U}_k^\top \mathbf{U}_k$ is of full rank with probability 1. From Lemma 2.9 and (2.44), we obtain $nG_n = op(1)$.

Assumption A3 implies $\frac{n}{k} + \frac{n}{n-k-1} \to \frac{1}{\rho(1-\rho)}$, as $n \to \infty$. Therefore, (2.43) leads to

$$\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sigma^2 \sqrt{2/n\rho(1-\rho)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

In addition, from $E(\frac{\boldsymbol{\epsilon}^\top (\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\boldsymbol{\epsilon}}{n-k-1}) = \sigma^2$, $Var(\frac{\boldsymbol{\epsilon}^\top (\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\boldsymbol{\epsilon}}{n-k-1}) \leq \frac{\mu_4 - \sigma^4}{n-k-1} \to 0$ and Markov's inequality, we have

$$\frac{\boldsymbol{\epsilon}^\top (\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_k)\boldsymbol{\epsilon}}{n-k-1} = \sigma^2 + o_p(1).$$

Hence, under $\mathbf{H}_0$,

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

which completes the proof. □

2.6.5 Proof of Theorem 2.2

Proof. First, we derive a decomposition of $\mathbf{x}_i^\top \boldsymbol{\beta}$. Let $\boldsymbol{\xi} = (\mathbf{P}_k^\top \mathbf{\Sigma} \mathbf{P}_k)^{-1} \mathbf{P}_k^\top \mathbf{\Sigma} \boldsymbol{\beta}$. For each i, define

$$r_i = E(\mathbf{x}_i^\top \boldsymbol{\beta} | \mathbf{P}_k^\top \mathbf{x}_i) - \mathbf{x}_i^\top \mathbf{P}_k \boldsymbol{\xi}, \qquad q_i = \mathbf{x}_i^\top \boldsymbol{\beta} - E(\mathbf{x}_i^\top \boldsymbol{\beta} | \mathbf{P}_k^\top \mathbf{x}_i).$$

Then, we have $\mathbf{x}_i^\top \boldsymbol{\beta} = \mathbf{x}_i^\top \mathbf{P}_k \boldsymbol{\xi} + r_i + q_i$, where q_i satisfies $E(q_i | \mathbf{P}_k^\top \mathbf{x}_i) = 0$. Let $\omega^2 = \boldsymbol{\beta}^\top \mathbf{\Sigma} \boldsymbol{\beta} - \boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{\Sigma} \mathbf{P}_k \boldsymbol{\xi}$ and $\tau_i = Var(q_i | \mathbf{P}_k^\top \mathbf{x}_i) - \omega^2$. According to Lemma 2.10 and the condition $\boldsymbol{\beta}^\top \mathbf{\Sigma} \boldsymbol{\beta} = o(1)$, it shows

$$\sum_{i=1}^n r_i^2 = o_p(1) \text{ and } \frac{1}{\sqrt{n}} \sum_{i=1}^n |\tau_i| = o_p(1), \tag{2.45}$$

when the event $\mathbf{A} \in F_n$ is satisfied, where F_n is a series of sets that satisfy $v_{m,k}(F_n) \to 1$, as $n \to \infty$, and $\mathbf{A} = \mathbf{\Gamma}^\top \mathbf{P}_k (\mathbf{P}_k^\top \mathbf{\Sigma} \mathbf{P}_k)^{-1/2}$. The probability of the event tends to 1, based on the randomness of $\mathbf{P}_k$.

Define a new error term $e_i = q_i + \epsilon_i$. Let $\sigma^2 = Var(\epsilon_i)$. The model can be denoted as

$$\mathbf{y} = \alpha \mathbf{1} + \mathbf{X}\mathbf{P}_k \boldsymbol{\xi} + \mathbf{r} + \mathbf{e}, \tag{2.46}$$

where $\mathbf{r} = (r_1, \ldots, r_n)^\top$, and $\mathbf{e} = (e_1, \ldots, e_n)^\top$ with each element of $\mathbf{e}$ satisfying $E(e_i) = 0$, $E(e_i|\mathbf{P}_k^\top \mathbf{x}_i) = 0$, $Var(e_i|\mathbf{P}_k^\top \mathbf{x}_i) = \sigma^2 + \omega^2 + \tau_i$, and $E(e_i^4|\mathbf{P}_k^\top \mathbf{x}_i) = \mu_4 + 6\sigma^2 Var(q_i|\mathbf{P}_k^\top \mathbf{x}_i) + E(q_i^4|\mathbf{P}_k^\top \mathbf{x}_i)$. For matrix $\mathbf{M} = (m_{ij}) = \frac{\mathbf{H}_k}{k} - \frac{\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_k}{n-k-1}$, calculation shows

$$
\begin{aligned}
E(\mathbf{e}^\top \mathbf{M}\mathbf{e}|\mathbf{X}\mathbf{P}_k) &= \sum_{i=1}^n m_{ii}\tau_i, \\
Var(\mathbf{e}^\top \mathbf{M}\mathbf{e}|\mathbf{X}\mathbf{P}_k) &= \sum_{i=1}^n m_{ii}^2 \left\{ E(e_i^4|\mathbf{X}\mathbf{P}_k) - 3E(e_i^2|\mathbf{X}\mathbf{P}_k)^2 \right\} \\
&\quad + 2\sum_{i,j} m_{ij}^2 E(e_i^2|\mathbf{X}\mathbf{P}_k) E(e_j^2|\mathbf{X}\mathbf{P}_k) \\
&= 2(\sigma^2 + \omega^2)^2 tr(\mathbf{M}^\top \mathbf{M}) + g(\mathbf{M}, \mathbf{X}, \boldsymbol{\epsilon}, \mathbf{P}_k),
\end{aligned}
$$

where $g(\mathbf{M}, \mathbf{X}, \boldsymbol{\epsilon}, \mathbf{P}_k) = \sum_{i=1}^n m_{ii}^2 \left\{ E(e_i^4|\mathbf{X}\mathbf{P}_k) - 3E(e_i^2|\mathbf{X}\mathbf{P}_k)^2 \right\} + 2\sum_{i,j} m_{ij}^2 \{(\sigma^2 + \omega^2)(\tau_i + \tau_j) + \tau_i \tau_j\}$. For a constant $a \le 2/\rho(1-\rho)$ and large n, $\mathbf{M}$ satisfies $||\mathbf{M}||_{sp} \le a/n$ and $|m_{ii}| = |\mathbf{e}_i^\top \mathbf{M} \mathbf{e}_i| \le ||\mathbf{M}||_{sp}$. Then, (2.45) leads to

$$\sqrt{n} E(\mathbf{e}^\top \mathbf{M}\mathbf{e}|\mathbf{X}\mathbf{P}_k) = o_p(1). \tag{2.47}$$

To investigate the conditional variance, based on (2.44) and Lemma 2.9, we can derive

$$\sum_{i=1}^n m_{ii}^2 \left\{ E(e_i^4|\mathbf{X}\mathbf{P}_k) - 3E(e_i^2|\mathbf{X}\mathbf{P}_k)^2 \right\} \le \sum_{i=1}^n m_{ii}^2 \left\{ \mu_4 - 3\sigma^4 + E(q_i^4|\mathbf{X}\mathbf{P}_k) \right\} = o_p(n^{-1}).$$

In addition, $\sum_{j=1}^{n} m_{ij}^2 = \mathbf{e}_i^\top \mathbf{M}\mathbf{M}^\top \mathbf{e}_i \leq ||\mathbf{M}||_{sp}^2 \leq a^2/n^2$ and (2.45) lead to

$$\sum_{i,j} m_{ij}^2 \left\{(\sigma^2+\omega^2)(\tau_i+\tau_j)+\tau_i\tau_j\right\} \leq 2(\sigma^2+\omega^2)a^2\frac{\sum_{i=1}^{n}|\tau_i|}{n^2}+a^2\frac{(\sum_{i=1}^{n}|\tau_i|)^2}{n^2} = o_p(n^{-1}).$$

Therefore, $g(\mathbf{M}, \mathbf{X}, \boldsymbol{\epsilon}, \mathbf{P}_k) = o_p(n^{-1})$, from which we obtain

$$Var(\mathbf{e}^\top \mathbf{M}\mathbf{e}|\mathbf{X}\mathbf{P}_k) = 2(\sigma^2+\omega^2)^2 tr(\mathbf{M}^\top\mathbf{M}) + o_p(n^{-1}). \tag{2.48}$$

According to $tr(\mathbf{M}^\top\mathbf{M}) = \frac{1}{k} + \frac{1}{n-1-k}$, (2.47), (2.48) and the condition $k/n \to \rho$, Lemma 2.4 shows

$$\frac{\sqrt{\frac{n\rho(1-\rho)}{2}}\mathbf{e}^\top\mathbf{M}\mathbf{e} - o_p(1)}{(\sigma^2+\omega^2)\sqrt{1+o_p(1)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{2.49}$$

To investigate the numerator of the test statistic, (2.45) shows that $\mathbf{r}$ satisfies

$$\frac{1}{\sqrt{n}}\mathbf{r}^\top\mathbf{E}\mathbf{r} \leq \frac{1}{\sqrt{n}}\mathbf{r}^\top\mathbf{r} = o_p(n^{-1/2}), \tag{2.50}$$

for any $n \times n$ idempotent matrix $\mathbf{E}$. Based on Jensen's inequality, the fourth moment of q_i satisfies $E(q_i^4) \leq 16E\{(\mathbf{x}_i^\top\boldsymbol{\beta})^4\}$. According to

$$E\{(\mathbf{x}_1^\top\boldsymbol{\beta})^4\} = \sum_{i=1}^{m}(\boldsymbol{\Gamma}^\top\boldsymbol{\beta})_i^4 E(z_{1i}^4) + 3\sum_{i\neq j}^{m}(\boldsymbol{\Gamma}^\top\boldsymbol{\beta})_i^2(\boldsymbol{\Gamma}^\top\boldsymbol{\beta})_j^2 E(z_{1i}^2 z_{1j}^2),$$

and $Var(q_i) \leq \omega^2 \leq \boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta}$, the condition $\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta} = o(1)$ leads to $E(q_i^4) = o(1)$ and

$$|E(e_i^4) - \mu_4| \leq c_1\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta} = o(1), \quad E\{\tau_i^2\} \leq E(q_i^4) + \omega^4 \leq c_1(\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta})^2 = o(1), \tag{2.51}$$

for a constant c_1. In addition, the calculation shows

$$\left|E\left(\frac{\mathbf{e}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\mathbf{e}}{n-1-k}\right) - (\sigma^2+\omega^2)\right| = \left|E\left\{\sum_{i=1}^{n}\left(\frac{\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k}{n-1-k}\right)_{ii}\tau_i\right\}\right|$$

$$\leq \sum_{i=1}^{n}\frac{1}{n-1-k}\sqrt{E\{\tau_i^2\}} = o(1),$$

$$E\left\{Var\left(\frac{\mathbf{e}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\mathbf{e}}{n-1-k}|\mathbf{X}\mathbf{P}_k\right)\right\}$$

$$\leq \frac{\sum_{i=1}^n E\left(e_i^4\right)}{(n-1-k)^2}+\frac{2(\sigma^2+\omega^2)^2}{n-1-k}+\frac{4n(\sigma^2+\omega^2)\sum_{i=1}^n\sqrt{E(\tau_i^2)}+2\sum_{i,j}\sqrt{E(\tau_i^2)E(\tau_j^2)}}{(n-1-k)^2}=o(1),$$

$$\begin{aligned}Var\left\{E\left(\frac{\mathbf{e}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\mathbf{e}}{n-1-k}|\mathbf{X}\mathbf{P}_k\right)\right\} &= Var\left(\sum_{i=1}^n\frac{(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)_{ii}}{n-1-k}\tau_i\right)\\ &\leq E\left\{\left(\sum_{i=1}^n\frac{(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)_{ii}^2}{(n-1-k)^2}\right)\left(\sum_{i=1}^n\tau_i^2\right)\right\}\\ &\leq \frac{n}{(n-1-k)^2}\sum_{i=1}^n E(\tau_i^2)=o(1).\end{aligned}$$

Consequently, Markov's inequality leads to

$$\frac{\mathbf{e}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\mathbf{e}}{n-1-k}=\sigma^2+\omega^2+o_p(1).$$

This combines with (2.50) shows

$$\frac{(\mathbf{e}+\mathbf{r})^\top\big(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k\big)(\mathbf{e}+\mathbf{r})}{n-1-k}=\sigma^2+\omega^2+o_p(1). \tag{2.52}$$

Next, we study $\frac{\sqrt{n}}{k}\boldsymbol{\xi}^\top\mathbf{P}_k^\top\mathbf{X}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{X}\mathbf{P}_k\boldsymbol{\xi}$. From Assumption A1, we have

$$E\left\{\frac{1}{\sqrt{n}}\boldsymbol{\xi}^\top\mathbf{P}_k^\top\mathbf{X}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{X}\mathbf{P}_k\boldsymbol{\xi}\right\}=\frac{n-1}{\sqrt{n}}\boldsymbol{\xi}^\top\mathbf{P}_k^\top\boldsymbol{\Sigma}\mathbf{P}_k\boldsymbol{\xi} \tag{2.53}$$

and the fourth moment of $\mathbf{x}_1^\top\mathbf{P}_k\boldsymbol{\xi}$ satisfies

$$E\big\{(\mathbf{x}_1^\top\mathbf{P}_k\boldsymbol{\xi})^4\big\}=\sum_{i=1}^m(\boldsymbol{\Gamma}^\top\mathbf{P}_k\boldsymbol{\xi})_i^4E(z_{1i}^4)+3\sum_{i\neq j}^m(\boldsymbol{\Gamma}^\top\mathbf{P}_k\boldsymbol{\xi})_i^2(\boldsymbol{\Gamma}^\top\mathbf{P}_k\boldsymbol{\xi})_j^2E(z_{1i}^2z_{1j}^2).$$

Based on $\boldsymbol{\xi}^\top\mathbf{P}_k^\top\boldsymbol{\Sigma}\mathbf{P}_k\boldsymbol{\xi}=\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\mathbf{P}_k(\mathbf{P}_k^\top\boldsymbol{\Sigma}\mathbf{P}_k)^{-1}\mathbf{P}_k^\top\boldsymbol{\Sigma}\boldsymbol{\beta}\leq\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta}=o(1)$, we have

$$Var\left(\frac{1}{\sqrt{n}}\boldsymbol{\xi}^\top\mathbf{P}_k^\top\mathbf{X}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{X}\mathbf{P}_k\boldsymbol{\xi}\right)\leq E\left\{(\mathbf{x}_1^\top\mathbf{P}_k\boldsymbol{\xi})^4\right\}+2(\boldsymbol{\xi}^\top\mathbf{P}_k^\top\boldsymbol{\Sigma}\mathbf{P}_k\boldsymbol{\xi})^2=o(1).$$

From Markov's inequality and $k/n \to \rho$, we have

$$\frac{\sqrt{n}}{k}\boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top (\mathbf{I} - \mathbf{P}_1)\mathbf{X}\mathbf{P}_k\boldsymbol{\xi} = \frac{\sqrt{n}}{\rho}\boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma}\mathbf{P}_k\boldsymbol{\xi} + o_p(1). \tag{2.54}$$

To investigate $\frac{\sqrt{n}}{k}\boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top (\mathbf{I} - \mathbf{P}_1)\mathbf{e}$, the condition $\boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta} = o(1)$, (2.51) and (2.53) lead to

$$\begin{aligned}
&E\left\{\left(\frac{1}{\sqrt{n}}\boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top \big(\mathbf{I} - \mathbf{P}_1\big)\mathbf{e}\right)^2\right\} \\
&= E\left[E\left\{\left(\frac{1}{\sqrt{n}}\boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top \big(\mathbf{I} - \mathbf{P}_1\big)\mathbf{e}\right)^2 \big| \mathbf{X}\mathbf{P}_k\right\}\right] \\
&= E\left\{\frac{1}{n}\sum_{i=1}^n (\sigma^2 + \omega^2 + \tau_i)\big(\mathbf{x}_i^\top \mathbf{P}_k\boldsymbol{\xi} - \frac{1}{n}\sum_{j=1}^n \mathbf{x}_j^\top \mathbf{P}_k\boldsymbol{\xi}\big)^2\right\} \\
&\leq (\sigma^2 + \omega^2)E\left\{\frac{1}{n}\boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top (\mathbf{I} - \mathbf{P}_1)\mathbf{X}\mathbf{P}_k\boldsymbol{\xi}\right\} \\
&\quad + \sqrt{E\left(\frac{1}{n}\sum_{i=1}^n \tau_i^2\right)}\sqrt{E\left\{\frac{1}{n}\sum_{i=1}^n \big(\mathbf{x}_i^\top \mathbf{P}_k\boldsymbol{\xi} - \sum_{j=1}^n \mathbf{x}_j^\top \mathbf{P}_k\boldsymbol{\xi}\big)^4\right\}} \\
&\leq (\sigma^2 + \omega^2)\boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma}\mathbf{P}_k\boldsymbol{\xi} + \sqrt{c_1}\boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta}\sqrt{E\left\{\frac{16}{n}\sum_{i=1}^n \big(\mathbf{x}_i^\top \mathbf{P}_k\boldsymbol{\xi}\big)^4\right\}} \\
&= o(1).
\end{aligned}$$

Therefore, Markov's inequality and $k/n \to \rho$ demonstrate

$$\frac{\sqrt{n}}{k}\boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top (\mathbf{I} - \mathbf{P}_1)\mathbf{e} = o_p(1).$$

This combines with (2.50) and (2.54) implies

$$\frac{\sqrt{n}}{k}\boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top (\mathbf{I} - \mathbf{P}_1)(\mathbf{e} + \mathbf{r}) = o_p(1). \tag{2.55}$$

Based on the new expression (2.46), together with (2.50), (2.52), (2.54) and (2.55), we have

$$\frac{T_n - 1}{\sqrt{2/[n\rho(1-\rho)]}} = \frac{\sqrt{\frac{n\rho(1-\rho)}{2}}\left\{\frac{\boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{X}\mathbf{P}_k\boldsymbol{\xi}}{k} + \frac{2\boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top(\mathbf{I}-\mathbf{P}_1)(\mathbf{e}+\mathbf{r})}{k} + (\mathbf{e}+\mathbf{r})^\top \mathbf{M}(\mathbf{e}+\mathbf{r})\right\}}{\frac{(\mathbf{e}+\mathbf{r})^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)(\mathbf{e}+\mathbf{r})}{n-k-1}}$$

$$= \frac{\sqrt{\frac{n\rho(1-\rho)}{2}}(\frac{1}{\rho}\boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma}\mathbf{P}_k\boldsymbol{\xi} + \mathbf{e}^\top \mathbf{M}\mathbf{e}) + o_p(1)}{\sigma^2 + \omega^2 + o_p(1)}.$$

Define $\delta_k^2 = \sigma^2 + \boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta} - \boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma}\mathbf{P}_k\boldsymbol{\xi}$. From (2.49), the asymptotic power function of the proposed test T_n is

$$\Psi_n^{RP}(\boldsymbol{\beta}; \mathbf{P}_k) = P(\frac{T_n - 1}{\sqrt{2/[n\rho(1-\rho)]}} > z_\alpha)$$

$$= \Phi(-z_\alpha + \sqrt{\frac{n(1-\rho)}{2\rho}}\frac{\boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma}\mathbf{P}_k\boldsymbol{\xi}}{\delta_k^2}) + o(1),$$

which completes the proof. □

2.6.6 Proof of Theorem 2.3

Proof. Recall the definitions of projection matrices.

$$\mathbf{P}_1 = \frac{1}{n}\mathbf{1}\mathbf{1}^\top,$$

$$\mathbf{P}_{\mathbf{X}_1} = (\mathbf{I}-\mathbf{P}_1)\mathbf{X}_1(\mathbf{X}_1^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{X}_1)^{-1}\mathbf{X}_1^\top(\mathbf{I}-\mathbf{P}_1),$$

$$\mathbf{H}_{k_2} = (\mathbf{I}-\mathbf{P}_1)\mathbf{W}(\mathbf{W}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{W})^{-1}\mathbf{W}^\top(\mathbf{I}-\mathbf{P}_1),$$

where $\mathbf{W} = (\mathbf{X}_1, \mathbf{X}_2\mathbf{P}_{k_2})$. Under $\mathbf{H}_{part,0}$, we have

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

Define $\mathbf{M} = (m_{ij}) = \frac{\mathbf{H}_{k_2}-\mathbf{P}_{\mathbf{X}_1}}{k_2} - \frac{\mathbf{I}-\mathbf{P}_1-\mathbf{H}_{k_2}}{n-1-p_1-k_2}$. From $Span\{(\mathbf{I}-\mathbf{P}_1)\mathbf{X}_1\} \subseteq Span\{(\mathbf{I}-\mathbf{P}_1)\mathbf{W}\}$ and properties of projection matrices, we have

$$\mathbf{P}_{\mathbf{X}_1}\mathbf{H}_{k_2} = \mathbf{H}_{k_2}\mathbf{P}_{\mathbf{X}_1} = \mathbf{P}_{\mathbf{X}_1}.$$

Hence, $tr(\mathbf{M}) = 0$, $\mathbf{M}^\top \mathbf{M} = \frac{\mathbf{H}_{k_2} - \mathbf{P}_{\mathbf{X}_1}}{k_2^2} + \frac{\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_{k_2}}{(n-1-p_1-k_2)^2}$, and

$$\frac{||\mathbf{M}||_{sp}^2}{||\mathbf{M}||_F^2} = \frac{\lambda_{\max}(\mathbf{M}^\top \mathbf{M})}{tr(\mathbf{M}^\top \mathbf{M})} \leq \frac{\lambda_{\max}(\frac{\mathbf{H}_{k_2} - \mathbf{P}_{\mathbf{X}_1}}{k_2^2}) + \lambda_{\max}(\frac{\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_{k_2}}{(n-1-p_1-k_2)^2})}{\frac{1}{k_2} + \frac{1}{n-1-p_1-k_2}} = O(n^{-1}).$$

For given $\mathbf{M}$, we have

$$E(\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon} | \mathbf{M}) = \sigma^2 tr(\mathbf{M}) = 0,$$

$$Var(\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon} | \mathbf{M}) = (\mu_4 - 3\sigma^4) \sum_{i=1}^n m_{ii}^2 + 2\sigma^4 \left(\frac{1}{k_2} + \frac{1}{n-1-p_1-k_2} \right).$$

Then, Lemma 2.4 leads to

$$\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sqrt{Var(\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon} | \mathbf{M})}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

This together with the law of total expectation and the dominated convergence theorem shows

$$P\left(\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sqrt{Var(\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon} | \mathbf{M})}} \leq \alpha \right) = E\left[P\left(\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sqrt{Var(\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon} | \mathbf{M})}} \leq \alpha | \mathbf{M} \right) \right] \rightarrow \Phi(\alpha),$$

for $\forall \alpha \in \mathbb{R}$. Therefore,

$$\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sigma^2 \sqrt{[E\{(\frac{\epsilon_1}{\sigma})^4\} - 3] \sum_{i=1}^n m_{ii}^2 + 2(\frac{1}{k_2} + \frac{1}{n-1-p_1-k_2})}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

When $n \sum_{i=1}^n m_{ii}^2 = o_p(1)$, Assumption S3 and Slutsky's lemma demonstrate

$$\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sigma^2 \sqrt{2(1-\rho_1)/n\rho_2(1-\rho_1-\rho_2)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \tag{2.56}$$

Let $G_n = \sum_{i=1}^n m_{ii}^2$. Next, we will verify $nG_n = o_p(1)$. From the definition, $m_{ii} =$

$\frac{(\mathbf{H}_{k_2})_{ii}-(\mathbf{P}_{\mathbf{X}_1})_{ii}}{k_2} - \frac{1-\frac{1}{n}-(\mathbf{H}_{k_2})_{ii}}{n-1-p_1-k_2}$. Then

$$nG_n = n\sum_{i=1}^{n} m_{ii}^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{(1-\frac{1}{n}-\frac{p_1}{n})\big((\mathbf{H}_{k_2})_{ii}-\frac{p_1+k_2}{n}\big)}{\frac{k_2}{n}(1-\frac{1}{n}-\frac{p_1}{n}-\frac{k_2}{n})} - \frac{(\mathbf{P}_{\mathbf{X}_1})_{ii}-\frac{p_1}{n}}{\frac{k_2}{n}}\right\}^2$$

$$\leq \frac{2h_1}{n}\sum_{i=1}^{n}\left\{(\mathbf{H}_{k_2})_{ii}-\frac{p_1+k_2}{n}\right\}^2 + \frac{2h_2}{n}\sum_{i=1}^{n}\left\{(\mathbf{P}_{\mathbf{X}_1})_{ii}-\frac{p_1}{n}\right\}^2 \tag{2.57}$$

where $h_1 = (1-\frac{1}{n}-\frac{p_1}{n})^2/(\frac{k_2}{n}(1-\frac{1}{n}-\frac{p_1}{n}-\frac{k_2}{n}))^2$ and $h_2 = n^2/k_2^2$. Based on Assumption S3, as $n\to\infty$,

$$h_1 \to \frac{(1-\rho_1)^2}{\rho_2^2(1-\rho_1-\rho_2)^2},\ h_2 \to \frac{1}{\rho_2^2}.$$

Consequently, we only need to consider the sum parts in (2.57). From the definition,

$$\mathbf{W} = (\mathbf{X}_1, \mathbf{X}_2\mathbf{P}_{k_2}) = \mathbf{Z}\boldsymbol{\Gamma}^\top \begin{pmatrix} \mathbf{I}_{p_1} & 0 \\ 0 & \mathbf{P}_{k_2} \end{pmatrix} \triangleq \mathbf{Z}\boldsymbol{\Gamma}^\top\mathbf{V},$$

where $\mathbf{Z} = (\mathbf{z}_1,\ldots,\mathbf{z}_n)^\top$ and $\mathbf{V}$ is a full column rank matrix with probability 1. Define $\boldsymbol{\Sigma}_2 = \mathbf{V}^\top\boldsymbol{\Sigma}\mathbf{V}$. The matrix $\boldsymbol{\Sigma}_2$ is of full rank with probability 1, then $\boldsymbol{\Gamma}^\top\mathbf{V}\boldsymbol{\Sigma}_2^{-1/2}$ is well-defined on the Stiefel manifold $\mathcal{V}_{p_1+k_2}(\mathbb{R}^m)$. Let $\mathbf{W}_1 = \mathbf{W}\boldsymbol{\Sigma}_2^{-1/2} = \mathbf{Z}\boldsymbol{\Gamma}^\top\mathbf{V}\boldsymbol{\Sigma}_2^{-1/2}$. The hat matrix $\mathbf{H}_{k_2}$ can be denoted as

$$\mathbf{H}_{k_2} = (\mathbf{I}-\mathbf{P}_1)\mathbf{W}_1(\mathbf{W}_1^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{W}_1)^{-1}\mathbf{W}_1^\top(\mathbf{I}-\mathbf{P}_1).$$

According to Lemma 2.9 and the condition $(p_1+k_2)/n \to \rho_1+\rho_2$, we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\left\{(\mathbf{H}_{k_2})_{ii}-\frac{p_1+k_2}{n}\right\}^2 = o_p(1).$$

Let $\mathbf{R}_1 = \mathbf{Z}\boldsymbol{\Gamma}_1\boldsymbol{\Sigma}_{11}^{-1/2}$. The hat matrix $\mathbf{P}_{\mathbf{X}_1}$ can be denoted as

$$\mathbf{P}_{\mathbf{X}_1} = (\mathbf{I}-\mathbf{P}_1)\mathbf{R}_1(\mathbf{R}_1^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{R}_1)^{-1}\mathbf{R}_1^\top(\mathbf{I}-\mathbf{P}_1).$$

Based on Lemma 2.9 and the condition $p_1/n \to \rho_1$, we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\left\{(\mathbf{P}_{\mathbf{X}_1})_{ii} - \frac{p_1}{n}\right\}^2 = o_p(1).$$

Therefore, $nG_n = o_p(1)$ is verified, and then (2.56) is demonstrated.

To study the denominator of T_{n,p_2}, calculation shows

$$E\left\{\frac{\boldsymbol{\epsilon}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_{k_2})\boldsymbol{\epsilon}}{n-1-p_1-k_2}\right\} = E\left[E\left\{\frac{\boldsymbol{\epsilon}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_{k_2})\boldsymbol{\epsilon}}{n-1-p_1-k_2}\big|\mathbf{H}_{k_2}\right\}\right] = \sigma^2,$$

$$Var\left(\frac{\boldsymbol{\epsilon}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_{k_2})\boldsymbol{\epsilon}}{n-1-p_1-k_2}\right) = E\left\{Var\left(\frac{\boldsymbol{\epsilon}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_{k_2})\boldsymbol{\epsilon}}{n-1-p_1-k_2}\big|\mathbf{H}_{k_2}\right)\right\} = o(1).$$

From Markov's inequality, we have

$$\frac{\boldsymbol{\epsilon}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_{k_2})\boldsymbol{\epsilon}}{n-1-p_1-k_2} = \sigma^2 + o_p(1).$$

Combining this with (2.56), we obtain

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

which completes the proof. □

2.6.7 Proof of Theorem 2.4

Proof. Define $\mathbf{V} = \mathrm{diag}(\mathbf{I}_{p_1}, \mathbf{P}_{k_2})$. The matrix is a full column rank matrix with probability 1, and $\mathbf{W} = \mathbf{X}\mathbf{V}$, with the i-th row $\mathbf{w}_i = \mathbf{V}^\top\mathbf{x}_i$. Let $\boldsymbol{\gamma} = (\mathbf{V}^\top\boldsymbol{\Sigma}\mathbf{V})^{-1}\mathbf{V}^\top\boldsymbol{\Gamma}\boldsymbol{\Gamma}_2^\top\boldsymbol{\beta}_2$. For each i, define

$$r_i = E\left(\mathbf{x}_{2i}^\top\boldsymbol{\beta}_2\big|\mathbf{V}^\top\mathbf{x}_i\right) - \mathbf{x}_i^\top\mathbf{V}\boldsymbol{\gamma}, \quad q_i = \mathbf{x}_{2i}^\top\boldsymbol{\beta}_2 - E\left(\mathbf{x}_{2i}^\top\boldsymbol{\beta}_2\big|\mathbf{V}^\top\mathbf{x}_i\right).$$

Then, a decomposition of $\mathbf{x}_{2i}^\top\boldsymbol{\beta}_2$ can be derived, given as $\mathbf{x}_{2i}^\top\boldsymbol{\beta}_2 = \mathbf{w}_i^\top\boldsymbol{\gamma} + r_i + q_i$. Let $\omega^2 = \boldsymbol{\beta}_2^\top\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_2 - \boldsymbol{\gamma}^\top\mathbf{V}^\top\boldsymbol{\Sigma}\mathbf{V}\boldsymbol{\gamma}$ and $\tau_i = Var(q_i|\mathbf{V}^\top\mathbf{x}_i) - \omega^2$. According to Lemma

2.10 and the condition $\boldsymbol{\beta}_2^\top \boldsymbol{\Sigma}_{22} \boldsymbol{\beta}_2 = o(1)$, we have

$$\sum_{i=1}^{n} r_i^2 = o_p(1) \text{ and } \frac{1}{\sqrt{n}} |\tau_i| = o_p(1), \tag{2.58}$$

when the event $\mathbf{A} \in F_n$ is satisfied, where $\mathbf{A} = \boldsymbol{\Gamma}^\top \mathbf{V}(\mathbf{V}^\top \boldsymbol{\Sigma} \mathbf{V})^{-1/2}$ and F_n is a series of sets that satisfy $\nu_{m,(p_1+k_2)}(F_n) \to 1$, as $n \to \infty$. The probability of the event tends to 1, based on the randomness of $\mathbf{P}_{k_2}$.

Define a new error term $e_i = q_i + \epsilon_i$. Let σ^2 denote the variance of ϵ_i. The model can be expressed as

$$\mathbf{y} = \alpha \mathbf{1} + \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{W} \boldsymbol{\gamma} + \mathbf{r} + \mathbf{e}, \tag{2.59}$$

where $\mathbf{r} = (r_1, \ldots, r_n)^\top$, and $\mathbf{e} = (e_1, \ldots, e_n)^\top$ with each element of $\mathbf{e}$ satisfying $E(e_i) = 0$, $E(e_i|\mathbf{V}^\top \mathbf{x}_i) = 0$, $Var(e_i|\mathbf{V}^\top \mathbf{x}_i) = \sigma^2 + \omega^2 + \tau_i$, and $E(e_i^4|\mathbf{V}^\top \mathbf{x}_i) = \mu_4 + 6\sigma^2 Var(q_i|\mathbf{V}^\top \mathbf{x}_i) + E(q_i^4|\mathbf{V}^\top \mathbf{x}_i)$. Define $\mathbf{M} = \frac{\mathbf{H}_{k_2} - \mathbf{P}_{\mathbf{X}_1}}{k_2} - \frac{\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_{k_2}}{n-1-p_1-k_2}$. The matrix satisfies $tr(\mathbf{M}) = 0$, $tr(\mathbf{M}\mathbf{M}^\top) = \frac{1}{k_2} + \frac{1}{n-1-p_1-k_2}$, and $||\mathbf{M}||_{sp}^2 \leq \frac{1}{k_2^2} + \frac{1}{(n-1-p_1-k_2)^2}$. Based on the condition $p_1/n \to \rho_1$ and $k_2/n \to \rho_2$, then for large n, there is a constant $a \leq 2/\rho_2(1-\rho_1-\rho_2)$ such that $||\mathbf{M}||_{sp} \leq a/n$. With a similar proof method in Section 2.6.5, we can derive

$$\frac{\sqrt{\frac{n\rho_2(1-\rho_1-\rho_2)}{2(1-\rho_1)}} \mathbf{e}^\top \mathbf{M} \mathbf{e} - o_p(1)}{(\sigma^2 + \omega^2)\sqrt{1 + o_p(1)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \tag{2.60}$$

The condition $\boldsymbol{\beta}_2^\top \boldsymbol{\Sigma}_{22} \boldsymbol{\beta}_2 = o(1)$ leads to $E(q_i^4) = o(1)$ as well as

$$|E(e_i^4) - \mu_4| \leq c_1 \boldsymbol{\beta}_2^\top \boldsymbol{\Sigma}_{22} \boldsymbol{\beta}_2 = o(1), \quad E\{\tau_i^2\} \leq E(q_i^4) + \omega^4 \leq c_1 (\boldsymbol{\beta}_2^\top \boldsymbol{\Sigma}_{22} \boldsymbol{\beta}_2)^2 = o(1), \tag{2.61}$$

for a constant c_1, from which we could obtain

$$\frac{(\mathbf{e} + \mathbf{r})^\top (\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_{k_2})(\mathbf{r} + \mathbf{e})}{n - 1 - p_1 - k_2} = \sigma^2 + \omega^2 + o_p(1). \tag{2.62}$$

Let $\mathbf{V}\boldsymbol{\gamma} = (\boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top)^\top$ with $\boldsymbol{\xi}_1 \in \mathbb{R}^{p_1}$ and $\boldsymbol{\xi}_2 \in \mathbb{R}^{p_2}$. Define $\nu^2 = \boldsymbol{\xi}_2^\top(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})\boldsymbol{\xi}_2$. Then

$$\nu^2 = \boldsymbol{\beta}_2^\top\boldsymbol{\Gamma}_2\left(\boldsymbol{\Gamma}^\top\mathbf{V}(\mathbf{V}^\top\boldsymbol{\Sigma}\mathbf{V})^{-1}\mathbf{V}^\top\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_1^\top\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Gamma}_1\right)\boldsymbol{\Gamma}_2^\top\boldsymbol{\beta}_2 \le \boldsymbol{\beta}_2^\top(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})\boldsymbol{\beta}_2 = o(1).$$

To investigate $\boldsymbol{\gamma}^\top\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{P}_{\mathbf{X}_1})\mathbf{W}\boldsymbol{\gamma}$, the term could be denoted as

$$\begin{aligned}&\boldsymbol{\gamma}^\top\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{P}_{\mathbf{X}_1})\mathbf{W}\boldsymbol{\gamma}\\ &= \boldsymbol{\phi}^\top\mathbf{Z}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{Z}\boldsymbol{\phi} - \boldsymbol{\phi}^\top\mathbf{Z}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{Z}\boldsymbol{\Gamma}_1^\top\left(\boldsymbol{\Gamma}_1\mathbf{Z}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{Z}\boldsymbol{\Gamma}_1^\top\right)^{-1}\boldsymbol{\Gamma}_1\mathbf{Z}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{Z}\boldsymbol{\phi}\end{aligned}$$

where $\boldsymbol{\phi} = (\mathbf{I} - \boldsymbol{\Gamma}_1^\top\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Gamma}_1)\boldsymbol{\Gamma}_2^\top\boldsymbol{\xi}_2$ and $\boldsymbol{\phi}^\top\boldsymbol{\phi} = \nu^2 = o(1)$. From the calculation

$$E\left\{\frac{1}{\sqrt{n}}\boldsymbol{\phi}^\top\mathbf{Z}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{Z}\boldsymbol{\phi}\right\} = \frac{n-1}{\sqrt{n}}\nu^2,$$

$$Var\left\{\frac{1}{\sqrt{n}}\boldsymbol{\phi}^\top\mathbf{Z}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{Z}\boldsymbol{\phi}\right\} \le 6\nu^4 = o(1)$$

Markov's inequality implies $\frac{1}{\sqrt{n}}\boldsymbol{\phi}^\top\mathbf{Z}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{Z}\boldsymbol{\phi} = \sqrt{n}\nu^2 + o_p(1)$. From a similar derivation method for (2.62), we obtain

$$\frac{1}{\sqrt{n}}\boldsymbol{\phi}^\top\mathbf{Z}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{Z}\boldsymbol{\Gamma}_1^\top\left(\boldsymbol{\Gamma}_1\mathbf{Z}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{Z}\boldsymbol{\Gamma}_1^\top\right)^{-1}\boldsymbol{\Gamma}_1\mathbf{Z}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{Z}\boldsymbol{\phi} = \frac{p_1}{\sqrt{n}}\nu^2 + o_p(1).$$

Therefore,

$$\frac{1}{\sqrt{n}}\boldsymbol{\gamma}^\top\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{P}_{\mathbf{X}_1})\mathbf{W}\boldsymbol{\gamma} = \frac{n - p_1}{\sqrt{n}}\nu^2 + o_p(1). \tag{2.63}$$

To study $\boldsymbol{\gamma}^\top\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{P}_{\mathbf{X}_1})(\mathbf{e} + \mathbf{r})$, (2.58) and (2.63) lead to

$$\left|\frac{1}{\sqrt{n}}\boldsymbol{\gamma}^\top\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{P}_{\mathbf{X}_1})\mathbf{r}\right| \le \sqrt{\mathbf{r}^\top\mathbf{r}} \cdot \sqrt{\frac{1}{n}\boldsymbol{\gamma}^\top\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{P}_{\mathbf{X}_1})\mathbf{W}\boldsymbol{\gamma}} = o_p(1).$$

The condition $\boldsymbol{\beta}_2^\top\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_2 = o(1)$, (2.61) and (2.63) lead to

$$\begin{aligned}E\left\{\left(\frac{1}{\sqrt{n}}\boldsymbol{\gamma}^\top\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{P}_{\mathbf{X}_1})\mathbf{e}\right)^2\right\} &= E\left[E\left\{\left(\frac{1}{\sqrt{n}}\boldsymbol{\gamma}^\top\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{P}_{\mathbf{X}_1})\mathbf{e}\right)^2 |\mathbf{W}\right\}\right]\\ &\le (c_3 + \sigma^2 + \omega^2)\boldsymbol{\beta}_2^\top\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_2\\ &= o(1),\end{aligned}$$

where c_3 is a constant. Therefore, we obtain

$$\frac{1}{\sqrt{n}}\boldsymbol{\gamma}^\top \mathbf{W}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{P}_{\mathbf{X}_1})(\mathbf{e}+\mathbf{r}) = o_p(1). \tag{2.64}$$

From the new expression (2.59), together with (2.58), (2.62), (2.63) and (2.64), we have

$$\begin{aligned}\frac{T_{n,p_2}-1}{\sqrt{2(1-\rho_1)/n\rho_2(1-\rho_1-\rho_2)}} &= \frac{\sqrt{\frac{n\rho_2(1-\rho_1-\rho_2)}{2(1-\rho_1)}}\left\{\frac{\boldsymbol{\gamma}^\top\mathbf{W}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{P}_{\mathbf{X}_1})(\mathbf{W}\boldsymbol{\gamma}+2\mathbf{e}+2\mathbf{r})}{k_2}+(\mathbf{r}+\mathbf{e})^\top\mathbf{M}(\mathbf{r}+\mathbf{e})\right\}}{\frac{(\mathbf{r}+\mathbf{e})^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_{k_2})(\mathbf{r}+\mathbf{e})}{n-1-p_1-k_2}} \\ &= \frac{\sqrt{\frac{n\rho_2(1-\rho_1-\rho_2)}{2(1-\rho_1)}}\left\{\frac{(1-\rho_1)}{\rho_2}\nu^2+\mathbf{e}^\top\mathbf{M}\mathbf{e}\right\}+o_p(1)}{\sigma^2+\omega^2+o_p(1)}.\end{aligned}$$

Define $\tau_k^2 = \sigma^2+\omega^2$. Then, ν^2 and τ_k^2 can also be calculated as follows. Let $\widetilde{\boldsymbol{\gamma}} = (\mathbf{V}^\top\boldsymbol{\Sigma}\mathbf{V})^{-1}\mathbf{V}^\top\boldsymbol{\Sigma}\boldsymbol{\beta}$ and $\mathbf{V}\widetilde{\boldsymbol{\gamma}} = (\widetilde{\boldsymbol{\xi}}_1^\top, \widetilde{\boldsymbol{\xi}}_2^\top)^\top$, where $\widetilde{\boldsymbol{\xi}}_1 \in \mathbb{R}^{p_1}$ and $\widetilde{\boldsymbol{\xi}}_2 \in \mathbb{R}^{p_2}$. Then,

$$\begin{aligned}\nu^2 &= \boldsymbol{\beta}_2^\top\boldsymbol{\Gamma}_2\left(\boldsymbol{\Gamma}^\top\mathbf{V}(\mathbf{V}^\top\boldsymbol{\Sigma}\mathbf{V})^{-1}\mathbf{V}^\top\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_1^\top\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Gamma}_1\right)\boldsymbol{\Gamma}_2^\top\boldsymbol{\beta}_2 \\ &= \boldsymbol{\beta}^\top\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^\top\mathbf{V}(\mathbf{V}^\top\boldsymbol{\Sigma}\mathbf{V})^{-1}\mathbf{V}^\top\boldsymbol{\Gamma}-\boldsymbol{\Gamma}_1^\top\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Gamma}_1\right)\boldsymbol{\Gamma}^\top\boldsymbol{\beta} \\ &= \widetilde{\boldsymbol{\xi}}_2^\top(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})\widetilde{\boldsymbol{\xi}}_2.\end{aligned}$$

and $\tau_k^2 = \sigma^2+\boldsymbol{\beta}_2^\top\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_2-\boldsymbol{\gamma}^\top\mathbf{V}^\top\boldsymbol{\Sigma}\mathbf{V}\boldsymbol{\gamma} = \sigma^2+\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta}-\widetilde{\boldsymbol{\gamma}}^\top\mathbf{V}^\top\boldsymbol{\Sigma}\mathbf{V}\widetilde{\boldsymbol{\gamma}}$.

From (2.60), the asymptotic power function of the proposed test T_{n,p_2} is

$$\begin{aligned}\Psi_{n,p_2}^{RP}(\boldsymbol{\beta}_2;\mathbf{P}_{k_2}) &= P\left(\frac{T_{n,p_2}-1}{\sqrt{2(1-\rho_1)/n\rho_2(1-\rho_1-\rho_2)}} > z_\alpha\right) \\ &= \Phi\left(-z_\alpha+\sqrt{\frac{n(1-\rho_1-\rho_2)(1-\rho_1)}{2\rho_2}}\frac{\nu^2}{\tau_k^2}\right)+o(1),\end{aligned}$$

which completes the proof. □

Chapter 3

A Random Projection Approach to Hypothesis Tests in High-Dimensional Single-Index Models

3.1 Introduction

In this chapter, we consider the single-index model (SIM)

$$y = f(\mathbf{x}^\top \boldsymbol{\theta}, \epsilon), \tag{3.1}$$

where y is a response variable, $\mathbf{x}$ is a $p \times 1$ covariate vector, $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown coefficients, ϵ is a random error independent of $\mathbf{x}$, and f is an unspecified link function. We are interested in testing the hypothesis

$$\mathbf{H}_0 : \boldsymbol{\theta} = \mathbf{0} \quad \text{versus} \quad \mathbf{H}_1 : \boldsymbol{\theta} \neq \mathbf{0}. \tag{3.2}$$

We propose new statistical tests for hypothesis (3.2) in high-dimensional SIM (3.1). First, the testing problem is investigated in a relatively high-dimensional regime, where $p/n \to \zeta$ with $\zeta \in (0, 1)$ and $\mathbf{x}$ is assumed to follow the normal distribution. According to the results in Li and Duan (1989), the vector of regression coefficients in SIM (3.1) can be obtained by the least square up to a scalar. Motivated by this, we propose the F-statistic, whose asymptotic normality and asymptotic local power function are derived in high-dimensional SIM. While this approach seems to ignore the nonlinear link function f for simplification, the theoretical results are asymptotically the same as one working on the linear regression model $y = c_0\mathbf{x}^\top\boldsymbol{\theta} + \sigma e$, where e is the standard normal error, and c_0, σ are constants depending on f. To investigate more high-dimensional settings, we study testing problem (3.2) in an ultrahigh-dimensional regime, where p can be much greater than n. The technique of random projection is used to reduce the data dimension, and the F-statistics is constructed based on the projected data which live in a lower-dimensional space. We prove that the proposed test statistic is asymptotically normal under the null hypothesis as $(n, p) \to \infty$. We also derive the asymptotic local power function of the proposed test. With no extra sparsity assumption required, our proposed test has a wide application range in terms of general model assumption and mild conditions on the distribution. And the test is still applicable to misspecified models. In addition, it is simple in form and easy to compute. Finally, we extend the proposed testing procedures for global hypothesis (3.2) to the problem of testing partial coefficients and derive their asymptotic null distributions and asymptotic local power functions.

The rest of this chapter is organized as follows. In Section 3.2, we introduce the model and establish the theoretical foundation for the design of our proposed methods. In Section 3.3, we focus on a relatively high-dimensional regime, where we consider the problem of testing global and partial regression coefficients, and derive the asymptotic normality and the asymptotic local power function of the F-test. In Section 3.4, for an ultrahigh-dimensional regime, we establish the asymptotic null distribution of the random-projection-based test statistics and derive the asymptotic local power functions. In particular, we compare the proposed test with other competing tests in Section 3.4.3. In Section 3.5.1, we conduct simulations to evaluate the finite-sample behaviors of the proposed test in terms of type I error and empirical power, and compare it with the competing tests. We also illustrate its applications in high-dimensional gene expression data in Section 3.5.2. The proofs of technical lemmas and theorems are relegated to Section 3.6.

3.2 Preliminaries

In this section, we introduce the target model and establish the theoretical foundation for the design of our proposed methods, which can be applied as the theoretical premise for our new tests both in relatively and extremely high-dimensional regimes.

For SIM in (3.1), the response variable y is generated from $\mathbf{x}$ based on a linear combination $\mathbf{x}^\top\boldsymbol{\theta}$, while the conditional distribution of y given $\mathbf{x}$ can be completely arbitrary, which includes a wide range of models, such as the GLM and the Cox model. When the information obtained from the data is insufficient, the high flexibility makes the above model an reasonable choice. However, with the introduction of general model assumption, the difficulty of statistical inference increases rapidly. In this paper, we focus on the hypothesis testing problem (3.2) in SIM.

To motivate the proposed test, we first consider a risk function for the estimation problem,

$$R(\alpha, \mathbf{b}) = E\left(L(\alpha + \mathbf{u}^\top\mathbf{b}, y)\right), \tag{3.3}$$

where $L(\mu, y)$ is a loss function, and $\mathbf{u}$ is a k-dimensional random variable with $k \leq \min\{n, p\}$. When $\mathbf{u} = \mathbf{x}$, this criterion is often used for estimating SIM in the classical settings, where the estimator of $\boldsymbol{\theta}$ can be obatined by solving a minimization problem of an empirical version of (3.3). For other types of $\mathbf{u}$, specifically, when $\mathbf{u}$ has a certain relationship with $\mathbf{x}$, Lemma 3.1 indicates the possibility to develop a testing procedure based on $\mathbf{u}$ for (3.2) in the $p > n$ settings. Therefore, the selection of $\mathbf{u}$ is an important ingredient in our proposed method and will be concretely discussed in Sections 3.3 and 3.4.

The following lemma is obtained from the proof in Li and Duan (1989, Theorem 2.2).

Lemma 3.1. *Assume the observation* $(\boldsymbol{x}, y)$ *is generated from SIM (3.1) with* $E(\boldsymbol{x}) = \boldsymbol{0}$ *and* $\boldsymbol{x}^\top \boldsymbol{\theta}$ *being normally distributed. Let* $\boldsymbol{u} \in \mathbb{R}^k$ *be a random vector independent of* ϵ *with* $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ *and* $k \leq \min\{n, p\}$. *Denote* $\boldsymbol{\eta} = E(\boldsymbol{u}\boldsymbol{x}^\top \boldsymbol{\theta})$. *Suppose there is an unique solution* $(\alpha^\star, \boldsymbol{b}^\star)$ *to the minimization problem* $\min_{\alpha \in \mathbb{R}, \boldsymbol{b} \in \mathbb{R}^k} R(\alpha, \boldsymbol{b})$, *when* $\boldsymbol{\eta} \neq \boldsymbol{0}$, *we have*

$$\boldsymbol{b}^\star = c^\star \boldsymbol{\eta},$$

where the scalar $c^\star = \boldsymbol{\eta}^\top \boldsymbol{b}^\star / ||\boldsymbol{\eta}||_2^2$.

The lemma indicates that the direction of $\boldsymbol{\eta}$ can be estimated correctly even if the model is misspecified, which is of great significance in practice, since the underlying true model is unknown in prior for most of the cases.

In this paper, the particular loss function $L(\mu, y) = (y - \mu)^2$ is analyzed. When $p < n$ and $\mathbf{u} = \mathbf{x}$, the minimization problem of (3.3) becomes the least square problem, for which the solution is unique and has a closed form. From Lemma 3.1, the least square solution $\mathbf{b}^{ls}$ is proportional to $\boldsymbol{\theta}$ with a scalar, expressed as

$$\mathbf{b}^{ls} = c_0 \boldsymbol{\theta}, \quad c_0 = E(\boldsymbol{\theta}^\top \mathbf{x} y)/||\boldsymbol{\theta}||_2^2. \tag{3.4}$$

In fact, c_0 might perform like a linear approximation coefficient for f, which can be explained from the special case below. When the response y and the covariate $\mathbf{x}$ satisfies $E(y|\mathbf{x}) = g(\mathbf{x}^\top \boldsymbol{\theta})$ for a differentiable function g, it can be found $c_0 = E\{g'(\mathbf{x}^\top \boldsymbol{\theta})\}$, which is obtained from the normality assumption and Stein's lemma. From the property of the least squares and (3.4), model (3.1) can be written in a linear form

$$y = E(y) + c_0 \mathbf{x}^\top \boldsymbol{\theta} + e, \tag{3.5}$$

where the residual e satisfies $E(e) = 0$ and $E(\mathbf{x}e) = \mathbf{0}$. The scalar coefficient c_0 is set to be 0 when $\boldsymbol{\theta} = \mathbf{0}$. It is noted that the linear coefficients in the linear form are $c_0\boldsymbol{\theta}$, which brings out implicit information of $\boldsymbol{\theta}$ contained in the unknown link function f. When $c_0 \neq 0$. This makes it possible for us to do statistical inference about $\boldsymbol{\theta}$ without estimation of the link function. Motiviated by this, we study the F-statistic in SIM. And its feasibility is investigated in Section 3.3.

3.3 The F-Test in Relatively High Dimensions

In this section, we focus on a relatively high-dimensional regime where p and n satisfy $p/n \to \zeta \in (0, 1)$. For the problems of testing global and partial coefficients, the F-statistic is investigated in SIM. In this setting, we derive the asymptotic null distribution and the asymptotic local power functions of the F-statistic. For simplicity, we assume $E(\mathbf{x}) = \mathbf{0}$ and $E(y) = 0$.

Suppose that $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ are i.i.d. copies of $(\mathbf{x}, y)$ from SIM (3.1). Let $\mathbf{x}_i$ be the i-th 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 F-statistic is defined as

$$F_n = \frac{\mathbf{y}^\top \mathbf{H} \mathbf{y}/p}{\mathbf{y}^\top (\mathbf{I} - \mathbf{H})\mathbf{y}/(n-p)}, \tag{3.6}$$

where the hat matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$. The test statistic F_n is well-defined, since the matrix $\mathbf{X}^\top\mathbf{X}$ is invertible with the probability 1 when $p < n$ and $\mathbf{x}$ follows the normal distribution.

3.3.1 Asymptotic Normality

First, the F-statistic is studied under $\mathbf{H}_0$. We make the following assumptions.

Assumption L1. $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ *and* $\boldsymbol{x}$ *is independent of* ϵ.

Assumption L2. $E(y) = 0$ *and* $E(y^4) < \infty$.

Assumption L3. *There is a constant* $\zeta \in (0,1)$ *such that* $\frac{p}{n} \to \zeta$.

The asymptotic normality of the test statistic F_n is established in the following theorem.

Theorem 3.1. *Under* $\boldsymbol{H}_0$ *and Assumptions L1–L3, as* $n \to \infty$*, we have*

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

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

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

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

3.3.2 Asymptotic Local Power Function

We study the property of F_n under $\mathbf{H}_1$ to derive the asymptotic local power function. We need the following additional assumption.

Assumption L4. $c_0^2||\boldsymbol{\theta}||_2^2 = o(1)$.

This is known as a local alternative, which is commonly used to study the asymptotic properties of a statistical test. Detailed discussions can be found in Van der Vaart (1998, Section 14.1). To derive the asymptotic local power function, our analysis method is based on the linear form (3.5) and the analysis method of the F-test in the linear model. Because the residual e in (3.5) does not satisfy the conditions in a linear model, where the residual is often assumed to be conditionally independent of $\mathbf{x}$. The method is further modified to adapt to our nonlinear high-dimensional settings. The normality assumption of $\mathbf{x}$ is of great significance. It allows us to establish the independence between $\mathbf{e}$ and a new hat matrix, which is derived from a decomposition of the hat matrix $\mathbf{H}$. The scalar $c_0 = E(\boldsymbol{\theta}^\top \mathbf{x} y)/||\boldsymbol{\theta}||_2^2$. The following theorem gives the asymptotic local power function.

Theorem 3.2. *Suppose Assumptions L1–L4 hold. Let $\Psi_n(\boldsymbol{\theta})$ denote the power function of the test statistic F_n. As $n \to \infty$, we have*

$$\Psi_n(\boldsymbol{\theta}) - \Phi(-z_\alpha + \sqrt{\frac{n(1-\zeta)}{2\zeta}} \frac{c_0^2||\boldsymbol{\theta}||_2^2}{\sigma^2}) \to 0,$$

where $\sigma^2 = Var(y) - c_0^2||\boldsymbol{\theta}||_2^2$, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and z_α is the upper α-quantile of Φ.

This result shows that the power of the test becomes stronger as $c_0^2||\boldsymbol{\theta}||_2^2$ increases or ζ decreases. As shown in the linear form (3.5), $c_0^2||\boldsymbol{\theta}||_2^2$ is related to the level of linearity between y and $\mathbf{x}$. Hence, it is reasonable to gain more testing power with larger $c_0^2||\boldsymbol{\theta}||_2^2$. We note that an increase in the value of ζ leads to decrease of testing power. Therefore, the F-test is adversely affected by the effect of high dimensionality and becomes powerless when the limit ζ of the ratio p/n is close to 1.

3.3.3 Partial Test

In this subsection, we investigate the problem of testing partial coefficients in a relatively high-dimensional regime. Specifically, the classical F-statistic designed for testing partial coefficients is studied. We derive the asymptotic normality and the asymptotic local power function of the proposed test.

Let $\mathbf{x} = (\mathbf{x}_1^\top, \mathbf{x}_2^\top)^\top$, where $\mathbf{x}_1$ is a p_1-dimensional nuisance covariate and $\mathbf{x}_2$ is a p_2-dimensional covariate of interest. SIM (3.1) is then denoted as

$$y = f(\mathbf{x}_1^\top \boldsymbol{\theta}_1 + \mathbf{x}_2^\top \boldsymbol{\theta}_2, \epsilon). \tag{3.7}$$

Suppose that $\mathbf{B}^\top$ is a $b \times p_2$ matrix of full row rank. We are interested in testing the following linear hypothesis problem

$$\mathbf{H}_{0,\mathbf{B}} : \mathbf{B}^\top \boldsymbol{\theta}_2 = \mathbf{0} \quad \text{versus} \quad \mathbf{H}_{1,\mathbf{B}} : \mathbf{B}^\top \boldsymbol{\theta}_2 \neq \mathbf{0}. \tag{3.8}$$

When $\mathbf{B}^\top = \mathbf{I}$, the problem is converted to detecting the significance of $\mathbf{x}_2$. Therefore, hypothesis testing problem (3.8) is general. We rewrite the model to highlight the part of testing interest. Let $\mathbf{H}_\mathbf{B} = \mathbf{B}(\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top$, $\mathbf{S}_\mathbf{B}$ be a $p_2 \times b$ matrix such that $\mathbf{H}_\mathbf{B} = \mathbf{S}_\mathbf{B}\mathbf{S}_\mathbf{B}^\top$ and $\mathbf{S}_\mathbf{B}^\top\mathbf{S}_\mathbf{B} = \mathbf{I}$, and $\mathbf{S}_{\mathbf{B}^\perp}$ be a $p_2 \times (p_2 - b)$ matrix such that $\mathbf{I} - \mathbf{H}_\mathbf{B} = \mathbf{S}_{\mathbf{B}^\perp}\mathbf{S}_{\mathbf{B}^\perp}^\top$ and $\mathbf{S}_{\mathbf{B}^\perp}^\top\mathbf{S}_{\mathbf{B}^\perp} = \mathbf{I}$. The existences of $\mathbf{S}_\mathbf{B}$ and $\mathbf{S}_{\mathbf{B}^\perp}$ are proved in Section 3.6.4. Let $\mathbf{w} = (\mathbf{x}_1^\top, \mathbf{x}_2^\top\mathbf{S}_{\mathbf{B}^\perp})^\top$ and $\boldsymbol{\gamma} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top\mathbf{S}_{\mathbf{B}^\perp})^\top$. Based on linear form (3.5), model (3.7) can be denoted as

$$y = c_0\mathbf{w}^\top\boldsymbol{\gamma} + c_0\mathbf{x}_2^\top\mathbf{S}_\mathbf{B}\mathbf{S}_\mathbf{B}^\top\boldsymbol{\theta}_2 + e, \tag{3.9}$$

where $\mathbf{w}$ is the nuisance covariate, $\mathbf{x}_2^\top\mathbf{S}_\mathbf{B}$ is the b-dimensional covariate of testing interest, and e is the residual term.

Suppose that observations $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ are i.i.d. from SIM (3.7). Let $\mathbf{x}_{\ell i}$ be the i-th row of the matrix $\mathbf{X}_\ell$, for $\ell = 1, 2$, and $\mathbf{y} = (y_1, \ldots, y_n)^\top$. Define $\mathbf{W} = (\mathbf{X}_1, \mathbf{X}_2\mathbf{S}_{\mathbf{B}^\perp})$. The F-statistic for testing problem (3.8) is defined as

$$F_{n,p_2} = \frac{\mathbf{y}^\top(\mathbf{H} - \mathbf{H}_\mathbf{W})\mathbf{y}/b}{\mathbf{y}^\top(\mathbf{I} - \mathbf{H})\mathbf{y}/(n-p)},$$

where $\mathbf{H}_\mathbf{W} = \mathbf{W}(\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top$. The test statistic F_{n,p_2} is well-defined, since the matrix $\mathbf{W}^\top\mathbf{W}$ is invertible with the probability 1 when $p < n$ and $\mathbf{w}$ has the normal distribution.

3.3.3.1 Asymptotic Normality

First, we derive the asymptotic normality of F_{n,p_2} under $\mathbf{H}_{0,\mathbf{B}}$. The following assumptions are needed.

Assumption L5. $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ *and* $\boldsymbol{x}$ *is independent of* ϵ.

Assumption L6. *There are constants* $\zeta, \zeta_1 \in (0,1)$*, with* $\zeta_1 < \zeta$*, such that* $\frac{p}{n} \to \zeta$ *and* $\frac{b}{n} \to \zeta_1$.

From Assumption L6, the orders of b and p are asymptotically the same, indicating that linear hypothesis testing (3.8) is a high-dimensional testing problem. The asymptotic normality of the test statistic is shown as follows.

Theorem 3.3. *Under* $\boldsymbol{H}_{0,\boldsymbol{B}}$ *and Assumptions L2, L5 and L6, as* $n \to \infty$*, we have*

$$\frac{F_{n,p_2} - 1}{\sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

This asymptotic normality result justifies the following test procedure. Given an α-level of significance, the test rejects $\mathbf{H}_{0,\mathbf{B}}$ if

$$\frac{F_{n,p_2} - 1}{\sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}} > z_\alpha,$$

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

3.3.3.2 Asymptotic Local Power Function

Then, we move to investigate the asymptotic local power function of F_{n,p_2}. Let $\mathbf{D} = \mathrm{diag}(\mathbf{I}, \mathbf{S}_{\mathbf{B}^\perp})$. Then $\mathbf{w} = \mathbf{D}^\top \mathbf{x}$. Let $\boldsymbol{\Sigma}_{22} = Var(\mathbf{x}_2)$, $\boldsymbol{\Sigma}_{12} = Cov(\mathbf{x}_1, \mathbf{x}_2)$ and $\boldsymbol{\Sigma}_{21} = Cov(\mathbf{x}_2, \mathbf{x}_1)$. Define

$$\tau^2 = \boldsymbol{\theta}_2^\top \mathbf{H}_\mathbf{B} \left[\boldsymbol{\Sigma}_{22} - (\boldsymbol{\Sigma}_{21}\ \boldsymbol{\Sigma}_{22})\, \mathbf{D}(\mathbf{D}^\top \boldsymbol{\Sigma} \mathbf{D})^{-1} \mathbf{D}^\top \begin{pmatrix} \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{22} \end{pmatrix} \right] \mathbf{H}_\mathbf{B} \boldsymbol{\theta}_2.$$

The scalar $c_0 = E(\boldsymbol{\theta}^\top \mathbf{x} y)/||\boldsymbol{\theta}||_2^2$. Additional assumption is needed for the study.

Assumption L7. $c_0^2 \tau^2 = o(1)$.

Theorem 3.4. *Suppose Assumptions L2, L5–L7 hold. Let* $\Psi_n(\boldsymbol{\theta}_2; \boldsymbol{B})$ *denote the power function of the test statistic* F_{n,p_2}*. As* $n \to \infty$*, we have*

$$\Psi_n(\boldsymbol{\theta}_2; \boldsymbol{B}) - \Phi(-z_\alpha + \sqrt{\frac{n(1-\zeta)(1-\zeta+\zeta_1)}{2\zeta_1}}\frac{c_0^2\tau^2}{\sigma^2}) \to 0,$$

where $\sigma^2 = Var(y) - c_0^2\boldsymbol{\theta}^\top\boldsymbol{\Sigma}\boldsymbol{\theta}$*,* $\Phi(\cdot)$ *is the cumulative distribution function of the standard normal distribution, and* z_α *is the upper* α*-quantile of* Φ*.*

It indicates that $\Psi_n(\boldsymbol{\theta}_2; \mathbf{B})$ is an increasing function of $c_0^2\tau^2$. In addition, when ζ increases, the loss of testing power demonstrates that the test is influenced by the effect of high dimensionality, and the test becomes powerless when ζ is close to 1.

3.4 New Test in Ultrahigh Dimensions

In this section, we consider a higher dimensional regime where the dimension p is much greater than the sample size n. In this case, the F-statistic cannot be well-defined due to the singularity of the matrix $\mathbf{X}^\top\mathbf{X}$, for which a new high-dimensional test statistic is required to address the problem. Using the technique of random projection, we propose a new test statistic based on the F-statistic of the projected data. The new test has a less restriction on the distribution of $\mathbf{x}$ and is applicable for many commonly used distributions. We derive the asymptotic normality and the asymptotic local power function of the proposed test. We also compare the properties of the proposed test with competing tests and derive sufficient conditions that guarantee its superior performance.

First, we concentrate on testing the global hypothesis

$$\mathbf{H}_0 : \boldsymbol{\theta} = \mathbf{0} \quad \text{versus} \quad \mathbf{H}_1 : \boldsymbol{\theta} \neq \mathbf{0}. \tag{3.10}$$

With the normality assumption of $\mathbf{x}$, when $n > p$, the feasibility of applying the F-test in SIM has been carefully investigated in Section 3.3. However, the F-test becomes invalid in the $p > n$ settings. To solve the problem, we randomly project the high-dimensional covariates into a lower-dimensional space, and then apply the F-test to the projected data. Specifically, for an integer $1 \leq k < \min\{n, p\}$, let $\mathbf{P}_k \in \mathbb{R}^{p\times k}$ denote a random projection matrix with random entries, drawn independently of the data. Define $\mathbf{u} = \mathbf{P}_k^\top \mathbf{x}$. We consider a model

$$y = f(\mathbf{u}^\top \boldsymbol{\eta}, \epsilon). \tag{3.11}$$

Note that the distribution of y in model (3.1) under hypothesis $\mathbf{H}_0 : \boldsymbol{\theta} = \mathbf{0}$ is the same as that in model (3.11) under hypothesis $\mathbf{H}_0 : \boldsymbol{\eta} = \mathbf{0}$. In addition, when $\boldsymbol{\eta} \neq \mathbf{0}$ and $\mathbf{P}_k$ has i.i.d. $\mathcal{N}(0, 1)$ entries, the probability of $\mathbf{P}_k\boldsymbol{\eta} \neq \mathbf{0}$ is 1. We propose a test statistic

$$T_{n,k} = \frac{\mathbf{y}^\top \mathbf{H}_k \mathbf{y}/k}{\mathbf{y}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_k)\mathbf{y}/(n - k - 1)}, \tag{3.12}$$

where $\mathbf{P}_1 = \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ 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 $\mathbf{U}_k = (\mathbf{I} - \mathbf{P}_1)\mathbf{X}\mathbf{P}_k$.

The test statistic $T_{n,k}$ can be well defined even when $p > n$, for the reason that the matrix $\mathbf{U}_k^\top \mathbf{U}_k$ is of full rank with probability 1, where $\mathbf{P}_k$ has i.i.d. $\mathcal{N}(0, 1)$ entries. This is shown in the proof of Theorem 3.5.

One of the convenient ways to construct $\mathbf{P}_k$ is to generate its i.i.d. entries from $\mathcal{N}(0, 1)$. Furthermore, Li, Hastie, and Church (2006) proposed that it is possible to generate other types of random projections $\mathbf{P}_k$, such as sparse random projections to obtain the same asymptotic performance as the normal random projections with a fast convergence speed. A sparse random projection is composed of entries p_{ij} that are i.i.d. from distributions satisfying

$$P(p_{ij} = \sqrt{l}) = P(p_{ij} = -\sqrt{l}) = \frac{1}{2l},\ P(p_{ij} = 0) = 1 - \frac{1}{l}, \tag{3.13}$$

where the recommended value of l is $\sqrt{p}$. In the theoretical analysis, we will focus on the random projection consisting of i.i.d. $\mathcal{N}(0,1)$ entries. The results are also applicable to some non-normal projections. The sparse random projections mentioned above will be used as the theoretical basis to evaluate the performance of non-normal projections in the simulation studies in Section 3.5.

3.4.1 Asymptotic Normality

The first main result demonstrates the asymptotic normality of the proposed test under $\mathbf{H}_0$. We work under the following assumptions.

Assumption H1. *$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{\Gamma}\boldsymbol{z}$, where $\boldsymbol{\Gamma}$ is a $p \times m$ matrix with $m \geq p$, $\boldsymbol{\mu}$ is a p-dimensional vector and $\boldsymbol{z} = (z_1, \ldots, z_m)^\top$ is an m-variate random vector with $E(\boldsymbol{z}) = \boldsymbol{0}, Var(\boldsymbol{z}) = \boldsymbol{I}$ and $Var(\frac{\boldsymbol{z}^\top \boldsymbol{z}}{m}) = O(m^{-1})$. For any nonnegative integers $q_1, \ldots, q_m$, with $\sum_{j=1}^m q_j = 4$, the mixed moments $E\big(\Pi_{j=1}^m z_j^{q_j}\big)$ are bounded, and equal to 0 when at least one of the q_j is odd.*

Assumption H2. *ϵ is independent of $\boldsymbol{x}$, and $E(y^4) < \infty$.*

Assumption H3. *$p \gg n$ and there is a constant $\rho \in (0,1)$ such that $\frac{k}{n} \to \rho$.*

As stated in Assumptions H1 and H3, there is no specific relationship between n and p, so that the dimension p, mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^\top$ implicitly vary as n goes to infinity, making our test applicable to ultrahigh-dimensional problems. The covariate $\mathbf{x}$ is generated from a similar structure of the factor model, which has no specific conditions for the covariance matrix $\boldsymbol{\Sigma}$. It includes a flexible family of distributions, with the elliptical distributions as special cases. The similar assumptions were adopted in Bai and Saranadasa (1996), Zhong and Chen (2011), Guo and Chen (2016) and Cui, Guo, and Zhong (2018), where stricter conditions were imposed on each element of $\mathbf{z}$. Suppose that the projection dimension k is asymptotically proportional to n with a coefficient ρ. The selection of ρ will be discussed in Section 3.5.

Since $T_{n,k}$ is invariant to the location shift of $\mathbf{y}$ and $\mathbf{X}$, we assume that $E(y) = 0$ and $\boldsymbol{\mu} = \mathbf{0}$ in the rest of the paper.

Under $\mathbf{H}_0$, the response y is independent of $\mathbf{u}$, as the result of the independence between y and $\mathbf{x}$. Therefore, it is sufficient to study the proposed test under the linear model. The asymptotic normality of the test statistic $T_{n,k}$ is established in the following theorem.

Theorem 3.5. *Suppose that the random projection matrix $\boldsymbol{P}_k$ consists of i.i.d. $\mathcal{N}(0,1)$ entries. Under $\boldsymbol{H}_0$ and Assumptions H1–H3, as $n \to \infty$, we have*

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

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

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

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

3.4.2 Asymptotic Local Power Function

In this section, we analyze the asymptotic local power function of the proposed test. Additional assumption is shown below.

Assumption H4. $E(y\boldsymbol{x}^\top)\boldsymbol{\Sigma}^{-1}E(y\boldsymbol{x}) = o(1)$.

This is known as a local alternative. In the linear model, Assumption H4 is converted to $\boldsymbol{\theta}^\top\boldsymbol{\Sigma}\boldsymbol{\theta} = o(1)$. Considering a family of models where $E(y|\mathbf{x}) = g(\mathbf{x}^\top\boldsymbol{\theta})$ for a differentiable function g while $\mathbf{x}$ follows the normal distribution, Assumption H4 can be denoted as $c_{0,k}^2\boldsymbol{\theta}^\top\boldsymbol{\Sigma}\boldsymbol{\theta} = o(1)$ with $c_{0,k} = E\{g'(\mathbf{x}^\top\boldsymbol{\theta})\}$. Specifically, GLM is included in this case.

In the study of the asymptotic local power function, the analysis method of the F-test proposed in Section 3.3 is considered. According to Diaconis and Freedman (1984), the empirical distribution of randomly projected data tends to be approximately normal. Therefore, it is expected that the result of the asymptotic local power function will be valid when the p-dimensional data are not generated from the normal distribution. The idea is illustrated by the simulation in Section 3.5. Let $\boldsymbol{\eta} = \boldsymbol{\Sigma}_1^{-1}\mathbf{P}_k^\top\boldsymbol{\Sigma}\boldsymbol{\theta}$ with $\boldsymbol{\Sigma}_1 = \mathbf{P}_k^\top\boldsymbol{\Sigma}\mathbf{P}_k$, the scalar $c_{0,k} = E(\boldsymbol{\eta}^\top\mathbf{u}y)/(\boldsymbol{\eta}^\top\boldsymbol{\Sigma}_1\boldsymbol{\eta})$ with $\mathbf{P}_k$ considered as deterministic in the expectation, and $\omega^2 = \boldsymbol{\eta}^\top\boldsymbol{\Sigma}_1\boldsymbol{\eta}$. The formal result is shown as follows.

Theorem 3.6. *Suppose that Assumptions H1–H4 hold and $\boldsymbol{z}$ follows the standard normal distribution. Let $\Psi_n^{RP}(\boldsymbol{\theta};\boldsymbol{P}_k)$ denote the power function of the proposed test $T_{n,k}$. Then*

$$\Psi_n^{RP}(\boldsymbol{\theta};\boldsymbol{P}_k) - \Phi(-z_\alpha + \sqrt{\frac{n(1-\rho)}{2\rho}\frac{c_{0,k}^2\omega^2}{\sigma^2}}) \to 0,$$

where $\sigma^2 = Var(y) - c_{0,k}^2\omega^2$, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and z_α is the upper α-quantile of Φ.

The asymptotic local function relies on $\mathbf{P}_k$ and is an increasing function of $c_{0,k}^2\omega^2$. When the vector $\boldsymbol{\Gamma}^\top\boldsymbol{\theta}$ is in the space generated by $\boldsymbol{\Gamma}^\top\mathbf{P}_k$, ω^2 can reach its upper bound $\boldsymbol{\theta}^\top\boldsymbol{\Sigma}\boldsymbol{\theta}$. To reach the boundary asymptotically, we give a sufficient condition.

Assumption H5. *(Tail eigenvalue condition) There is an integer s and a real number $\gamma > 0$ such that $s < k$ and $\frac{\sqrt{n}}{p}||\boldsymbol{\theta}||_2^2\sum_{i=s+1}^p d_i = o(n^{-\gamma})$, where d_i are the eigenvalues of $\boldsymbol{\Sigma}$ satisfying $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$.*

Assumption H5 is denoted as a tail eigenvalue condition, since it requires that the product of $||\boldsymbol{\theta}||_2^2$ and the sum of tail eigenvalues of $\boldsymbol{\Sigma}$ to be of order less than $p/\sqrt{n}$.

Lemma 3.2. *Let $\boldsymbol{P}_k \in \mathbb{R}^{p\times k}$ be composed of entries from i.i.d. $\mathcal{N}(0,1)$ entries. Suppose that Assumption H5 holds. Then we have*

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

for some $\boldsymbol{\zeta} \in \mathbb{R}^k$ with probability tending to 1.

This lemma indicates that we can approximate $\boldsymbol{\Gamma}^\top\boldsymbol{\theta}$ by $\boldsymbol{\Gamma}^\top\mathbf{P}_k\boldsymbol{\zeta}$ with a negligible approximation error. In this case, the asymptotic local power function for the proposed test is shown as follows.

Corollary 3.1. *Suppose that Assumptions H1–H5 hold and $\boldsymbol{z}$ follows the standard normal distribution. As $n \to \infty$, we have*

$$\Psi_n^{RP}(\boldsymbol{\theta};\boldsymbol{P}_k) - \Phi(-z_\alpha + \sqrt{\frac{n(1-\rho)}{2\rho}}\frac{c_{0,k}^2\boldsymbol{\theta}^\top\boldsymbol{\Sigma}\boldsymbol{\theta}}{\sigma^2}) \to 0,$$

where $\sigma^2 = Var(y) - c_{0,k}^2\boldsymbol{\theta}^\top\boldsymbol{\Sigma}\boldsymbol{\theta}$, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and z_α is the upper α-quantile of Φ.

In the corollary, ω^2 is regarded as a deterministic value inside the asymptotic local power function, even when $\mathbf{P}_k$ is randomly generated. In addition, when $E(y|\mathbf{x}) = g(\mathbf{x}^\top\boldsymbol{\theta})$ for a differentiable function g, we have $c_{0,k} = E\{g^{'}(\mathbf{x}^\top\boldsymbol{\theta})\}$, that is, $c_{0,k}$ can be determined by the model. Therefore, it is proved that $\Psi_n^{RP}(\boldsymbol{\theta};\mathbf{P}_k)$ can be a nonrandom function in some certain conditions, and it is a decreasing function of ρ.

We give some examples to illustrate the forms of $c_{0,k}$ and σ^2. Let $\lambda^2 = \boldsymbol{\theta}^\top\boldsymbol{\Sigma}\boldsymbol{\theta}$ and $\mathbf{x}$ follow the normal distribution.

Example 3.1 (Linear model)**.** *Suppose the observation $(\boldsymbol{x}, y)$ is generated from*

$$y = \boldsymbol{x}^\top\boldsymbol{\theta} + \epsilon,\ \epsilon \text{ is independent of } \boldsymbol{x}. \tag{3.14}$$

Then $c_{0,k} = 1$ and $\sigma^2 = Var(\epsilon)$.

Example 3.2 (Logistic model). *Suppose the observation* $(\boldsymbol{x}, y)$ *is generated from*

$$y|\boldsymbol{x} \sim Bernoulli\big(g(\boldsymbol{x}^\top \boldsymbol{\theta})\big),\ g(t) = \frac{\exp(t)}{1+\exp(t)}. \tag{3.15}$$

The scalar $c_{0,k}$ *and the variance* σ^2 *can be derived by*

$$c_{0,k} = E\{g^{'}(\boldsymbol{x}^\top \boldsymbol{\theta})\} \quad and$$
$$\sigma^2 = Var(y) - c_{0,k}^2\omega^2 = E\{g(\boldsymbol{x}^\top \boldsymbol{\theta})\}\left(1 - E\{g(\boldsymbol{x}^\top \boldsymbol{\theta})\}\right) - c_{0,k}^2\omega^2.$$

This result is also available to the probit model, where the link function satisfies $g(t) = \Phi(t)$.

Example 3.3 (Poisson model). *Suppose the observation* $(\boldsymbol{x}, y)$ *is generated from*

$$y|\boldsymbol{x} \sim Poisson\big(g(\boldsymbol{x}^\top \boldsymbol{\theta})\big),\ g(t) = \exp(t).$$

The scalar $c_{0,k}$ *and the variance* σ^2 *can be derived by*

$$c_{0,k} = E\{g^{'}(\boldsymbol{x}^\top \boldsymbol{\theta})\} = \exp(0.5\lambda^2) \quad and$$
$$\sigma^2 = Var(y) - c_{0,k}^2\omega^2 = \exp(\lambda^2)\left(\exp(\lambda^2) + \exp(-0.5\lambda^2) - 1 - \omega^2\right),$$

where $c_{0,k}$ *is an increasing function of* λ^2.

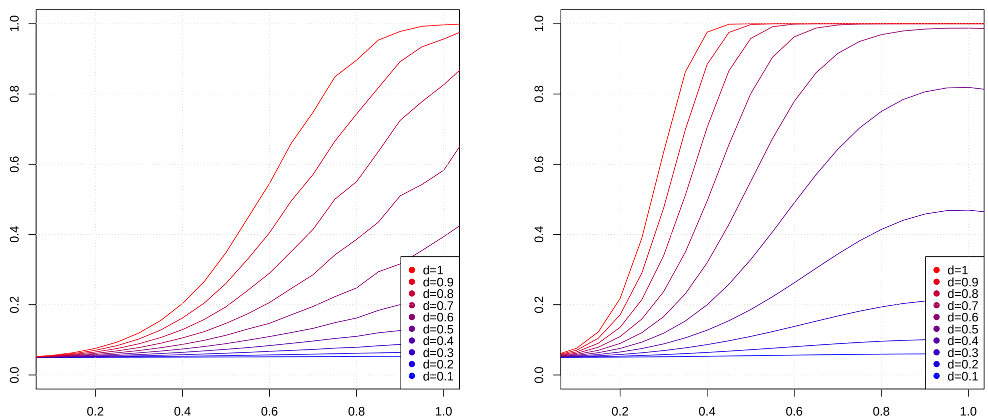
Example 3.4 (Sin model). *Suppose the response* y *is generated from*

$$y = \sin(\boldsymbol{x}^\top \boldsymbol{\theta}) + \epsilon,\ \epsilon \sim \mathcal{N}(0,1) \text{ and is independent of } \boldsymbol{x}.$$

Then, we have

$$c_{0,k} = E\{\cos(\boldsymbol{x}^\top \boldsymbol{\theta})\} = \exp(-0.5\lambda^2) \quad and$$
$$\sigma^2 = Var(y) - c_{0,k}^2\omega^2 = 1.5 - 0.5\exp(-2\lambda^2) - \omega^2\exp(-\lambda^2).$$

This indicates that $c_{0,k}$ *is a decreasing function of* λ^2.



(a) Logistic model. (b) Poisson model.

Figure 3.1: The asymptotic local power functions of logistic and Poisson models as λ^2 increases when $\omega^2 = d^2\lambda^2$ for a given d.

Therefore, explicit expressions of the asymptotic local power functions can be derived in these models. In addition, it is found that the ratio ω^2/λ^2 has a significant influence on the testing power. The ratio is determined by $\boldsymbol{\theta}$, $\boldsymbol{\Sigma}$ and $\mathbf{P}_k$ and is in the range $[0, 1]$. As shown in Lemma 3.2, the ratio is close to 1 when Assumption H5 is satisfied by $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$. We consider a simple situation where $\omega^2 = d^2\lambda^2$ for a specific value d, and we study the influence of the ratio ω^2/λ^2 on the asymptotic local power function. When $n = 600$ and $\rho = 0.4$, Figure 3.1 illustrates the asymptotic local power functions of the logistic and Poisson models for different values of d. It is found that, for a given λ^2, the largest value of the asymptotic local power can be derived when d is close to 1.

3.4.3 Theoretical Comparison

From the analysis in Sections 3.4.1 and 3.4.2, our new test is appropriate for extremely high-dimensional settings with mild assumptions about the model and the covariate. In this section, we will further research into the testing performance of the proposed test by comparing it with other tests. Specifically, we compare our test with the GLM test proposed in Guo and Chen (2016) and the test developed by Ma, Cai, and Li (2020), which focused on the logistic model in sparse settings. These tests are designed for high-dimensional testing problem (3.10) and have been demonstrated to have powerful testing performance. We denote these competing tests as GC test and MCL test, respectively. First, GC test is adopted in a certain class of GLM for comparison. For theoretical analysis, the criterion asymptotic relative efficiency (ARE) is used, for which we give a sufficient condition to guarantee that the proposed test is asymptotically more powerful than GC test. Since MCL test has no closed form of the asymptotic power function, principal comparison with MCL test is conducted through a simulation study, shown in Section 3.5, where GC test is also investigated. In this section, we compare the proposed test with MCL test in terms of model assumptions to demonstrate the general application range for our test.

For the GLM with the canonical link, the response variable y satisfies $E(y|\mathbf{x}) = g(\mathbf{x}^\top\boldsymbol{\theta})$ and $Var(y|\mathbf{x}) = \phi g^{'}(\mathbf{x}^\top\boldsymbol{\theta})$, where g is a monotone differentiable function and ϕ is a dispersion parameter. When $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, the asymptotic power function of GC test proposed by Guo and Chen (2016) is denoted as

$$\Psi_n^{GC}(\boldsymbol{\theta}) = \Phi(-z_\alpha + \frac{n||\Delta_{\boldsymbol{\theta},\mathbf{0}}||_2^2}{\sqrt{2tr\left(\{\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\mathbf{0}) + \Xi_{\boldsymbol{\theta},\mathbf{0}}\}^2\right)}}), \tag{3.16}$$

where the matrices satisfy $\Delta_{\boldsymbol{\theta},\mathbf{0}} = \phi^{-1}E\{g^{'}(\mathbf{x}^\top\boldsymbol{\theta})\}\boldsymbol{\Sigma}\boldsymbol{\theta}$, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\mathbf{0}) = \phi^{-1}E\{g^{'}(\mathbf{x}^\top\boldsymbol{\theta})\mathbf{x}\mathbf{x}^\top\}$, and $\Xi_{\boldsymbol{\theta},\mathbf{0}} = \phi^{-2}E\left[\{g(\mathbf{x}^\top\boldsymbol{\theta}) - g(0)\}^2\mathbf{x}\mathbf{x}^\top\right]$.

With a slight abuse of notation, we also denote the asymptotic power function of our random-projection-based (RP) test by

$$\Psi_n^{RP}(\boldsymbol{\theta}; \mathbf{P}_k) = \Phi(-z_\alpha + \sqrt{\frac{n(1-\rho)}{2\rho}}\frac{c_{0,k}^2\omega^2}{\sigma^2}), \tag{3.17}$$

where $c_{0,k} = E\{g^{'}(\mathbf{x}^\top \boldsymbol{\theta})\}$ in the current setting.

Since the term added to $-z_\alpha$ inside function Φ is the component controlling power, the ratio of such terms is defined as the ARE. More explicitly, we define

$$ARE(\Psi_n^{RP}; \Psi_n^{GC}) = \frac{1-\rho}{n\rho} \frac{\omega^4}{(\boldsymbol{\theta}^\top \Sigma^2 \boldsymbol{\theta})^2} \frac{\phi^4 tr(\{\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\mathbf{0}) + \Xi_{\boldsymbol{\theta},\mathbf{0}}\}^2)}{\sigma^4}. \tag{3.18}$$

Whenever the ARE is larger than 1, our procedure is considered to have a greater asymptotic power than the competing test. For this purpose, an inequality for (3.18) is derived below. Let $r_0 = y - g(0)$ and $\mathbf{F}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}^{1/2} (\mathbf{I} - \mathbf{P}_{\boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}}) \boldsymbol{\Sigma}^{1/2}$, where $\mathbf{P}_{\boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}}$ denotes the projection matrix for $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}$. Then,

$$\begin{aligned}\phi^2\{\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\mathbf{0}) + \Xi_{\boldsymbol{\theta},\mathbf{0}}\} &= E(r_0^2 \mathbf{x}\mathbf{x}^\top) \\ &= E(r_0^2)\mathbf{F}(\boldsymbol{\theta}) + E(r_0^2 \frac{(\mathbf{x}^\top \boldsymbol{\theta})^2}{\boldsymbol{\theta}^\top \boldsymbol{\Sigma} \boldsymbol{\theta}}) \frac{\boldsymbol{\Sigma}\boldsymbol{\theta}\boldsymbol{\theta}^\top \boldsymbol{\Sigma}}{\boldsymbol{\theta}^\top \boldsymbol{\Sigma} \boldsymbol{\theta}}.\end{aligned}$$

Plugging this into (3.18), under the conditions in Guo and Chen (2016), where $\boldsymbol{\Sigma}$ satisfies $tr(\boldsymbol{\Sigma}^4) = o(tr^2(\boldsymbol{\Sigma}^2))$, we obtain

$$\begin{aligned}ARE(\Psi_n^{RP}; \Psi_n^{GC}) &= \frac{1-\rho}{n\rho} \frac{\omega^4}{(\boldsymbol{\theta}^\top \boldsymbol{\Sigma}^2 \boldsymbol{\theta})^2} \frac{\phi^4 tr(\{\boldsymbol{\Sigma}_{\boldsymbol{\theta}}(\mathbf{0}) + \Xi_{\boldsymbol{\theta},\mathbf{0}}\}^2)}{\sigma^4} \\ &\geq \frac{1-\rho}{n\rho} \frac{\omega^4}{(\boldsymbol{\theta}^\top \boldsymbol{\Sigma}^2 \boldsymbol{\theta})^2} \frac{E(r_0^2)^2}{\sigma^4} tr(\mathbf{F}(\boldsymbol{\theta})^2) \qquad (3.19) \\ &\geq C_n \frac{1-\rho}{n\rho} \frac{\omega^4}{(\boldsymbol{\theta}^\top \boldsymbol{\Sigma}^2 \boldsymbol{\theta})^2} tr(\boldsymbol{\Sigma}^2),\end{aligned}$$

where C_n converges to 1 as $n \to \infty$. Let $\boldsymbol{\delta} = \boldsymbol{\Sigma}\boldsymbol{\theta}$. Then $\omega^2 = \boldsymbol{\delta}^\top \mathbf{P}_k (\mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k)^{-1} \mathbf{P}_k^\top \boldsymbol{\delta}$. Clearly, when the right side in the last inequality in (3.19) is larger than 1, it is sufficient for the ARE to give the conclusion that the proposed test has a superior performance. To derive this, with the analysis method given in Theorem 2 of Lopes, Jacob, and Wainwright (2011), we obtain the following result.

Lemma 3.3. *Suppose that $\boldsymbol{\delta}$ follows a spherical distribution with $P(\boldsymbol{\delta} = \mathbf{0}) = 0$, and is independent of $\boldsymbol{P}_k$. There is a constant $\gamma \in [0, 1)$ such that $k/p \to \gamma$. And assume*

$\frac{1}{\sqrt{k}}\frac{tr(\mathbf{\Sigma})}{p\lambda_{\min}(\mathbf{\Sigma})} = o(1)$. *For a fixed value* $\epsilon_1 > 0$, *let* $c(\epsilon_1)$ *be any constant strictly greater than* $\frac{1}{\epsilon_1\rho(1-\rho)(1-\sqrt{\gamma})^4}$. *If the condition*

$$n \geq c(\epsilon_1)\frac{tr(\mathbf{\Sigma})^2}{tr(\mathbf{\Sigma}^2)}$$

holds for all large n, *then* $P\Big(ARE(\Psi_n^{RP}; \Psi_n^{GC}) \geq 1/\epsilon_1\Big) \to 1$, *as* $n \to \infty$.

It is remarked that $\boldsymbol{\delta}$ is assumed to follow a spherical distribution, that is, the direction $\boldsymbol{\delta}/||\boldsymbol{\delta}||_2$ is uniformly generated on the unit sphere. In this case, when $tr(\mathbf{\Sigma})^2/tr(\mathbf{\Sigma}^2)$ has slower growing speed compared with the sample size n, the ARE will be large enough to demonstrate the advantages of our test.

Under the conditions in Corollary 3.1, ω^2 becomes a deterministic value, for which (3.19) leads to the following inequality.

$$ARE(\Psi_n^{RP}; \Psi_n^{GC}) \geq C_n\frac{1-\rho}{n\rho}\frac{(\boldsymbol{\theta}^\top\Sigma\boldsymbol{\theta})^2}{(\boldsymbol{\theta}^\top\Sigma^2\boldsymbol{\theta})^2}tr(\mathbf{\Sigma}^2). \tag{3.20}$$

A sufficient condition can be derived to ensure that the right hand side in (3.20) is larger than 1. Specifically, when $\frac{n^2(\boldsymbol{\xi}^\top\mathbf{\Sigma}^2\boldsymbol{\xi})^2}{p(\boldsymbol{\xi}^\top\mathbf{\Sigma}\boldsymbol{\xi})^4} = o(n^{-2\gamma})$, where $\boldsymbol{\xi} = \boldsymbol{\theta}/||\boldsymbol{\theta}||_2$ and γ is a sufficiently small positive constant, the right hand side in (3.20) goes to infinity as $n \to \infty$. This adequately demonstrates that the proposed test is asymptotically more powerful than GC test.

Next, we conduct a comparison with MCL test. Ma, Cai, and Li (2020) studied global testing problem (3.10) in the high-dimensional logistic model and constructed MCL test based on a bias-corrected estimator. The distribution of covariate $\mathbf{x}$ is limited to the multivariate normal and the bounded design, with a strong assumption about its covariance matrix. In particular, it is assumed to have bounded eigenvalues and sparse inverse. Therefore, for the reason that a larger class of models and more flexible families for the distribution of $\mathbf{x}$ are applicable in our setting, the new proposed test has a wider range of application. For alternative hypothesis, Ma, Cai, and Li (2020) was concerned about sparse alternative, shown as $\mathbf{H}_1 : \boldsymbol{\theta} \in \big\{\boldsymbol{\theta} \in \mathbb{R}^p : ||\boldsymbol{\theta}||_\infty \geq \gamma, ||\boldsymbol{\theta}||_0 \leq s\big\}$ for some $\gamma > 0$. It indicates that our test have testing property over more general alternatives. A further study via simulation is given in Section 3.5.

3.4.4 Partial Test

In this subsection, we focus on the problem of testing partial regression coefficients in an ultrahigh-dimensional regime. Specifically, we propose a new testing procedure and derive its asymptotic null distribution and asymptotic local power function. While our theoretical analysis focuses on the normal distribution, the proposed testing procedure can also be applicable for general cases, based on the property of randomly projected data, which can be demonstrated by the simulation in Section 3.5.

To emphasize the target model, SIM is restated as follows.

$$y = f(\mathbf{x}_1^\top \boldsymbol{\theta}_1 + \mathbf{x}_2^\top \boldsymbol{\theta}_2, \epsilon), \tag{3.21}$$

where $\mathbf{x}_1$ is a p_1-dimensional nuisance covariate and $\mathbf{x}_2$ is a p_2-dimensional covariate of interest. Suppose that $\mathbf{B}^\top$ is a $b \times p_2$ matrix of full row rank, we are interested in testing the linear hypothesis

$$\mathbf{H}_{0,\mathbf{B}} : \mathbf{B}^\top \boldsymbol{\theta}_2 = \mathbf{0} \quad \text{versus} \quad \mathbf{H}_{1,\mathbf{B}} : \mathbf{B}^\top \boldsymbol{\theta}_2 \neq \mathbf{0}. \tag{3.22}$$

With the same definitions of $\mathbf{H}_\mathbf{B}$, $\mathbf{S}_\mathbf{B}$, $\mathbf{S}_{\mathbf{B}^\perp}$, $\mathbf{w}$ and $\boldsymbol{\gamma}$ in Section 3.3.3, we obtain

$$\mathbf{x}_1^\top \boldsymbol{\theta}_1 + \mathbf{x}_2^\top \boldsymbol{\theta}_2 = \mathbf{w}^T \boldsymbol{\gamma} + \mathbf{x}_2^\top \mathbf{S}_\mathbf{B} \mathbf{S}_\mathbf{B}^\top \boldsymbol{\theta}_2.$$

For an integer $1 \leq k_2 < \min\{n + b - p, b\}$, let $\mathbf{P}_{k_2} \in \mathbb{R}^{b \times k_2}$ denote a random projection matrix with random entries, drawn independently of the data. Define

$\mathbf{W} = (\mathbf{X}_1, \mathbf{X}_2\mathbf{S}_{\mathbf{B}^\perp})$ and $\mathbf{U}_{k2} = (\mathbf{W}, \mathbf{X}_2\mathbf{S}_\mathbf{B}\mathbf{P}_{k_2})$. Their projection matrices are given as

$$\mathbf{H}_\mathbf{W} = (\mathbf{I} - \mathbf{P}_1)\mathbf{W}(\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{W})^{-1}\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1),$$

$$\mathbf{H}_{\mathbf{U}_{k2}} = (\mathbf{I} - \mathbf{P}_1)\mathbf{U}_{k2}(\mathbf{U}_{k2}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{U}_{k2})^{-1}\mathbf{U}_{k2}^\top(\mathbf{I} - \mathbf{P}_1).$$

We propose a test statistic

$$T_{n,k_2} = \frac{\mathbf{y}^\top(\mathbf{H}_{\mathbf{U}_{k2}} - \mathbf{H}_\mathbf{W})\mathbf{y}/k_2}{\mathbf{y}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_{\mathbf{U}_{k2}})\mathbf{y}/(n + b - p - k_2 - 1)}.$$

3.4.4.1 Asymptotic Normality

We first derive the asymptotic normality of the test statistic under $\mathbf{H}_{0,\mathbf{B}}$, for which the following assumption is made to facilitate our analysis.

Assumption H6. *$p = p_1 + p_2 \gg n$, $b \gg p_1$, and there are constants $\rho_1, \rho_2 \in (0, 1)$, with $\rho_1 + \rho_2 < 1$, such that $\frac{p-b}{n} \to \rho_1$ and $\frac{k_2}{n} \to \rho_2$.*

Since T_{n,k_2} is invariant to the location shift of $\mathbf{y}$, $\mathbf{X}_1$ and $\mathbf{X}_2$, we assume $E(y) = 0$ and $E(\mathbf{x}) = \mathbf{0}$ in the following.

Theorem 3.7. *Suppose that Assumptions H1, H2 and H6 hold and $\boldsymbol{z}$ follows the standard normal distribution. Under $\boldsymbol{H}_{0,\boldsymbol{B}}$, as $n \to \infty$, we have*

$$\frac{T_{n,k_2} - 1}{\sqrt{2(1 - \rho_1)/n\rho_2(1 - \rho_1 - \rho_2)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

This asymptotic normality result justifies the following test procedure. Given a α-level of significance, $\mathbf{H}_{0,\mathbf{B}}$ is rejected when

$$\frac{T_{n,k_2} - 1}{\sqrt{2(1 - \rho_1)/n\rho_2(1 - \rho_1 - \rho_2)}} > z_\alpha,$$

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

3.4.4.2 Asymptotic Local Power Function

We now investigate the asymptotic local power function of the test statistic. With the same definitions of $\mathbf{D}$, $\mathbf{\Sigma}_{22}$, $\mathbf{\Sigma}_{12}$, and $\mathbf{\Sigma}_{21}$ in Section 3.3.3, we define $\mathbf{R}_{k_2} = (\mathbf{D}, (\mathbf{0}, \mathbf{P}_{k_2}^\top \mathbf{S}_\mathbf{B}^\top)^\top)$ and $\boldsymbol{\xi} = (\mathbf{R}_{k_2}^\top \mathbf{\Sigma} \mathbf{R}_{k_2})^{-1} \mathbf{R}_{k_2}^\top \mathbf{\Sigma} \boldsymbol{\theta}$. Then $\mathbf{U}_{k2} = \mathbf{X}\mathbf{R}_{k_2}$. We divide the p-dimensional vector $\mathbf{R}_{k_2}\boldsymbol{\xi} = (\boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)^\top$, where $\boldsymbol{\eta}_1 \in \mathbb{R}^{p_1}$ and $\boldsymbol{\eta}_2 \in \mathbb{R}^{p_2}$. Let

$$\nu^2 = \boldsymbol{\eta}_2^\top \left[\mathbf{\Sigma}_{22} - (\mathbf{\Sigma}_{21}\ \mathbf{\Sigma}_{22})\, \mathbf{D}(\mathbf{D}^\top \mathbf{\Sigma} \mathbf{D})^{-1} \mathbf{D}^\top \begin{pmatrix} \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{22} \end{pmatrix} \right] \boldsymbol{\eta}_2,$$

and $\tau^2 = \boldsymbol{\theta}_2^\top \mathbf{H}_\mathbf{B} \left[\mathbf{\Sigma}_{22} - (\mathbf{\Sigma}_{21}\ \mathbf{\Sigma}_{22})\, \mathbf{D}(\mathbf{D}^\top \mathbf{\Sigma} \mathbf{D})^{-1} \mathbf{D}^\top \begin{pmatrix} \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{22} \end{pmatrix} \right] \mathbf{H}_\mathbf{B} \boldsymbol{\theta}_2$. Conditional on $\mathbf{P}_{k_2}$, define $c_{0,k_2} = E(\boldsymbol{\xi}^\top \mathbf{R}_{k_2}^\top \mathbf{x} y)/\boldsymbol{\xi}^\top \mathbf{R}_{k_2}^\top \mathbf{\Sigma} \mathbf{R}_{k_2} \boldsymbol{\xi}$. Additional assumption is needed to facilitate the study.

Assumption H7. $(\boldsymbol{\theta}^\top \mathbf{\Sigma} \boldsymbol{\theta})^{-1} E(y\boldsymbol{x})^\top \mathbf{\Sigma}^{-1} E(y\boldsymbol{x}) \tau^2 = o(1)$.

This is known as a local alternative. In the linear model, Assumption H7 is converted to $\tau^2 = o(1)$. Considering a family of models where $E(y|\mathbf{x}) = g(\mathbf{x}^\top \boldsymbol{\theta})$ for a differentiable function g and $\mathbf{x}$ follows the normal distribution, Assumption H7 is converted to $E\{g'(\mathbf{x}^\top \boldsymbol{\theta})\}^2 \tau^2 = o(1)$.

Theorem 3.8. *Suppose that Assumptions H1, H2, H6 and H7 hold and* $\boldsymbol{z}$ *follows the standard normal distribution. Let* $\Psi_n^{RP}(\boldsymbol{\theta}_2; \boldsymbol{B}, \boldsymbol{P}_{k_2})$ *denote the power function of the proposed test* T_{n,k_2}*. As* $n \to \infty$*, we have*

$$\Psi_n^{RP}(\boldsymbol{\theta}_2; \boldsymbol{B}, \boldsymbol{P}_{k_2}) - \Phi\left(-z_\alpha + \sqrt{\frac{n(1-\rho_1-\rho_2)(1-\rho_1)}{2\rho_2}} \frac{c_{0,k_2}^2 \nu^2}{\sigma^2} \right) \to 0,$$

where $\sigma^2 = Var(y) - c_{0,k_2}^2 \boldsymbol{\xi}^\top \boldsymbol{R}_{k_2}^\top \mathbf{\Sigma} \boldsymbol{R}_{k_2} \boldsymbol{\xi}$*,* $\Phi(\cdot)$ *is the cumulative distribution function of the standard normal distribution, and* z_α *is the upper* α*-quantile of* Φ*.*

From the result, the asymptotic local power function is an increasing function of $c_{0,k_2}^2\nu^2$. It is found that $\nu^2 \leq \tau^2$. We give a sufficient condition when the upper bound is asymptotically reached.

Assumption H8. *There is an integer $s_2 < k_2$ and a real number $\gamma_2 > 0$, such that $\frac{\sqrt{n}}{b}||\boldsymbol{S}_{\boldsymbol{B}}^\top\boldsymbol{\theta}_2||_2^2\sum_{i=s_2+1}^{b} d_i = o(n^{-\gamma_2})$, where d_i are the eigenvalues of $\boldsymbol{S}_{\boldsymbol{B}}^\top\boldsymbol{\Sigma}_{22}\boldsymbol{S}_{\boldsymbol{B}}$ satisfying $d_1 \geq d_2 \geq \cdots \geq d_b \geq 0$.*

This assumption ensures that Lemma 3.2 is valid for $\mathbf{S}_{\mathbf{B}}^\top\boldsymbol{\theta}_2$ and $\mathbf{S}_{\mathbf{B}}^\top\boldsymbol{\Sigma}_{22}\mathbf{S}_{\mathbf{B}}$. In this case, the asymptotic local power function of the proposed test statistic is shown as below.

Corollary 3.2. *Suppose that Assumptions H1, H2, H6, H7 and H8 hold and $\boldsymbol{z}$ follows the standard normal distribution. As $n \to \infty$, we have*

$$\Psi_n^{RP}(\boldsymbol{\theta}_2; \boldsymbol{B}, \boldsymbol{P}_{k_2}) - \Phi\left(-z_\alpha + \sqrt{\frac{n(1-\rho_1-\rho_2)(1-\rho_1)}{2\rho_2}\frac{c_{0,k_2}^2\tau^2}{\sigma^2}}\right) \to 0,$$

where $\sigma^2 = Var(y) - c_{0,k_2}^2\boldsymbol{\theta}^\top\boldsymbol{\Sigma}\boldsymbol{\theta}$, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution, and z_α is the upper α-quantile of Φ.

When $E(y|\mathbf{x}) = g(\mathbf{x}^\top\boldsymbol{\theta})$ for a differentiable function g, the scalar $c_{0,k_2} = E\{g^{'}(\mathbf{x}^\top\boldsymbol{\theta})\}$. Hence, it is proved that $\Psi_n^{RP}(\boldsymbol{\theta}_2; \mathbf{B}, \mathbf{P}_{k_2})$ can be a nonrandom function in some certain conditions, and it is a decreasing function of ρ_1. According to the examples illustrated in Section 3.4.2, for a given τ^2, large values of the asymptotic local power function can be reached when the ratio ν^2/τ^2 is close to 1.

3.5 Numerical Studies

3.5.1 Simulation Studies

We conducted simulations to evaluate the finite-sample performance of the proposed tests and compare it with GC and MCL tests.

The first simulation study was designed for testing the global hypothesis

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

Two SIMs were analyzed: logistic model and Poisson model. For both the experimented models, the covariate $\mathbf{x}$ was generated from $\boldsymbol{\Sigma}^{1/2}\mathbf{z}$, where each entry of $\mathbf{z}$ was i.i.d. from $\mathcal{N}(0,1)$ or $U(-\sqrt{3},\sqrt{3})$. The condition of $\boldsymbol{\Sigma}$ with different category was studied, which includes the settings that $\boldsymbol{\Sigma}$ is sparse and non-sparse. Concretely, $\boldsymbol{\Sigma}$ was generated based on $\mathbf{ODO}^\top$, where $\mathbf{O}$ was an orthogonal matrix and $\mathbf{D} = \operatorname{diag}(d_1, \ldots, d_p)$. Let $s = \lfloor n^{0.8} \rfloor$ and $L = n$. The diagonal entries of $\mathbf{D}$ were set as $d_i = 1$, for $i \leq s$, and $d_i = (L-s)w_i/W$, for $i = s+1, \ldots, p$, where $w_i = (i-s)^{-4}$ and $W = \sum_{i=s+1}^{p} w_i$. The design of $\mathbf{O}$ came from the block-wise diagonal matrix structure. Specifically, $\mathbf{O} = \operatorname{diag}(\mathbf{O}_1, \ldots, \mathbf{O}_B)$ was a block-wise diagonal matrix including B blocks, and each block was independently and uniformly generated on the $m \times m$ orthogonal group, with $p = Bm$. To study different types of $\boldsymbol{\Sigma}$, two different settings were analyzed as follows: (i) $\boldsymbol{\Sigma}_1$: $B = 1$. (ii) $\boldsymbol{\Sigma}_2$: $B = 100$ for strong sparsity. To create regimes of high dimensionality, we considered $(n,p) = (400, 1000)$ and $(n,p) = (600, 3000)$ in the simulation.

For alternative, both sparse and non-sparse cases were investigated. The vector of coefficients was generated by $\boldsymbol{\theta} = b\boldsymbol{\delta}/\sqrt{\boldsymbol{\delta}^\top\boldsymbol{\Sigma}\boldsymbol{\delta}}$, where b was a positive real value and $\boldsymbol{\delta}$ was an p-dimensional vector determining the sparsity of $\boldsymbol{\theta}$. In the simulation, two different types of $\boldsymbol{\delta}$ were considered as follows: (i) $\boldsymbol{\delta}_1$: $\delta_{1,j} = 1$, for $j \in S$, where the set S was randomly selected over $\{1, \ldots, p\}$ and had size $|S| = 10$, otherwise, $\delta_{1,j} = 0$. (ii) $\boldsymbol{\delta}_2$: randomly selected from $\operatorname{Span}\{\mathbf{u}_1, \ldots, \mathbf{u}_{100}\}$, where $\mathbf{u}_i$ was the i-th column of $\mathbf{O}$. The value of b^2 was selected in the range such that $c_{0,k}^2 b^2 \leq 0.25$. Specifically, for the logistic model, $b^2 = 0.4, 0.8$ were investigated in the simulation. For the Poisson model, $b^2 = 0.1, 0.2$ were considered in the simulation.

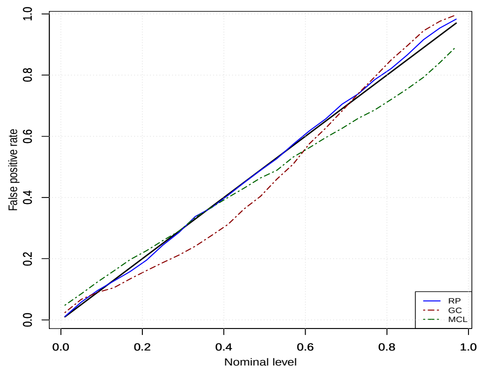
In Tables 3.1 and 3.2, we report the type I errors and the empirical powers of RP test, GC test and MCL test in the experimented models and settings based on 1000 simulations. It shows that the type I errors for RP and GC tests are around 0.05, which verifies the result in Theorem 3.5. And for MCL test, when the covariance matrix is set to $\mathbf{\Sigma}_1$, the type I errors might exceed 0.1. The results illustrate the strong sparsity requirement of MCL test, indicating that our proposed test has a wider range of application as discussed in Section 3.4.3. For empirical powers, since MCL test is designed for the binary response, a comparison with MCL test was only considered in the logistic model. As shown in the tables, the empirical power increases as b^2 grows. And for RP test, for different ρ, larger empirical powers are shown in the cases with $\rho = 0.4$. This may be the result of a higher probability of deriving a larger value of ω^2 in a larger projection space. However, for the settings with fixed ω^2, that is, Assumption H5 is satisfied by $\boldsymbol{\theta}$ and $\mathbf{\Sigma}$, larger empirical powers will be given in the case with smaller ρ, which is demonstrated in the second simulation study. Hence, the selection of ρ depends on the setting. According to Lopes, Jacob, and Wainwright (2011), the ratio $tr(\mathbf{\Sigma})^2/tr(\mathbf{\Sigma}^2)$ can be viewed as measuring the decay rate of the spectrum of $\mathbf{\Sigma}$, with the tail eigenvalue condition satisfied when $tr(\mathbf{\Sigma})^2/tr(\mathbf{\Sigma}^2) \ll p$. It indicates that ρ could be determined by an estimation of the ratio, which is available based on the results in Chen, Zhang, and Zhong (2010). Compared with GC test, the advantages of our test are illustrated in all the experimented models and settings. This demonstrates our theoretical comparison in Section 3.4.3. In comparison with MCL test, our proposed test shows higher testing power in most settings, including the alternative generated from $\boldsymbol{\delta}_1$, which has strong sparsity. Consequently, our test has powerful performance and provides an applicable solution for the condition that the property of the vector of coefficients is unknown in advance, which often happens in practice.

Table 3.1: Type I errors and empirical powers of RP, GC and MCL tests at the significance level 0.05 in logistic models.

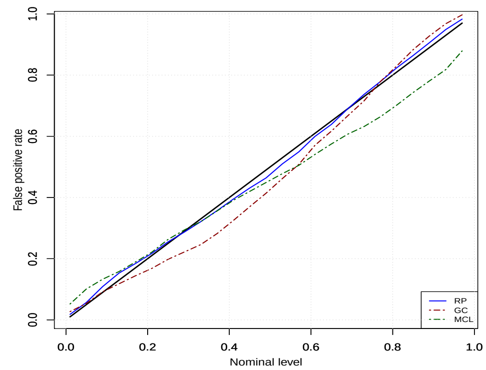
Type of $\mathbf{z}$	Type of $\mathbf{\Sigma}$	Type of $\boldsymbol{\theta}$	b^2	RP $\rho = 0.2$	RP $\rho = 0.4$	GC	MCL
Logistic model when $(n, p) = (400, 1000)$.							
$\mathcal{N}(0,1)$	$\mathbf{\Sigma}_1$	0	0	0.053	0.058	0.058	0.112
		$\boldsymbol{\delta}_1$	0.4	0.445	0.483	0.371	0.260
			0.8	0.765	0.811	0.737	0.561
		$\boldsymbol{\delta}_2$	0.4	0.425	0.480	0.272	0.136
			0.8	0.767	0.829	0.595	0.263
	$\mathbf{\Sigma}_2$	0	0	0.049	0.054	0.062	0.071
		$\boldsymbol{\delta}_1$	0.4	0.446	0.469	0.287	0.306
			0.8	0.784	0.836	0.654	0.683
		$\boldsymbol{\delta}_2$	0.4	0.451	0.434	0.261	0.078
			0.8	0.760	0.824	0.613	0.179
$U(-\sqrt{3}, \sqrt{3})$	$\mathbf{\Sigma}_1$	0	0	0.054	0.056	0.058	0.091
		$\boldsymbol{\delta}_1$	0.4	0.451	0.442	0.321	0.248
			0.8	0.756	0.806	0.711	0.575
		$\boldsymbol{\delta}_2$	0.4	0.463	0.481	0.278	0.151
			0.8	0.776	0.812	0.629	0.265
	$\mathbf{\Sigma}_2$	0	0	0.051	0.055	0.059	0.065
		$\boldsymbol{\delta}_1$	0.4	0.522	0.471	0.300	0.296
			0.8	0.855	0.839	0.690	0.703
		$\boldsymbol{\delta}_2$	0.4	0.471	0.462	0.265	0.083
			0.8	0.773	0.839	0.588	0.193
Logistic model when $(n, p) = (600, 3000)$.							
$\mathcal{N}(0,1)$	$\mathbf{\Sigma}_1$	0	0	0.052	0.054	0.057	0.051
		$\boldsymbol{\delta}_1$	0.4	0.582	0.583	0.339	0.288
			0.8	0.930	0.938	0.815	0.684
		$\boldsymbol{\delta}_2$	0.4	0.586	0.574	0.347	0.174
			0.8	0.909	0.929	0.794	0.385
	$\mathbf{\Sigma}_2$	0	0	0.051	0.058	0.057	0.042
		$\boldsymbol{\delta}_1$	0.4	0.614	0.580	0.339	0.733
			0.8	0.914	0.926	0.772	0.993
		$\boldsymbol{\delta}_2$	0.4	0.606	0.569	0.372	0.081
			0.8	0.939	0.926	0.812	0.180
$U(-\sqrt{3}, \sqrt{3})$	$\mathbf{\Sigma}_1$	0	0	0.052	0.057	0.054	0.060
		$\boldsymbol{\delta}_1$	0.4	0.601	0.562	0.345	0.300
			0.8	0.937	0.940	0.810	0.709
		$\boldsymbol{\delta}_2$	0.4	0.630	0.609	0.381	0.225
			0.8	0.938	0.934	0.819	0.397
	$\mathbf{\Sigma}_2$	0	0	0.054	0.058	0.061	0.030
		$\boldsymbol{\delta}_1$	0.4	0.595	0.582	0.346	0.722
			0.8	0.914	0.919	0.760	0.990
		$\boldsymbol{\delta}_2$	0.4	0.607	0.588	0.343	0.086
			0.8	0.921	0.926	0.787	0.187

Table 3.2: Type I errors and empirical powers of RP, GC and MCL tests at the significance level 0.05 in Poisson models.

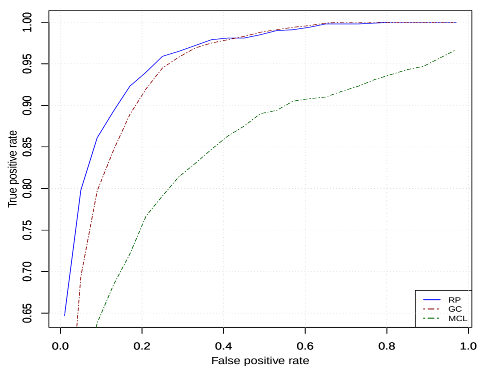
Type of $\mathbf{z}$	Type of $\boldsymbol{\Sigma}$	Type of $\boldsymbol{\theta}$	b^2	RP $\rho = 0.2$	RP $\rho = 0.4$	GC
Poisson model when $(n, p) = (400, 1000)$.						
$\mathcal{N}(0,1)$	$\boldsymbol{\Sigma}_1$	0	0	0.059	0.058	0.059
		$\boldsymbol{\delta}_1$	0.1	0.509	0.545	0.391
			0.2	0.872	0.920	0.838
		$\boldsymbol{\delta}_2$	0.1	0.468	0.563	0.332
			0.2	0.866	0.916	0.747
	$\boldsymbol{\Sigma}_2$	0	0	0.050	0.052	0.060
		$\boldsymbol{\delta}_1$	0.1	0.521	0.542	0.343
			0.2	0.876	0.935	0.809
		$\boldsymbol{\delta}_2$	0.1	0.505	0.549	0.317
			0.2	0.872	0.932	0.767
$U(-\sqrt{3}, \sqrt{3})$	$\boldsymbol{\Sigma}_1$	0	0	0.056	0.057	0.055
		$\boldsymbol{\delta}_1$	0.1	0.507	0.517	0.386
			0.2	0.873	0.928	0.844
		$\boldsymbol{\delta}_2$	0.1	0.506	0.533	0.323
			0.2	0.873	0.908	0.765
	$\boldsymbol{\Sigma}_2$	0	0	0.058	0.056	0.061
		$\boldsymbol{\delta}_1$	0.1	0.527	0.525	0.353
			0.2	0.879	0.920	0.807
		$\boldsymbol{\delta}_2$	0.1	0.511	0.513	0.331
			0.2	0.872	0.914	0.758
Poisson model when $(n, p) = (600, 3000)$.						
$\mathcal{N}(0,1)$	$\boldsymbol{\Sigma}_1$	0	0	0.057	0.054	0.060
		$\boldsymbol{\delta}_1$	0.1	0.671	0.659	0.445
			0.2	0.974	0.983	0.913
		$\boldsymbol{\delta}_2$	0.1	0.677	0.668	0.442
			0.2	0.977	0.986	0.927
	$\boldsymbol{\Sigma}_2$	0	0	0.055	0.059	0.062
		$\boldsymbol{\delta}_1$	0.1	0.615	0.657	0.305
			0.2	0.940	0.983	0.775
		$\boldsymbol{\delta}_2$	0.1	0.698	0.667	0.438
			0.2	0.978	0.985	0.923
$U(-\sqrt{3}, \sqrt{3})$	$\boldsymbol{\Sigma}_1$	0	0	0.055	0.060	0.058
		$\boldsymbol{\delta}_1$	0.1	0.689	0.681	0.451
			0.2	0.980	0.985	0.920
		$\boldsymbol{\delta}_2$	0.1	0.685	0.648	0.457
			0.2	0.984	0.985	0.918
	$\boldsymbol{\Sigma}_2$	0	0	0.059	0.056	0.060
		$\boldsymbol{\delta}_1$	0.1	0.586	0.645	0.298
			0.2	0.941	0.981	0.769
		$\boldsymbol{\delta}_2$	0.1	0.689	0.673	0.437
			0.2	0.980	0.987	0.902



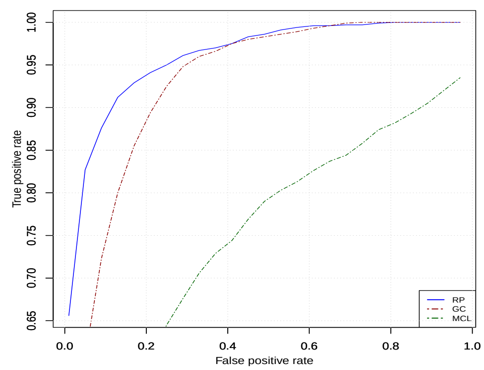
(a) The null hypothesis with $\boldsymbol{\Sigma}_1$.



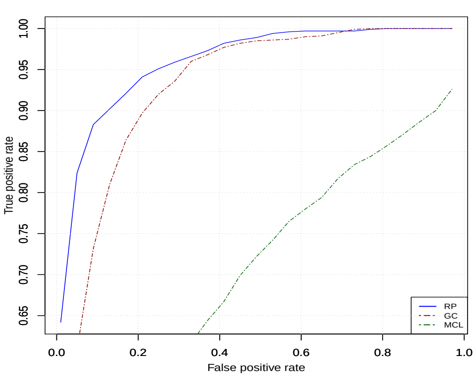
(b) The null hypothesis with $\boldsymbol{\Sigma}_2$.



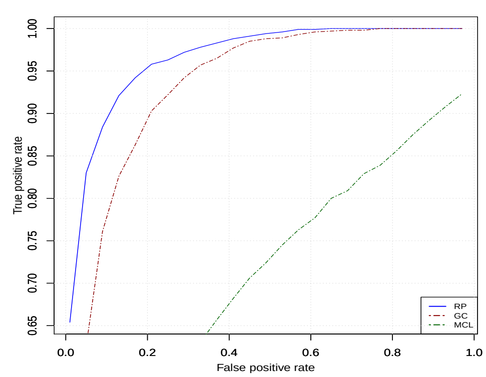
(c) Alternative $\boldsymbol{\delta}_1$ with $\boldsymbol{\Sigma}_1$.



(d) Alternative $\boldsymbol{\delta}_1$ with $\boldsymbol{\Sigma}_2$.

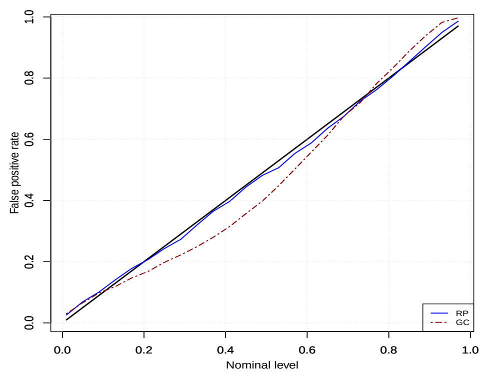


(e) Alternative $\boldsymbol{\delta}_2$ with $\boldsymbol{\Sigma}_1$.

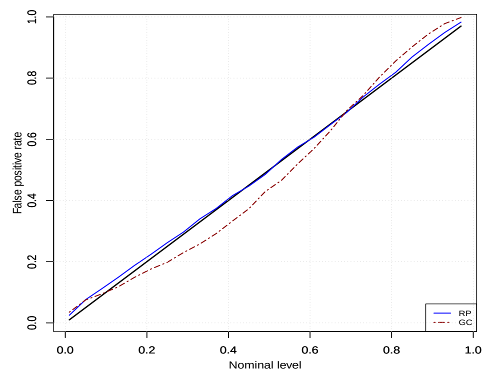


(f) Alternative $\boldsymbol{\delta}_2$ with $\boldsymbol{\Sigma}_2$.

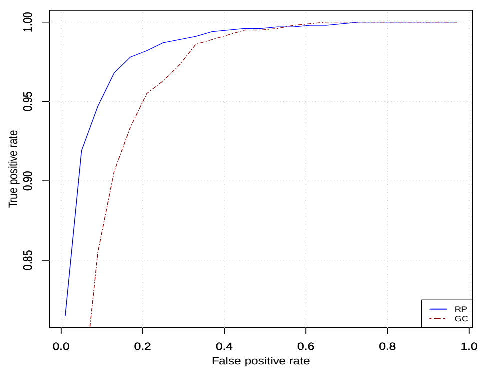
Figure 3.2: Type I errors and empirical powers of RP, GC and MCL tests for the logistic model when $\rho = 0.4$ and $(n, p) = (400, 1000)$.



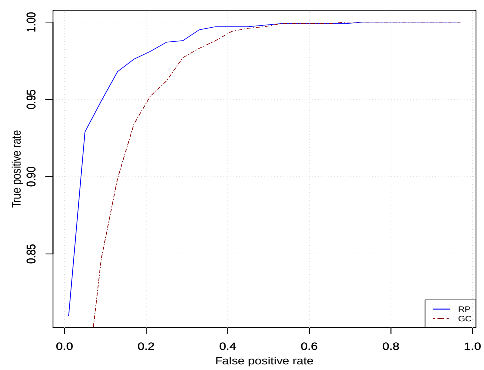
(a) The null hypothesis with $\boldsymbol{\Sigma}_1$.



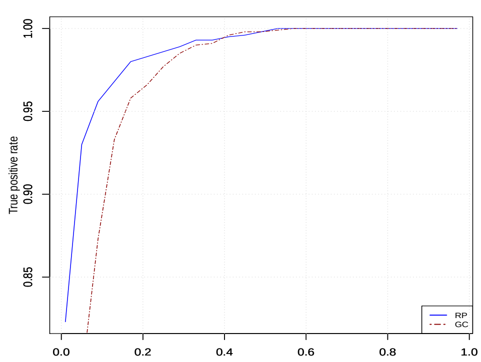
(b) The null hypothesis with $\boldsymbol{\Sigma}_2$.



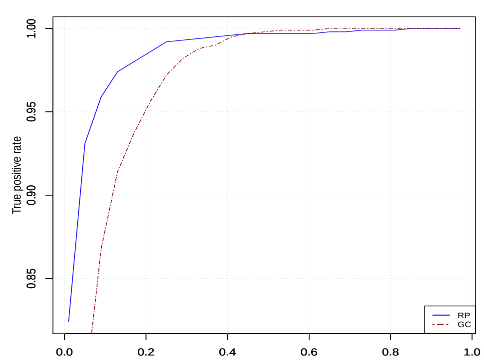
(c) Alternative $\boldsymbol{\delta}_1$ with $\boldsymbol{\Sigma}_1$.



(d) Alternative $\boldsymbol{\delta}_1$ with $\boldsymbol{\Sigma}_2$.



(e) Alternative $\boldsymbol{\delta}_2$ with $\boldsymbol{\Sigma}_1$.



(f) Alternative $\boldsymbol{\delta}_2$ with $\boldsymbol{\Sigma}_2$.

Figure 3.3: Type I errors and empirical powers of RP and GC tests for the Poisson model when $\rho = 0.4$ and $(n, p) = (400, 1000)$.

Figure 3.2 illustrates the numerical comparison between RP test and the competing tests in the logistic model when the nominal significance level varies from 0.1 to 0.9. Figures 3.2(a) and 3.2(b) report the type I errors of the tests. In the figures, the blue lines are close to the diagonal black lines, which indicates that RP test has well control of the type I error. In Figures 3.2(c)–3.2(f), the empirical powers of tests under different alternatives and covariance matrices of $\mathbf{x}$ are presented. As shown in the figures, the blue lines keep above the other lines. It illustrates the higher testing power of our tests, which clearly demonstrates the advantages. In Figure 3.3, the comparison between RP test and GC test in the Poisson model is reported.

The second simulation study was designed for testing the hypothesis

$$\mathbf{H}_{0,p_2} : \boldsymbol{\theta}_2 = \mathbf{0} \quad \text{versus} \quad \mathbf{H}_{1,p_2} : \boldsymbol{\theta}_2 \neq \mathbf{0}.$$

The logistic model was analyzed and the covariate $\mathbf{x}$ was generated from $\boldsymbol{\Sigma}^{1/2}\mathbf{z}$. Each entry of $\mathbf{z}$ was i.i.d. from $\mathcal{N}(0,1)$, $U(-\sqrt{3},\sqrt{3})$, or Rademacher distribution. A specific type of $\boldsymbol{\Sigma}^{1/2}$ was considered. We generated the matrix from $\boldsymbol{\Sigma}^{1/2} = \text{diag}\big(\mathbf{O}_1\sqrt{\mathbf{D}_1}\mathbf{O}_1^\top, \mathbf{O}_2\sqrt{\mathbf{D}_2}\mathbf{O}_2^\top\big)$, where $\mathbf{O}_1$ ($\mathbf{O}_2$) was an orthogonal matrix that was generated from the same way as the matrix $\mathbf{O}$ in the first simulation study, when p took p_1 (p_2) and $B = 1$ ($B = 100$). The entries of diagonal matrix $\mathbf{D}_1$ were from $\mathcal{N}(\mathbf{0},\mathbf{I})$ with absolute values taken and the entries of diagonal matrix $\mathbf{D}_2$ were generated in the same way as the first simulation study, when $s = \lfloor n^{0.72} \rfloor$ and $L = \lfloor n^{0.8} \rfloor$. In the simulation, we considered $(n, p_1, p_2) = (400, 40, 1000)$.

For the alternative, we generated $\boldsymbol{\theta}_1$ from $\mathcal{N}(\mathbf{0},\mathbf{I})$ and then scaled it to have $||\boldsymbol{\theta}_1||_2 = 1$. The vector of coefficients $\boldsymbol{\theta}_2$ was generated from $\boldsymbol{\theta}_2 = b_2\boldsymbol{\delta}_2/\sqrt{\boldsymbol{\delta}_2^\top\boldsymbol{\Sigma}_{22}\boldsymbol{\delta}_2}$, where b_2^2 took values $0.4, 0.8$ in the simulation and $\boldsymbol{\delta}_2$ was a p_2-dimensional vector determining the sparsity level of $\boldsymbol{\theta}_2$. Two types of $\boldsymbol{\delta}_2$ were analyzed as follows: (i) $\boldsymbol{\delta}_{2,1}$: $\delta_{2,1,j} = 1$, for $j \in S$, where the set S was randomly selected over $\{1,\ldots,p_2\}$ and had size $|S| = 10$, otherwise, $\delta_{2,1,j} = 0$. (ii) $\boldsymbol{\delta}_{2,2}$: randomly selected from the space that was spanned by the eigenvectors corresponding to the largest 100 eigenvalues of $\mathbf{\Sigma}_{22}$.

To demonstrate the feasibility of implementing other random projection, we investigated two other random-projection-based tests: (i) multi-RP test: independently generating normal random projection for 10 times and utilizing their mean; (ii) S-RP test: applying sparse random projection defined in (3.13) with $l = 400$.

Table 3.3: Type I errors and empirical powers of RP, multi-RP and S-RP tests at the significance level 0.05 when $(n, p_1, p_2) = (400, 40, 1000)$.

Type of $\mathbf{z}$	Type of $\boldsymbol{\theta}_2$	b_2^2	$\rho_2 = 0.2$			$\rho_2 = 0.4$		
			RP	multi-RP	S-RP	RP	multi-RP	S-RP
$\mathcal{N}(0,1)$	-	0	0.053	0.057	0.054	0.062	0.064	0.050
	$\boldsymbol{\delta}_{2,1}$	0.4	0.575	0.558	0.518	0.357	0.389	0.366
		0.8	0.906	0.905	0.870	0.716	0.700	0.686
	$\boldsymbol{\delta}_{2,2}$	0.4	0.559	0.553	0.548	0.351	0.349	0.354
		0.8	0.905	0.893	0.887	0.683	0.668	0.687
$U(-\sqrt{3}, \sqrt{3})$	-	0	0.061	0.056	0.050	0.060	0.053	0.052
	$\boldsymbol{\delta}_{2,1}$	0.4	0.554	0.554	0.539	0.387	0.368	0.371
		0.8	0.892	0.894	0.857	0.683	0.694	0.691
	$\boldsymbol{\delta}_{2,2}$	0.4	0.527	0.532	0.517	0.336	0.366	0.336
		0.8	0.904	0.901	0.890	0.701	0.710	0.689
Rademacher	-	0	0.052	0.050	0.058	0.060	0.056	0.057
	$\boldsymbol{\delta}_{2,1}$	0.4	0.555	0.548	0.544	0.400	0.362	0.365
		0.8	0.901	0.897	0.875	0.706	0.716	0.711
	$\boldsymbol{\delta}_{2,2}$	0.4	0.556	0.558	0.541	0.328	0.358	0.342
		0.8	0.909	0.903	0.885	0.692	0.700	0.713

Table 3.3 reports the type I errors and empirical powers of the random-projection-based tests based on 1000 simulations. The type I errors are around 0.05, and the performances of three tests have negligible differences, indicating the feasible implementing of other random projection. For both alternatives, the empirical power increases as b_2^2 grows. And larger empirical powers are shown with smaller ρ_2, which is consistent with the result in Corollary 3.1. When $\mathbf{z}$ belongs to other types of

distribution rather than the normal distribution, the results are similar to that when $\mathbf{z}$ is normal. Therefore, it is possible for the proposed method to be applied in general situations while controlling the type I error well. And for alternatives, the results indicate that the derived asymptotic power function might also be valid for non-normal distribution.

3.5.2 Application

The breast cancer is known to have biologically heterogeneity characterized by variant pathological features, disparate response to therapeutics, etc. Traditional breast cancer treatment methods are guided by a classification based on the expression levels of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). In the last decade, research on global gene expression analyses indicated a more complex breast cancer portrait and identified at least four intrinsic molecular subtypes, which have significant differences in terms of response to therapies. To identify the subtypes, a clinically gene expression-based test, named as prediction analysis of microarray 50 (PAM50) was introduced (Parker et al. , 2009), which is based on the expression measurement of 50 genes and shows a high accuracy of identification.

In this study, we illustrate our proposed methods by analyzing a real data set of breast cancer, which is available under accession number GSE50948 in the Gene Expression Omnibus (GEO). In the data set, the gene expression analysis was performed for 156 samples from the enrolled patients in the NOAH trail, which consists of 114 patients with HER2+ locally advanced or inflammatory breast cancer and 42 patients with HER2-disease. According to completeness of the information, $n = 152$ samples were used in the study. The detail of the data set was given in Prat et al. (2014). As multiple probes might represent the same gene, the measurement for each gene was from the probe with the highest interquartile range. Then, expression values of 20592 genes were obtained. In addition, some standard clinicopathologic variables were also considered in the analysis, including age at diagnosis, histologic grade, tumor size, histology, and hormonal receptor status with values of 1 and 0 corresponding to positive and negative status, respectively.

First, we studied the overall association between the status of HER2, ER, PR with other standard clinicopathologic features and gene expression levels. To illustrate our analysis method, the case considering HER2 was given as an example. The status of HER2 was set as the response variable and the covariates were composed of the rest standard clinicopathologic features and gene expression levels. The testing problem considered a extremely high-dimensional setting where $n = 152$ and $p = 20601$. The calculation of our proposed testing procedure reported a significance association with a p-value < 0.001. And the same conclusion was also obtained by GC test and MCL test under the assumption that the data follows the logistic model. Hence, it showed that HER2 was associated with the clinicopathologic features and gene expression. As the dimension of the genes was extremely high, it was impractical to examine all of them in practice. Then, we proposed a conjecture that there was a representative subsets of genes to explain almost all the influence from the genes given in the first study. Specifically, the set of 50 genes from the PAM50 was considered. We divided the covariates by setting the clinicopathologic features and the set of genes as the nuisance variables, and the rest genes as the interested variables. The result of our proposed methods are given in Table 3.4, which indicates a weak association between HER2 and the rest genes. This confirms our conjecture and demonstrates that HER2 can be explained by a much smaller amount of genes together with some clinicopathologic features. It also helps to understand the powerful classification ability of PAM50. Similarly, the cases where the response variable was the status of ER or PR were also considered. The results are reported in the Table 3.4.

Table 3.4: The p-values of the proposed tests, CG test and MCL test in the global and partial testing problems.

The p-values of the following tests in the global testing problem:			
Response variable	RP	GC	MCL
HER2	<0.001	<0.001	<0.001
ER	<0.001	<0.001	<0.001
PR	0.002	<0.001	<0.001
The p-values of the following tests in the partial testing problem:			
Response variable	RP	multi-RP	S-RP
HER2	>0.999	>0.999	>0.999
ER	>0.999	>0.999	>0.999
PR	>0.999	>0.999	>0.999

3.6 Proofs of the Technical Results

3.6.1 Proof of Lemma 3.1

Proof. As $\mathbf{x}^\top\boldsymbol{\theta}$ follows the normal distribution and $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, we could make a division

$$\mathbf{x}^\top\boldsymbol{\theta} = \mathbf{u}^\top\boldsymbol{\eta} + q,$$

where $\boldsymbol{\eta} = E(\mathbf{u}\mathbf{x}^\top\boldsymbol{\theta})$ and q is independent of $\mathbf{u}$.

When $\boldsymbol{\eta} \neq \mathbf{0}$, for any $\mathbf{b} \in \mathbb{R}^k$, it could be expressed as a sum $\mathbf{b} = c\boldsymbol{\eta} + \mathbf{r}$, where $c = \boldsymbol{\eta}^\top\mathbf{b}/||\boldsymbol{\eta}||_2^2$ and $\mathbf{r} = \mathbf{b} - c\boldsymbol{\eta}$. The orthogonality between $\boldsymbol{\eta}$ and $\mathbf{r}$ implies that $\mathbf{u}^\top\boldsymbol{\eta}$ and $\mathbf{u}^\top\mathbf{r}$ are independent, which further leads to the independence between $\mathbf{u}^\top\mathbf{r}$ and the response y. Consequently, we have

$$\begin{aligned} R(\alpha, \mathbf{b}) &= E\left\{L\left(\alpha + \mathbf{u}^\top\mathbf{b}, y\right)\right\} \\ &= E\left\{L\left(\alpha + c\mathbf{u}^\top\boldsymbol{\eta} + \mathbf{u}^\top\mathbf{r}, y\right)\right\} \\ &= E\left[E\left\{L\left(\alpha + \mathbf{u}^\top\mathbf{r} + c\mathbf{u}^\top\boldsymbol{\eta}, y\right)\middle|\mathbf{u}^\top\mathbf{r}\right\}\right] \end{aligned}$$

$$\geq \min_{\tilde{\alpha}} E\left\{L(\tilde{\alpha} + c\mathbf{u}^\top \boldsymbol{\eta}, y)\right\}$$

$$\geq \min_{\tilde{\alpha}, \tilde{c}} E\left\{L(\tilde{\alpha} + \tilde{c}\mathbf{u}^\top \boldsymbol{\eta}, y)\right\}.$$

Since there is an unique solution to the minimization problem, we have

$$\mathbf{b}^\star = c^\star \boldsymbol{\eta},$$

where $c^\star$ is a constant and could be calculated through $c^\star = \boldsymbol{\eta}^\top \mathbf{b}^\star / ||\boldsymbol{\eta}||_2^2$. □

3.6.2 Proof of Theorem 3.1

Lemma 3.4. *Suppose* $(x_1, y_1), \ldots, (x_n, y_n)$ *are i.i.d. from a distribution satisfying* $E(x_i) = E(y_i) = 0$, $Var(x_i) = \sigma_x^2$, $Var(y_i) = \sigma_y^2$ *and* $E(x_i y_i) = \tau$. *Let* $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ *and* $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$. *For a symmetric* $n \times n$ *matrix* $\boldsymbol{M} = (m_{ij})$, *we have*

$$E\left(\boldsymbol{x}^\top \boldsymbol{M} \boldsymbol{y}\right) = \tau tr(\boldsymbol{M}) \quad \text{and}$$

$$Var\left(\boldsymbol{x}^\top \boldsymbol{M} \boldsymbol{y}\right) = \left(E(x_1^2 y_1^2) - 2\tau^2 - \sigma_x^2 \sigma_y^2\right) \sum_{i=1}^n m_{ii}^2 + \left(\tau^2 + \sigma_x^2 \sigma_y^2\right) tr(\boldsymbol{M}^2).$$

The property of the diagonal entries of the hat matrix is investigated by the following lemma, whose proof is deferred to Section 3.6.11.

Lemma 3.5. *Suppose* $\boldsymbol{X}$ *is an* $n \times p$ *random matrix, where each entry is i.i.d. from* $\mathcal{N}(0, 1)$. *Let* $\boldsymbol{H} = (h_{ij}) = \boldsymbol{X}\left(\boldsymbol{X}^\top \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top$. *As* $(p, n) \to \infty$ *with* $p/n \to \zeta \in (0, 1)$, *we have*

$$\max_{i=1,\ldots,n} E\left\{(h_{ii} - \zeta)^2\right\} \to 0.$$

We present a result of asymptotic normality of quadratic form that was discussed by Bhansali, Giraitis, and Kokoszka (2007).

Lemma 3.6. *Consider a general quadratic form*

$$Q_n = \boldsymbol{z}^\top \boldsymbol{A}_n \boldsymbol{z} = \sum_{i,j=1}^{n} z_i a_{ij} z_j,$$

where $\boldsymbol{A}_n = (a_{ij})$ *is a symmetric matrix, and* z_i *are i.i.d. variables satisfying* $E(z_i) = 0$ *and* $Var(z_i) = 1$. *When* $E(z_1^4) < \infty$ *and* $\frac{||\boldsymbol{A}_n||_{sp}}{||\boldsymbol{A}_n||_F} \to 0$, *then*

$$Var(Q_n)^{-1/2}\big(Q_n - E(Q_n)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Proof of Theorem 3.1. Under $\mathbf{H}_0$, the response y is independent of $\mathbf{x}$ and satisfies $y = e$, where e is the residual defined in the linear form (3.5). Let $\mathbf{M} = (m_{ij}) = \frac{\mathbf{H}}{p} - \frac{\mathbf{I}-\mathbf{H}}{n-p}$. The matrix is independent of $\mathbf{y}$ and satisfies $tr(\mathbf{M}) = 0$ and $\mathbf{M}^2 = \frac{\mathbf{H}}{p^2} + \frac{\mathbf{I}-\mathbf{H}}{(n-p)^2}$. Hence, $||\mathbf{M}||_{sp}^2/||\mathbf{M}||_F^2 \leq O(n^{-1})$ and $tr(\mathbf{M}^2) = \frac{1}{p} + \frac{1}{n-p}$. From Lemma 3.4, we have

$$E\left(\mathbf{y}^\top\mathbf{M}\mathbf{y}|\mathbf{M}\right) = 0 \quad \text{and}$$

$$Var\left(\mathbf{y}^\top\mathbf{M}\mathbf{y}|\mathbf{M}\right) = \left(E(y^4) - 3E(y^2)^2\right)\sum_{i=1}^{n} m_{ii}^2 + 2E(y^2)^2 tr(\mathbf{M}^2).$$

Under the condition $n\sum_{i=1}^{n} m_{ii}^2 = o_p(1)$, which will be verified subsequently, Lemma 3.6 demonstrates that

$$\frac{\mathbf{y}^\top\mathbf{M}\mathbf{y}}{\sigma^2\sqrt{2/n\zeta(1-\zeta)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{3.23}$$

where $\sigma^2 = Var(e) = Var(y)$.

Let $G_n = n\sum_{i=1}^{n} m_{ii}^2$. It can be written as

$$G_n = n\sum_{i=1}^{n}\left\{\frac{h_{ii}}{p} - \frac{1-h_{ii}}{n-p}\right\}^2 = \frac{n^3}{p^2(n-p)^2}\sum_{i=1}^{n}\left\{h_{ii} - \frac{p}{n}\right\}^2,$$

where h_{ij} denotes the ij-th entry of $\mathbf{H}$. Based on the assumption $p/n \to \zeta$, Lemma 3.5 implies that $E(G_n) = o(1)$. Therefore, we have $G_n = o_p(1)$ by Markov's inequality.

To study $\mathbf{y}^\top(\mathbf{I}-\mathbf{H})\mathbf{y}$, the calculation shows

$$E\left(\mathbf{y}^\top(\mathbf{I}-\mathbf{H})\mathbf{y}\right)=\sigma^2(n-p),\ Var\left(\mathbf{y}^\top(\mathbf{I}-\mathbf{H})\mathbf{y}\right)\leq 3E(y^4)(n-p).$$

From Markov's inequality,

$$\frac{\mathbf{y}^\top(\mathbf{I}-\mathbf{H})\mathbf{y}}{n-p}=\sigma^2+o_p(1). \tag{3.24}$$

Consequently, (3.23) and (3.24) lead to

$$\frac{F_n-1}{\sqrt{2/n\zeta(1-\zeta)}}\xrightarrow{\mathcal{D}}\mathcal{N}(0,1),$$

which completes the proof. □

3.6.3 Proof of Theorem 3.2

First, we introduce a decomposition of the projection matrix in the following lemma, whose proof is deferred to Section 3.6.11.

Lemma 3.7. *Suppose $\boldsymbol{X}$ is an $n\times p$ random matrix, where each entry is i.i.d. from $\mathcal{N}(0,1)$. Consider a decomposition $\boldsymbol{X}=(\boldsymbol{w},\boldsymbol{G})$, where $\boldsymbol{w}\in\mathbb{R}^n$ and $\boldsymbol{G}\in\mathbb{R}^{n\times(p-1)}$. Let $\boldsymbol{H}=\boldsymbol{X}\big(\boldsymbol{X}^\top\boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top$. We have*

$$\boldsymbol{H}=\boldsymbol{H_w}+(\boldsymbol{I}-\boldsymbol{H_w})[\boldsymbol{H_G}+\frac{\boldsymbol{H_G}\boldsymbol{H_w}\boldsymbol{H_G}}{1-tr(\boldsymbol{H_G}\boldsymbol{H_w})}](\boldsymbol{I}-\boldsymbol{H_w}),$$

where $\boldsymbol{H_w}=\boldsymbol{w}(\boldsymbol{w}^\top\boldsymbol{w})^{-1}\boldsymbol{w}^\top$ and $\boldsymbol{H_G}=\boldsymbol{G}(\boldsymbol{G}^\top\boldsymbol{G})^{-1}\boldsymbol{G}^\top$. Suppose $\boldsymbol{e}=(e_1,\ldots,e_n)^\top$ is a random vector composed of i.i.d. entries, satisfying $E(e_1)=0$ and $E(e_1^4)<\infty$. In addition, $\boldsymbol{e}$ is independent of $\boldsymbol{G}$. For $i\neq j$, e_i is independent of w_j, otherwise, $E(e_iw_i)=0$, where w_i denotes the i-th entry of $\boldsymbol{w}$. As $(p,n)\to\infty$ with $p/n\to\zeta\in(0,1)$, for any $\delta>0$, we have

$$\boldsymbol{e}^\top\boldsymbol{H}\boldsymbol{e}=\boldsymbol{e}^\top\boldsymbol{H_G}\boldsymbol{e}+Re,$$

where $Re=o_p(n^\delta)$.

Proof of Theorem 3.2. From the linear form (3.5) and the condition $E(y) = 0$, the model can be written as

$$\mathbf{y} = c_0 \mathbf{X}\boldsymbol{\theta} + \mathbf{e},$$

where $\mathbf{e} = (e_1, \dots, e_n)^\top$. The direct calculation shows that

$$\begin{aligned}
&\mathbf{y}^\top \mathbf{H}\mathbf{y} = c_0^2 \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\theta} + 2c_0 \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{e} + \mathbf{e}^\top \mathbf{H}\mathbf{e} \text{ and} \\
&\mathbf{y}^\top (\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{e}^\top (\mathbf{I} - \mathbf{H})\mathbf{e}.
\end{aligned}$$

First, we investigate the quadratic form $\mathbf{e}^\top \mathbf{H}\mathbf{e}$. Let $\mathbf{w} = \frac{\mathbf{X}\boldsymbol{\theta}}{||\boldsymbol{\theta}||_2}$ and $\mathbf{G} = \mathbf{X}(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_{p-1})$, where $\{\frac{\boldsymbol{\theta}}{||\boldsymbol{\theta}||_2}, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_{p-1}\}$ forms an orthonormal basis of $\mathbb{R}^p$. The normality of $\mathbf{X}$ implies that $\mathbf{w}$ is independent of $\mathbf{G}$, which further indicates that $\mathbf{e}$ and $\mathbf{G}$ are independent, since e_i is determined by $\mathbf{x}_i^\top \boldsymbol{\theta}$ and ϵ_i. When $\mathbf{X}$ is replaced by $\mathbf{X}\mathbf{O}$, where $\mathbf{O}$ is a $p \times p$ orthogonal matrix, $\mathbf{H}$ stays the same. Based on Lemma 3.7, we have

$$\mathbf{e}^\top \mathbf{H}\mathbf{e} = \mathbf{e}^\top \mathbf{H}_\mathbf{G} \mathbf{e} + Re,$$

where $Re = o_p(n^\delta)$ with sufficiently small $\delta > 0$. Define $\mathbf{M} = \frac{\mathbf{H}_\mathbf{G}}{p-1} - \frac{\mathbf{I} - \mathbf{H}_\mathbf{G}}{n-p+1}$. The matrix is independent of $\mathbf{e}$. With a similar method in 3.6.2, we have

$$\frac{\mathbf{e}^\top \mathbf{M}\mathbf{e}}{\sigma^2 \sqrt{2/n\zeta(1-\zeta)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

and

$$\frac{\mathbf{e}^\top (\mathbf{I} - \mathbf{H}_\mathbf{G})\mathbf{e}}{n - p} = \sigma^2 + o_p(1),$$

where $\sigma^2 = Var(e_1) = Var(y) - c_0^2 ||\boldsymbol{\theta}||_2^2$.

To investigate $c_0^2 \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\theta} + 2c_0 \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{e}$, the calculation shows

$$\begin{aligned}
&E(c_0^2 \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\theta}) = nc_0^2 ||\boldsymbol{\theta}||_2^2, \quad Var(c_0^2 \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\theta}) = 2nc_0^4 ||\boldsymbol{\theta}||_2^4, \\
&E(c_0 \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{e}) = 0, \quad Var(c_0 \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{e}) \le 4nc_0^2 ||\boldsymbol{\theta}||_2^2 E(e_1^4)^{1/2}.
\end{aligned}$$

Under the condition $c_0^2||\boldsymbol{\theta}||_2^2 = o(1)$, Markov's inequality implies that

$$\frac{1}{\sqrt{n}} c_0^2 \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\theta} = \sqrt{n} c_0^2 ||\boldsymbol{\theta}||_2^2 + o_p(1), \quad \frac{1}{\sqrt{n}} c_0 \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{e} = o_p(1).$$

Consequently,

$$\begin{aligned}
\frac{F_n - 1}{\sqrt{2/n\zeta(1-\zeta)}} &= \frac{\sqrt{\frac{n\zeta(1-\zeta)}{2}} \left\{ \frac{c_0^2 \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\theta}}{p} + \frac{2c_0 \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{e}}{p} + \mathbf{e}^\top \left(\frac{\mathbf{H}}{p} - \frac{\mathbf{I}-\mathbf{H}}{n-p} \right) \mathbf{e} \right\}}{\frac{\mathbf{e}^\top (\mathbf{I}-\mathbf{H}) \mathbf{e}}{n-p}} \\
&= \frac{\sqrt{\frac{n(1-\zeta)}{2\zeta}} c_0^2 ||\boldsymbol{\theta}||_2^2 + \sqrt{\frac{n\zeta(1-\zeta)}{2}} \mathbf{e}^\top \mathbf{M} \mathbf{e} + o_p(1)}{\sigma^2 + o_p(1)}.
\end{aligned}$$

It shows that the power function satisfies

$$\begin{aligned}
\Psi_n(\boldsymbol{\theta}) &= P\left(\frac{F_n - 1}{\sqrt{2/n\zeta(1-\zeta)}} > z_\alpha\right) \\
&= \Phi\left(-z_\alpha + \sqrt{\frac{n(1-\zeta)}{2\zeta}} \frac{c_0^2 ||\boldsymbol{\theta}||_2^2}{\sigma^2}\right) + o(1),
\end{aligned}$$

which completes the proof. □

3.6.4 Proof of Theorem 3.3

Proof. First, we prove the existences of $\mathbf{S}_\mathbf{B}$ and $\mathbf{S}_{\mathbf{B}^\perp}$. For the matrix $\mathbf{B}$, based on the Gram-Schmidt process, we can derive an orthonormal basis of $\mathbb{R}^{p_2}$, denoted as $\{\mathbf{b}_1, \ldots, \mathbf{b}_{p_2}\}$, such that $Span\{\mathbf{B}\} = Span\{\mathbf{b}_1, \ldots, \mathbf{b}_b\}$. Let $\mathbf{S}_\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_b)$ and $\mathbf{S}_{\mathbf{B}^\perp} = (\mathbf{b}_{b+1}, \ldots, \mathbf{b}_{p_2})$. The required conditions are satisfied by the matrices. Therefore, the existences of $\mathbf{S}_\mathbf{B}$ and $\mathbf{S}_{\mathbf{B}^\perp}$ are shown.

Under $\mathbf{H}_{0,\mathbf{B}}$, according to the linear form (3.9), the model can be written as

$$\mathbf{y} = c_0 \mathbf{W} \boldsymbol{\gamma} + \mathbf{e},$$

where $\mathbf{e} = (e_1, \ldots, e_n)^\top$. Then,

$$\mathbf{y}^\top (\mathbf{H} - \mathbf{H}_\mathbf{W}) \mathbf{y} = \mathbf{e}^\top (\mathbf{H} - \mathbf{H}_\mathbf{W}) \mathbf{e}, \quad \mathbf{y}^\top (\mathbf{I} - \mathbf{H}) \mathbf{y} = \mathbf{e}^\top (\mathbf{I} - \mathbf{H}) \mathbf{e}.$$

Hence, we have

$$F_{n,p2} = \frac{\mathbf{e}^\top(\mathbf{H} - \mathbf{H_W})\mathbf{e}/b}{\mathbf{e}^\top(\mathbf{I} - \mathbf{H})\mathbf{e}/(n-p)}. \tag{3.25}$$

First, we investigate $\mathbf{e}^\top\mathbf{He}$ and $\mathbf{e}^\top\mathbf{H_W e}$. Let $\widetilde{\mathbf{X}} = \mathbf{X}\boldsymbol{\Sigma}^{-1/2}$ and $\widetilde{\mathbf{W}} = \mathbf{W}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1/2}$, where $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{\mathbf{w}}$ are the covariance matrices of $\mathbf{x}$ and $\mathbf{w}$, respectively. Then,

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top = \widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^\top \quad \text{and}$$

$$\mathbf{H_W} = \mathbf{W}(\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top = \widetilde{\mathbf{W}}(\widetilde{\mathbf{W}}^\top\widetilde{\mathbf{W}})^{-1}\widetilde{\mathbf{W}}^\top,$$

where the entries of $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{W}}$ are i.i.d. from $\mathcal{N}(0,1)$, respectively. Let $\widetilde{\boldsymbol{\gamma}} = \boldsymbol{\Sigma}_{\mathbf{w}}^{1/2}\boldsymbol{\gamma}$ and $\widetilde{\boldsymbol{\xi}} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\xi}$, where $\boldsymbol{\xi} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top(\mathbf{I} - \mathbf{H_B}))^\top$. Then, $\widetilde{\mathbf{X}}\widetilde{\boldsymbol{\xi}} = \widetilde{\mathbf{W}}\widetilde{\boldsymbol{\gamma}}$, since $\mathbf{X}\boldsymbol{\xi} = \mathbf{W}\boldsymbol{\gamma}$. Based on this relationship, we make decomposition to the projection matrices. For the matrix $\mathbf{H}$, let $\mathbf{v} = \frac{\widetilde{\mathbf{X}}\widetilde{\boldsymbol{\xi}}}{||\widetilde{\boldsymbol{\xi}}||_2}$ and $\mathbf{G} = \widetilde{\mathbf{X}}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{p-1})$, where $\{\frac{\widetilde{\boldsymbol{\xi}}}{||\widetilde{\boldsymbol{\xi}}||_2}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{p-1}\}$ forms an orthonormal basis of $\mathbb{R}^p$. The orthogonality implies that $\mathbf{v}$ and $\mathbf{G}$ are independent, which further shows that $\mathbf{e}$ is independent of $\mathbf{G}$, since e_i is determined by $\mathbf{w}_i^\top\boldsymbol{\gamma}$ and ϵ_i. From Lemma 3.7, we obtain

$$\mathbf{e}^\top\mathbf{He} = \mathbf{e}^\top\mathbf{H_G e} + Re_1, \tag{3.26}$$

where $\mathbf{H_G} = \mathbf{G}(\mathbf{G}^\top\mathbf{G})^{-1}\mathbf{G}^\top$, and $Re_1 = o_p(n^{\delta_1})$ with sufficiently small $\delta_1 > 0$. For the matrix $\mathbf{H_W}$, let $\mathbf{F} = \widetilde{\mathbf{W}}(\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{p-b-1})$, where $\{\frac{\widetilde{\boldsymbol{\gamma}}}{||\widetilde{\boldsymbol{\gamma}}||_2}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{p-b-1}\}$ forms an orthonormal basis of $\mathbb{R}^{p-b}$. According to $\widetilde{\mathbf{X}}\widetilde{\boldsymbol{\xi}} = \widetilde{\mathbf{W}}\widetilde{\boldsymbol{\gamma}}$, there is a constant c such that $\frac{\widetilde{\mathbf{W}}\widetilde{\boldsymbol{\gamma}}}{||\widetilde{\boldsymbol{\gamma}}||_2} = c\mathbf{v}$. The orthogonality implies that $\mathbf{v}$ and $\mathbf{F}$ are independent, which then shows that $\mathbf{e}$ is independent of $\mathbf{F}$. From Lemma 3.7, we obtain

$$\mathbf{e}^\top\mathbf{H_W e} = \mathbf{e}^\top\mathbf{H_F e} + Re_2, \tag{3.27}$$

where $\mathbf{H_F} = \mathbf{F}(\mathbf{F}^\top\mathbf{F})^{-1}\mathbf{F}^\top$ and $Re_2 = o_p(n^{\delta_2})$ with sufficiently small $\delta_2 > 0$. Let $\sigma^2 = Var(e_1)$. The calculation yields

$$E\{\mathbf{e}^\top(\mathbf{I} - \mathbf{H_G})\mathbf{e}\} = \sigma^2(n+1-p), \; Var\{\mathbf{e}^\top(\mathbf{I} - \mathbf{H_G})\mathbf{e}\} \le 3E(e_1^4)(n+1-p).$$

Markov's inequality implies

$$\frac{\mathbf{e}^\top(\mathbf{I}-\mathbf{H_G})\mathbf{e}}{n+1-p} = \sigma^2 + o_p(1).$$

Combining (3.25), (3.26) and (3.27), we obtain

$$\frac{F_{n,p_2}-1}{\sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}} = \frac{\sqrt{\frac{n\zeta_1(1-\zeta)}{2(1-\zeta+\zeta_1)}}\mathbf{e}^\top\mathbf{M}\mathbf{e} + o_p(1)}{\sigma^2+o_p(1)}, \tag{3.28}$$

where $\mathbf{M} = (m_{ij}) = \frac{\mathbf{H_G}-\mathbf{H_F}}{b} - \frac{\mathbf{I}-\mathbf{H_G}}{n+1-p}$.

To study $\mathbf{e}^\top\mathbf{M}\mathbf{e}$, we follow a similar method as that given in Section 3.6.2. First, note that $Span\{\mathbf{F}\} \subseteq Span\{\mathbf{G}\}$. The property of the projection matrix shows that

$$\mathbf{H_G}\mathbf{H_F} = \mathbf{H_F}\mathbf{H_G} = \mathbf{H_F},$$

and $tr(\mathbf{M}) = 0$. Then, $\mathbf{M}^\top\mathbf{M} = \frac{\mathbf{H_G}-\mathbf{H_F}}{b^2} + \frac{\mathbf{I}-\mathbf{H_G}}{(n+1-p)^2}$, leading to $||\mathbf{M}||_{sp}^2/||\mathbf{M}||_F^2 \leq O(n^{-1})$ and

$$Var(\mathbf{e}^\top\mathbf{M}\mathbf{e}|\mathbf{M}) = (E(e_1^4)-3\sigma^4)\sum_{i=1}^n m_{ii}^2 + 2\sigma^4(\frac{1}{b}+\frac{1}{n+1-p}).$$

We show that $n\sum_{i=1}^n m_{ii}^2 = o_p(1)$. Let $(\mathbf{H_G})_{ii}$ and $(\mathbf{H_F})_{ii}$ denote the i-th diagonal entries of $\mathbf{H_G}$ and $\mathbf{H_F}$, respectively. Then,

$$n\sum_{i=1}^n m_{ii}^2 \leq \frac{2h_1^2}{n}\sum_{i=1}^n\left\{(\mathbf{H_G})_{ii}-\frac{p-1}{n}\right\}^2 + \frac{2h_2^2}{n}\sum_{i=1}^n\left\{(\mathbf{H_F})_{ii}-\frac{p-b-1}{n}\right\}^2,$$

where $h_1 = \frac{1-\frac{p}{n}+\frac{b}{n}+\frac{1}{n}}{\frac{b}{n}(1-\frac{p}{n}+\frac{1}{n})} \to \frac{1+\zeta_1-\zeta}{\zeta_1(1-\zeta)}$ and $h_2 = \frac{n}{b} \to \frac{1}{\zeta_1}$ as $n\to\infty$. From the definitions of $\mathbf{G}$ and $\mathbf{F}$, Lemma 3.5 demonstrates that

$$E\left\{(\mathbf{H_G})_{ii}-\frac{p-1}{n}\right\}^2 \to 0,\ E\left\{(\mathbf{H_F})_{ii}-\frac{p-b-1}{n}\right\}^2 \to 0.$$

It leads to $E(n\sum_{i=1}^n m_{ii}^2) = o(1)$. Hence, $n\sum_{i=1}^n m_{ii}^2 = o_p(1)$ by Markov's inequality. From Lemma 3.6, we obtain

$$\frac{\mathbf{e}^\top \mathbf{M}\mathbf{e}}{\sigma^2\sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Taking this into (3.28), we have

$$\frac{F_{n,p_2}-1}{\sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

which completes the proof. □

3.6.5 Proof of Theorem 3.4

Proof. Let

$$\boldsymbol{\psi} = \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{D}^\top \begin{pmatrix} \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{22} \end{pmatrix} \mathbf{H}_{\mathbf{B}}\boldsymbol{\theta}_2, \quad \mathbf{D} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{\mathbf{B}^\perp} \end{pmatrix},$$

where $\boldsymbol{\Sigma}_{\mathbf{w}}$ is the covariance matrix of $\mathbf{w}$. Then, $\mathbf{w} = \mathbf{D}^\top\mathbf{x}$ and $\boldsymbol{\Sigma}_{\mathbf{w}} = \mathbf{D}^\top\boldsymbol{\Sigma}\mathbf{D}$. From the normality assumption of $\mathbf{x}$, we derive a decomposition

$$\mathbf{x}_2^\top \mathbf{H}_{\mathbf{B}}\boldsymbol{\theta}_2 = \mathbf{w}^\top\boldsymbol{\psi} + q,$$

where q is independent of $\mathbf{w}$. Let $\tau^2 = Var(q)$. It satisfies

$$\tau^2 = \boldsymbol{\theta}_2^\top \mathbf{H}_{\mathbf{B}} \left[\boldsymbol{\Sigma}_{22} - (\boldsymbol{\Sigma}_{21}\ \boldsymbol{\Sigma}_{22})\,\mathbf{D}(\mathbf{D}^\top\boldsymbol{\Sigma}\mathbf{D})^{-1}\mathbf{D}^\top \begin{pmatrix} \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{22} \end{pmatrix}\right] \mathbf{H}_{\mathbf{B}}\boldsymbol{\theta}_2.$$

Based on the linear form (3.9), the model can be written as

$$y = c_0\mathbf{w}^\top(\boldsymbol{\gamma}+\boldsymbol{\psi}) + c_0 q + e.$$

Then, we obtain

$$\mathbf{y}^\top(\mathbf{H}-\mathbf{H}_{\mathbf{W}})\mathbf{y} = c_0^2\mathbf{q}^\top(\mathbf{I}-\mathbf{H}_{\mathbf{W}})\mathbf{q} + 2c_0\mathbf{q}^\top(\mathbf{I}-\mathbf{H}_{\mathbf{W}})\mathbf{e} + \mathbf{e}^\top(\mathbf{H}-\mathbf{H}_{\mathbf{W}})\mathbf{e} \quad \text{and}$$

$$\mathbf{y}^\top(\mathbf{I}-\mathbf{H})\mathbf{y} = \mathbf{e}^\top(\mathbf{I}-\mathbf{H})\mathbf{e},$$

where $\mathbf{q} = (q_1, \ldots, q_n)^\top$ and $\mathbf{e} = (e_1, \ldots, e_n)^\top$.

To investigate $c_0^2 \mathbf{q}^\top (\mathbf{I} - \mathbf{H_W})\mathbf{q}$, Lemma 3.4 indicates

$$E\left\{\mathbf{q}^\top (\mathbf{I} - \mathbf{H_W})\mathbf{q}\right\} = \tau^2(n-p+b),\ Var\left\{\mathbf{q}^\top (\mathbf{I} - \mathbf{H_W})\mathbf{q}\right\} = 2\tau^4(n-p+b).$$

Under the condition $c_0^2\tau^2 = o(1)$, Markov's inequality shows

$$\frac{\sqrt{n} c_0^2 \mathbf{q}^\top (\mathbf{I} - \mathbf{H_W})\mathbf{q}}{b} = \frac{\sqrt{n}(1-\zeta+\zeta_1)}{\zeta_1} c_0^2\tau^2 + o_p(1). \tag{3.29}$$

Next, we study $\mathbf{e}^\top(\mathbf{H} - \mathbf{H_W})\mathbf{e}$ and $\mathbf{e}^\top(\mathbf{I} - \mathbf{H})\mathbf{e}$. Let $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{W}}$ have the same definitions as that in Section 3.6.4 and define $\widetilde{\boldsymbol{\theta}} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\theta}$. We introduce another matrix $\mathbf{R} = (\mathbf{W}, \mathbf{q})$ with its projection matrix defined as $\mathbf{H_R} = \mathbf{R}(\mathbf{R}^\top\mathbf{R})^{-1}\mathbf{R}^\top$. Let $\mathbf{v} = \mathbf{X}\boldsymbol{\theta}$. Then, $\mathbf{v} \in Span\{\mathbf{R}\} \subseteq Span\{\mathbf{X}\}$. Let $\mathbf{F} = \widetilde{\mathbf{X}}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{p-b})$ and $\mathbf{G} = \widetilde{\mathbf{X}}(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{p-1})$, where $\boldsymbol{\xi}_i$ are selected to make $\{\frac{\widetilde{\boldsymbol{\theta}}}{||\widetilde{\boldsymbol{\theta}}||_2}, \boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{p-1}\}$ form an orthonormal basis for $\mathbb{R}^p$ and $Span\{\mathbf{R}\} = Span\{\mathbf{F}, \mathbf{v}\}$. It indicates that $\mathbf{v}$ is independent of $\mathbf{F}$ and $\mathbf{G}$. Then, $\mathbf{e}$ is independent of $\mathbf{F}$ and $\mathbf{G}$, since e_i is determined by ϵ_i and $\mathbf{x}_i^\top\boldsymbol{\theta}$. From Lemma 3.7, we have

$$\mathbf{e}^\top\mathbf{H_R}\mathbf{e} = \mathbf{e}^\top\mathbf{H_F}\mathbf{e} + Re_1, \quad \mathbf{e}^\top\mathbf{H}\mathbf{e} = \mathbf{e}^\top\mathbf{H_G}\mathbf{e} + Re_2, \tag{3.30}$$

where $Re_1 = o_p(n^{\delta_1})$ and $Re_2 = o_p(n^{\delta_2})$ for sufficiently small $\delta_1, \delta_2 > 0$. Since $Span\{\mathbf{F}\} \subseteq Span\{\mathbf{G}\}$,

$$\mathbf{H_G}\mathbf{H_F} = \mathbf{H_F}\mathbf{H_G} = \mathbf{H_F}.$$

Let $\mathbf{M} = \frac{\mathbf{H_G}-\mathbf{H_F}}{b-1} - \frac{\mathbf{I}-\mathbf{H_G}}{n+1-p}$. The matrix is independent of $\mathbf{e}$ and satisfies $tr(\mathbf{M}) = 0$ and $\mathbf{M}^\top\mathbf{M} = \frac{\mathbf{H_G}-\mathbf{H_F}}{(b-1)^2} + \frac{\mathbf{I}-\mathbf{H_G}}{(n+1-p)^2}$, leading to

$$\frac{||\mathbf{M}||_{sp}^2}{||\mathbf{M}||_F^2} \leq O(n^{-1}), \quad n||\mathbf{M}||_F^2 = \frac{1}{\zeta_1} + \frac{1}{1-\zeta} + o(1).$$

With a similar method in Section 3.6.4, we have

$$\frac{\mathbf{e}^\top\mathbf{M}\mathbf{e}}{\sigma^2\sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{3.31}$$

where $\sigma^2 = Var(e) = Var(y) - c_0^2\boldsymbol{\theta}^\top\boldsymbol{\Sigma}\boldsymbol{\theta}$.

We then study the relationship between $\mathbf{e}^\top\mathbf{H_R}\mathbf{e}$ and $\mathbf{e}^\top\mathbf{H_W}\mathbf{e}$, since the test statistic contains $\mathbf{H_W}$ rather than $\mathbf{H_R}$. From the definition of $\mathbf{R}$, the matrix $\mathbf{H_R}$ could be divided based on $\mathbf{q}$ and $\mathbf{H_W}$. Under the condition that $\tau^{-1}\mathbf{e}^\top\mathbf{H_W}\mathbf{q} = o_p(n^{\gamma+0.5})$ for any $\gamma > 0$, which will be verified subsequently, the proof of Lemma 3.7 shows

$$\mathbf{e}^\top\mathbf{H_R}\mathbf{e} = \mathbf{e}^\top\mathbf{H_W}\mathbf{e} + Re_3, \tag{3.32}$$

with $Re_3 = o_p(n^{\delta_3})$ for any $\delta_3 > 0$. Note that $\mathbf{e}$ might not be independent of $\mathbf{W}$. To verify the condition, let $\boldsymbol{\ell} = \boldsymbol{\Sigma}_{\mathbf{w}}^{1/2}(\boldsymbol{\gamma} + \boldsymbol{\psi})$ and $\mathbf{E} = \widetilde{\mathbf{W}}(\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{p-b-1})$, where $\boldsymbol{\gamma}_i$ are selected to make $\{\frac{\boldsymbol{\ell}}{||\boldsymbol{\ell}||_2}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{p-b-1}\}$ form an orthonormal basis for $\mathbb{R}^{p-b}$. Then, $\mathbf{e}$ is independent of $\mathbf{E}$, since $\widetilde{\mathbf{W}}\boldsymbol{\ell}$ and $\mathbf{E}$ are independent. From Lemma 3.7, it gives a decomposition, $\mathbf{H_W} = \mathbf{H_E} + \mathbf{K}$, where $\mathbf{K}$ satisfies $\mathbf{e}^\top\mathbf{K}\mathbf{e} = o_p(n^{\delta_4})$ for any $\delta_4 > 0$. The calculation shows

$$E(\mathbf{e}^\top\mathbf{H_E}\mathbf{q}) = 0,\ Var(\mathbf{e}^\top\mathbf{H_E}\mathbf{q}) \le 4(p-b-1)\tau^2E(e^4)^{1/2}$$

and

$$|\mathbf{e}^\top\mathbf{K}\mathbf{q}| \le \sqrt{\mathbf{e}^\top\mathbf{K}\mathbf{e}}\sqrt{\mathbf{q}^\top\mathbf{q}} = \tau o_p(n^{\delta_4+0.5}).$$

Hence, $\tau^{-1}\mathbf{e}^\top\mathbf{H_W}\mathbf{q} = o_p(n^{\gamma+0.5})$ for any $\gamma > 0$. This together with the condition $c_0^2\tau^2 = o(1)$ implies

$$\frac{1}{\sqrt{n}}c_0\mathbf{q}^\top(\mathbf{I} - \mathbf{H_W})\mathbf{e} = o_p(1). \tag{3.33}$$

Finally, we combine the above results to derive the asymptotic local power function. For any idempotent matrix $\mathbf{A}$, Lemma 3.4 shows

$$E(\mathbf{e}^\top\mathbf{A}\mathbf{e}) = \sigma^2 tr(\mathbf{A}),\ Var(\mathbf{e}^\top\mathbf{A}\mathbf{e}) \le 3E(e^4)tr(\mathbf{A}).$$

This together with Markov's inequality and the derived independence indicates

$$\frac{\mathbf{e}^\top(\mathbf{H_G} - \mathbf{H_F})\mathbf{e}}{b-1} = \sigma^2 + o_p(1) \ \text{ and } \ \frac{\mathbf{e}^\top(\mathbf{I} - \mathbf{H_G})\mathbf{e}}{n-p+1} = \sigma^2 + o_p(1).$$

Combining it with (3.29), (3.30), (3.31), (3.32) and (3.33), we have

$$\frac{F_{n,p_2}-1}{\sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}}$$

$$= \frac{\sqrt{\frac{n\zeta_1(1-\zeta)}{2(1-\zeta+\zeta_1)}}\left\{\frac{c_0^2\mathbf{q}^\top(\mathbf{I}-\mathbf{H_W})\mathbf{q}}{b}+\frac{2c_0\mathbf{q}^\top(\mathbf{I}-\mathbf{H_W})\mathbf{e}}{b}+\mathbf{e}^\top(\frac{\mathbf{H}-\mathbf{H_W}}{b}-\frac{\mathbf{I}-\mathbf{H}}{n-p})\mathbf{e}\right\}}{\frac{\mathbf{e}^\top(\mathbf{I}-\mathbf{H})\mathbf{e}}{n-p}}$$

$$= \frac{\sqrt{\frac{n(1-\zeta)(1-\zeta+\zeta_1)}{2\zeta_1}}c_0^2\tau^2+\sqrt{\frac{n\zeta_1(1-\zeta)}{2(1-\zeta+\rho_1)}}\mathbf{e}^\top\mathbf{M}\mathbf{e}+o_p(1)}{\sigma^2+o_p(1)}.$$

Therefore, the asymptotic power function satisfies

$$\Psi_n(\boldsymbol{\theta}_2;\mathbf{B}) = P(\frac{F_{n,p_2}-1}{\sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}} > z_\alpha)$$

$$= \Phi(-z_\alpha+\sqrt{\frac{n(1-\zeta)(1-\zeta+\zeta_1)}{2\zeta_1}}\frac{c_0^2\tau^2}{\sigma^2})+o(1),$$

which completes the proof. □

3.6.6 Proof of Theorem 3.5

An upper-bound and a lower-bound on the extreme eigenvalues of Wishart matrices were given in Davidson and Szarek (2001, Theorem 2.13) .

Lemma 3.8. *For $k \le p$, let $\boldsymbol{P}_k \in \mathbb{R}^{p\times k}$ be a random matrix with i.i.d. $\mathcal{N}(0,1)$ entries. Then, for all $t \ge 0$, we have*

$$P\left[\lambda_{\max}(\frac{1}{p}\boldsymbol{P}_k^\top\boldsymbol{P}_k) \ge (1+\sqrt{k/p}+t)^2\right] \le \exp(-pt^2/2) \quad \textit{and}$$

$$P\left[\lambda_{\min}(\frac{1}{p}\boldsymbol{P}_k^\top\boldsymbol{P}_k) \le (1-\sqrt{k/p}-t)^2\right] \le \exp(-pt^2/2).$$

Proof of Theorem 3.5. Under $\mathbf{H}_0$, the model becomes

$$y = f(\epsilon). \tag{3.34}$$

Consider a linear regression model

$$y = \alpha + \mathbf{u}^\top \boldsymbol{\eta} + e, \tag{3.35}$$

where α is an intercept, $\mathbf{u} = \mathbf{P}_k^\top \mathbf{x}$, $\boldsymbol{\eta}$ is a vector of coefficients and e is a error term independent of $\mathbf{x}$. When $\alpha = E(f(\epsilon))$, $\boldsymbol{\eta} = \mathbf{0}$, and $e = f(\epsilon) - E(f(\epsilon))$, $(\mathbf{x}, y)$ from model (3.34) has the same distribution as that of $(\mathbf{x}, y)$ from model (3.35). Hence, it suffices to study the test statistic under (3.35). Let $\mathbf{M} = (m_{ij}) = \frac{\mathbf{H}_k}{k} - \frac{\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k}{n-k-1}$. Under the condition $n\sum_{i=1}^n m_{ii}^2 = o_p(1)$, which will be verified subsequently, with a similar proof method in Section 3.6.2, we can derive

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

Therefore, the asymptotic normality of the test statistic is demonstrated.

To verify the condition $n\sum_{i=1}^n m_{ii}^2 = o_p(1)$, it is sufficient to prove

$$\frac{1}{n}\sum_{i=1}^n (h_{ii} - \rho)^2 = o_p(1),$$

where h_{ii} is the i-th diagonal entry of $\mathbf{H}_k$. For simplicity of notation, we denote $\mathbf{H}_k$ by $\mathbf{H}$ in the following. Based on Assumption H1, the matrix $\mathbf{H}$ can be denoted as

$$\mathbf{H} = (\mathbf{I} - \mathbf{P}_1)\mathbf{Z}\mathbf{A}\left(\mathbf{A}^\top \mathbf{Z}^\top (\mathbf{I} - \mathbf{P}_1)\mathbf{Z}\mathbf{A}\right)^{-1} \mathbf{A}^\top \mathbf{Z}^\top (\mathbf{I} - \mathbf{P}_1),$$

where $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)^\top$ and $\mathbf{A}$ belongs to the Stiefel manifold $\mathcal{V}_k(\mathbb{R}^m) = \{\mathbf{A} \in \mathbb{R}^{m\times k} : \mathbf{A}^\top \mathbf{A} = \mathbf{I}\}$. Considering the randomness of $\mathbf{P}_k$, to cover general cases, we assume that the matrix $\mathbf{A}$ is uniformly distributed on $\mathcal{V}_k(\mathbb{R}^m)$ and is independent of $\mathbf{Z}$. Let $\mathbf{U}\boldsymbol{\Lambda}\mathbf{O}^\top$ be the SVD of $\mathbf{Z}$, where $\mathbf{U}$ is an $n \times n$ orthogonal matrix, $\mathbf{O}$ is an $m \times m$ orthogonal matrix, and $\boldsymbol{\Lambda} = (\mathbf{D}, \mathbf{0})$ with $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$. Let $\mathbf{O}_n$ be the matrix consisting of first n columns of $\mathbf{O}$, then $\mathbf{Z}$ can be denoted as $\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{O}_n^\top$.

The diagonal entries of $\frac{1}{m}\mathbf{D}^2$ are the eigenvalues of $\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top$. The calculation shows

$$E\left[\max_{i=1,\dots,n}\left(\frac{d_i^2}{m}-1\right)^2\right] \le E\left[tr\left\{\left(\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top - \mathbf{I}\right)^2\right\}\right] = O(n^2m^{-1}).$$

Therefore, from Markov's inequality, for any $t>0$, we have

$$P\left\{\max_{i=1,\dots,n}\left(\frac{d_i}{\sqrt{m}}-1\right)^2 > t\right\} \le P\left\{\max_{i=1,\dots,n}\left(\frac{d_i^2}{m}-1\right)^2 > t\right\} \le O(n^2m^{-1}t^{-1}). \tag{3.36}$$

Let $\mathbf{V} = (\mathbf{I}-\mathbf{P}_1)\,\mathbf{U}\mathbf{O}_n^\top\mathbf{A}$ and $\tilde{\mathbf{Z}} = (\mathbf{I}-\mathbf{P}_1)\,\mathbf{U}\frac{\mathbf{D}}{\sqrt{m}}\mathbf{O}_n^\top\mathbf{A}$. The hat matrix for $\tilde{\mathbf{Z}}$ and $\mathbf{V}$ can be denoted as

$$\mathbf{H} = (h_{ij}) = \tilde{\mathbf{Z}}\left(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}}\right)^{-1}\tilde{\mathbf{Z}}^\top \ \text{ and } \ \mathbf{S} = (s_{ij}) = \mathbf{V}\left(\mathbf{V}^\top\mathbf{V}\right)^{-1}\mathbf{V}^\top,$$

where $\mathbf{H}$ is the target matrix. We will show that h_{ii} and s_{ii} are close. Let $\mathbf{e}_i$ denote the unit vector with 1 in the i-th coordinate. Define $\hat{\boldsymbol{\gamma}}_i^{ls} = \left(\mathbf{V}^\top\mathbf{V}\right)^{-1}\mathbf{V}^\top\mathbf{e}_i$ and $\hat{\boldsymbol{\eta}}_i^{ls} = \left(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}}\right)^{-1}\tilde{\mathbf{Z}}^\top\mathbf{e}_i$. Then, $\hat{\boldsymbol{\gamma}}_i^{ls}$ and $\hat{\boldsymbol{\eta}}_i^{ls}$ are the solutions to the least square problems $\min_{\boldsymbol{\gamma}\in\mathbb{R}^k}\left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \mathbf{V}\boldsymbol{\gamma}\right|\right|_2^2$ and $\min_{\boldsymbol{\eta}\in\mathbb{R}^k}\left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \tilde{\mathbf{Z}}\boldsymbol{\eta}\right|\right|_2^2$, respectively. Therefore,

$$\begin{aligned}\left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls}\right|\right|_2^2 &\le \left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \tilde{\mathbf{Z}}\hat{\boldsymbol{\gamma}}_i^{ls}\right|\right|_2^2 \\ &\le \left(\left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \mathbf{V}\hat{\boldsymbol{\gamma}}_i^{ls}\right|\right|_2 + \left|\left|(\mathbf{V}-\tilde{\mathbf{Z}})\hat{\boldsymbol{\gamma}}_i^{ls}\right|\right|_2\right)^2\end{aligned} \tag{3.37}$$

and

$$\begin{aligned}\left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \mathbf{V}\hat{\boldsymbol{\gamma}}_i^{ls}\right|\right|_2^2 &\le \left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \mathbf{V}\hat{\boldsymbol{\eta}}_i^{ls}\right|\right|_2^2 \\ &\le \left(\left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls}\right|\right|_2 + \left|\left|(\tilde{\mathbf{Z}}-\mathbf{V})\hat{\boldsymbol{\eta}}_i^{ls}\right|\right|_2\right)^2.\end{aligned} \tag{3.38}$$

To study (3.37) and (3.38), we first investigate the value of $\left|\left|(\mathbf{V}-\tilde{\mathbf{Z}})\hat{\boldsymbol{\gamma}}_i^{ls}\right|\right|_2$ and $\left|\left|(\tilde{\mathbf{Z}}-\mathbf{V})\hat{\boldsymbol{\eta}}_i^{ls}\right|\right|_2$. From Theorem 2.2.1 in Chikuse (2003), matrix $\mathbf{A}$ can be expressed as $\mathbf{A} = \mathbf{G}\left(\mathbf{G}^\top\mathbf{G}\right)^{-1/2}$, where each element of $m \times k$ matrix $\mathbf{G}$ is i.i.d. from $\mathcal{N}(0,1)$. Let $\mathbf{E} = \mathbf{O}_n^\top\mathbf{G}$. Then $\mathbf{O}_n^\top\mathbf{A} = \mathbf{E}\left(\mathbf{G}^\top\mathbf{G}\right)^{-1/2}$. From Lemma 3.8 and the independence between $\mathbf{A}$ and $\mathbf{Z}$, for any $h_1 > 0$ and $h_2 > 0$, we have

$$\begin{aligned} &P\left[\lambda_{\max}\left(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\right) \geq (1+\sqrt{k/n}+h_1)^2\right] \leq \exp\left(-nh_1^2/2\right) \quad \text{and} \\ &P\left[\lambda_{\min}\left(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\right) \leq (1-\sqrt{k/n}-h_2)^2\right] \leq \exp\left(-nh_2^2/2\right). \end{aligned} \tag{3.39}$$

Based on SVD, for any matrix $\mathbf{B}$, the nonzero eigenvalues of $\mathbf{B}^\top\mathbf{B}$ and $\mathbf{B}\mathbf{B}^\top$ are the same. Hence, we have

$$\begin{aligned} &\lambda_{\max}\left(\mathbf{V}\big(\mathbf{V}^\top\mathbf{V}\big)^{-1}\mathbf{A}^\top\mathbf{O}_n\mathbf{O}_n^\top\mathbf{A}\big(\mathbf{V}^\top\mathbf{V}\big)^{-1}\mathbf{V}^\top\right) \\ &\leq \lambda_{\max}\big(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\big)\frac{1}{\lambda_{\min}\big(\frac{1}{n}\mathbf{E}^\top\mathbf{U}^\top\big(\mathbf{I}-\mathbf{P}_1\big)\mathbf{U}\mathbf{E}\big)} \\ &\leq \frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2} \end{aligned} \tag{3.40}$$

and

$$\begin{aligned} &\lambda_{\max}\left(\tilde{\mathbf{Z}}\big(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}}\big)^{-1}\mathbf{A}^\top\mathbf{O}_n\mathbf{O}_n^\top\mathbf{A}\big(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}}\big)^{-1}\tilde{\mathbf{Z}}^\top\right) \\ &\leq \lambda_{\max}\big(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\big)\frac{1}{\lambda_{\min}\big(\frac{1}{n}\mathbf{E}^\top\frac{\mathbf{D}}{\sqrt{m}}\mathbf{U}^\top\big(\mathbf{I}-\mathbf{P}_1\big)\mathbf{U}\frac{\mathbf{D}}{\sqrt{m}}\mathbf{E}\big)} \\ &\leq \frac{1}{\lambda_{\min}\big(\frac{\mathbf{D}^2}{m}\big)}\cdot\frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2} \end{aligned} \tag{3.41}$$

with probability at least $1-\exp\left(-nh_1^2/2\right)-\exp\left(-nh_2^2/2\right)$. Based on (3.36), (3.40) and (3.41), upper bounds can be derived as follows:

$$\begin{aligned} \big|\big|\big(\mathbf{V}-\tilde{\mathbf{Z}}\big)\hat{\boldsymbol{\gamma}}_i^{ls}\big|\big|_2^2 &= \big|\big|\big(\mathbf{I}-\mathbf{P}_1\big)\mathbf{U}\big(\mathbf{I}-\frac{\mathbf{D}}{\sqrt{m}}\big)\mathbf{O}_n^\top\mathbf{A}\big(\mathbf{V}^\top\mathbf{V}\big)^{-1}\mathbf{V}^\top\mathbf{e}_i\big|\big|_2^2 \\ &\leq \max_{i=1,\dots,n}\big(1-\frac{d_i}{\sqrt{m}}\big)^2\big|\big|\mathbf{O}_n^\top\mathbf{A}\big(\mathbf{V}^\top\mathbf{V}\big)^{-1}\mathbf{V}^\top\mathbf{e}_i\big|\big|_2^2 \\ &\leq t\cdot\frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2} \end{aligned} \tag{3.42}$$

and

$$||(\tilde{\mathbf{Z}} - \mathbf{V})\hat{\boldsymbol{\eta}}_i^{ls}||_2^2 = ||(\mathbf{I} - \mathbf{P}_1)\mathbf{U}(\mathbf{I} - \frac{\mathbf{D}}{\sqrt{m}})\mathbf{O}_n^\top \mathbf{A}(\tilde{\mathbf{Z}}^\top \tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}^\top \mathbf{e}_i||_2^2$$

$$\leq \max_{i=1,\dots,n} (1 - \frac{d_i}{\sqrt{m}})^2 ||\mathbf{O}_n^\top \mathbf{A}(\tilde{\mathbf{Z}}^\top \tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}^\top \mathbf{e}_i||_2^2$$

$$\leq \max_{i=1,\dots,n} \left(1 - \frac{d_i}{\sqrt{m}}\right)^2 \cdot \frac{1}{\min\limits_{i=1,\dots,n} \left(\frac{d_i^2}{m}\right)} \cdot \frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2} \tag{3.43}$$

$$\leq \frac{t}{(1-\sqrt{t})^2} \cdot \frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2}$$

with probability at least $1 - O(n^2 m^{-1} t^{-1}) - \exp(-nh_1^2/2) - \exp(-nh_2^2/2)$. Combining (3.37), (3.38), (3.42) and (3.43), with $h_1 = n^{-1/4}$, $h_2 = n^{-1/4}$ and $t = n^{-c}$, where c is a positive constant, we have

$$||\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{e}_i - \tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls}||_2^2 \leq ||\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{e}_i - \mathbf{V}\hat{\boldsymbol{\gamma}}_i^{ls}||_2^2 + 3n^{-c/2} \cdot \frac{1+\sqrt{k/n}+n^{-1/4}}{1-\sqrt{k/n}-n^{-1/4}} \quad \text{and}$$

$$||\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{e}_i - \mathbf{V}\hat{\boldsymbol{\gamma}}_i^{ls}||_2^2 \leq ||\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{e}_i - \tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls}||_2^2 + \frac{3}{n^{c/2}-1} \cdot \frac{1+\sqrt{k/n}+n^{-1/4}}{1-\sqrt{k/n}-n^{-1/4}}$$

with probability at least $1 - O(n^{2+c}m^{-1}) - 2\exp\left(-n^{1/2}/2\right)$. As the above derivation is valid for any $\mathbf{e}_i$, when $n \to \infty$ and m is sufficiently large, we obtain

$$\max_{i=1\dots,n} |h_{ii} - s_{ii}|^2 = o_p(1). \tag{3.44}$$

According to the definitions of $\mathbf{V}$ and $\mathbf{A}$, the hat matrix $\mathbf{S}$ can be denoted as

$$\mathbf{S} = (\mathbf{I} - \mathbf{P}_1)\mathbf{U}\mathbf{O}_n^\top \mathbf{G}\left(\mathbf{G}^\top \mathbf{O}_n \mathbf{U}^\top (\mathbf{I} - \mathbf{P}_1)\mathbf{U}\mathbf{O}_n^\top \mathbf{G}\right)^{-1} \mathbf{G}^\top \mathbf{O}_n \mathbf{U}^\top (\mathbf{I} - \mathbf{P}_1),$$

where $\mathbf{U}\mathbf{O}_n^\top$ is independent of $\mathbf{G}$ and satisfies $\mathbf{U}\mathbf{O}_n^\top \mathbf{O}_n \mathbf{U}^\top = \mathbf{I}$. From the definition of $\mathbf{G}$, Lemma 3.5 and the dominated convergence theorem, we obtain

$$E\left\{\frac{1}{n}\sum_{i=1}^{n}\left(s_{ii} - \rho\right)^2\right\} \to 0.$$

Combining this with (3.44) and Markov's inequality, we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\big(h_{ii}-\rho\big)^2 = \frac{1}{n}\sum_{i=1}^{n}\big(h_{ii}-s_{ii}+s_{ii}-\rho\big)^2$$

$$\leq \max_{i=1\ldots,n} 2\big(h_{ii}-s_{ii}\big)^2 + \frac{2}{n}\sum_{i=1}^{n}\big(s_{ii}-\rho\big)^2$$

$$= o_p(1),$$

which completes the proof.

□

3.6.7 Proof of Theorem 3.6

Proof. Let $\mathbf{U} = \mathbf{X}\mathbf{P}_k\boldsymbol{\Sigma}_1^{-1/2}$, with the i-th row denoted by $\mathbf{u}_i$ and $\boldsymbol{\Sigma}_1 = \mathbf{P}_k^\top\boldsymbol{\Sigma}\mathbf{P}_k$. Define $\boldsymbol{\eta} = \boldsymbol{\Sigma}_1^{-1/2}\mathbf{P}_k^\top\boldsymbol{\Sigma}\boldsymbol{\theta}$ and $\omega^2 = ||\boldsymbol{\eta}||_2^2$. Based on the normality assumption, we derive a decomposition, $\mathbf{x}_i^\top\boldsymbol{\theta} = \mathbf{u}_i^\top\boldsymbol{\eta} + q_i$, where q_i is independent of $\mathbf{u}_i$. According to Lemma 3.1, the model can be written as

$$y_i = c_{0,k}\mathbf{u}_i^\top\boldsymbol{\eta} + e_i, \tag{3.45}$$

where $c_{0,k} = E(\boldsymbol{\eta}^\top\mathbf{u}_1 y_1)/\omega^2$ when $\boldsymbol{\eta} \neq \mathbf{0}$, and e_i satisfies $E(e_i) = 0$ and $E(\mathbf{u}_i e_i) = \mathbf{0}$. Let σ^2 denote the variance of e_i. It satisfies $\sigma^2 = Var(y) - c_{0,k}^2\omega^2$. The calculation shows that

$$\frac{T_{n,k}-1}{\sqrt{2/n\rho(1-\rho)}} = \frac{\sqrt{\frac{n\rho(1-\rho)}{2}}\left(\frac{c_{0,k}^2\boldsymbol{\eta}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{U}\boldsymbol{\eta}}{k} + \frac{2c_{0,k}\boldsymbol{\eta}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{e}}{k} + \mathbf{e}^\top\mathbf{F}\mathbf{e}\right)}{\frac{\mathbf{e}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\mathbf{e}}{n-1-k}}, \tag{3.46}$$

where $\mathbf{F} = \frac{\mathbf{H}_k}{k} - \frac{\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k}{n-1-k}$ and $\mathbf{e} = (e_1,\ldots,e_n)^\top$.

According to Assumption H4, $c_{0,k}^2\omega^2 = o(1)$. The calculation shows

$$E(\boldsymbol{\eta}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{U}\boldsymbol{\eta}) = (n-1)\omega^2,\ Var(\boldsymbol{\eta}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{U}\boldsymbol{\eta}) \leq 10(n-1)\omega^4,$$

$$E(\boldsymbol{\eta}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{e}) = 0,\ Var(\boldsymbol{\eta}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{e}) \leq 16\sqrt{2+E(y^4)}(n-1)\omega^2.$$

By Markov's inequality, we have

$$\frac{\sqrt{n}}{k} c_{0,k}^2 \boldsymbol{\eta}^\top \mathbf{U}^\top (\mathbf{I} - \mathbf{P}_1) \mathbf{U} \boldsymbol{\eta} = \frac{\sqrt{n}}{\rho} c_{0,k}^2 \omega^2 + o_p(1), \ \frac{\sqrt{n}}{k} c_{0,k} \boldsymbol{\eta}^\top \mathbf{U}^\top (\mathbf{I} - \mathbf{P}_1) \mathbf{e} = o_p(1). \tag{3.47}$$

For the study of $\mathbf{e}^\top \mathbf{F} \mathbf{e}$, we follow a similar method in Section 3.6.3 and derive

$$\frac{T_{n,k} - 1}{\sqrt{2/n\rho(1-\rho)}} = \frac{\sqrt{\frac{n(1-\rho)}{2\rho}} c_{0,k}^2 \omega^2 + \sqrt{\frac{n\rho(1-\rho)}{2}} \mathbf{e}^\top \mathbf{M} \mathbf{e} + o_p(1)}{\sigma^2 + o_p(1)},$$

where $\mathbf{M} = \frac{\mathbf{H_G}}{k-1} - \frac{\mathbf{I} - \mathbf{P}_1 - \mathbf{H_G}}{n-k}$, with the projection matrix $\mathbf{H_G}$ independent of $\mathbf{e}$. Based on a similar method in Section 3.6.6, we can derive

$$\frac{\mathbf{e}^\top \mathbf{M} \mathbf{e}}{\sigma^2 \sqrt{2/n\rho(1-\rho)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Consequently, the asymptotic power function satisfies

$$\begin{aligned}\Psi_n^{RP}(\boldsymbol{\theta}; \mathbf{P}_k) &= P\left(\frac{T_{n,k} - 1}{\sqrt{2/n\rho(1-\rho)}} > z_\alpha\right) \\ &= \Phi\left(-z_\alpha + \sqrt{\frac{n(1-\rho)}{2\rho}} \frac{c_{0,k}^2 \omega^2}{\sigma^2}\right) + o(1),\end{aligned}$$

which completes the proof. □

3.6.8 Proof of Lemma 3.2

Proof. The proof follows a similar method in Section 2.6.3. □

3.6.9 Proof of Theorem 3.7

Proof. Under $\mathbf{H}_{0,\mathbf{B}}$, the model becomes

$$y = f(\mathbf{w}^T \boldsymbol{\gamma}, \epsilon). \tag{3.48}$$

Define

$$\mathbf{D} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{\mathbf{B}^\perp} \end{pmatrix}, \ \mathbf{R}_{k_2} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{\mathbf{B}^\perp} & \mathbf{S}_{\mathbf{B}}\mathbf{P}_{k_2} \end{pmatrix}.$$

Then, $\mathbf{w} = \mathbf{D}^\top \mathbf{x}$ and $\mathbf{u}_{k2} = \mathbf{R}_{k_2}^\top \mathbf{x}$. Consider a working model

$$y = f(\mathbf{u}_{k2}^\top \boldsymbol{\xi}, \epsilon), \tag{3.49}$$

where $\mathbf{u}_{k2}^\top \boldsymbol{\xi} = \mathbf{w}^\top \boldsymbol{\gamma} + \mathbf{x}_2^\top \mathbf{S}_{\mathbf{B}} \mathbf{P}_{k_2} \boldsymbol{\xi}_2$ with $\boldsymbol{\xi}_2$ being a k_2-dimensional vector. When $\boldsymbol{\xi}_2 = \mathbf{0}$, the distribution of y under model (3.49) is the same as that in model (3.48). Therefore, it is sufficient to study the test statistic under model (3.49) with $\boldsymbol{\xi}_2 = \mathbf{0}$. With a similar analysis in Section 3.6.4, we can derive

$$\frac{T_{n,k_2} - 1}{\sqrt{2(1-\rho_1)/n\rho_2(1-\rho_1-\rho_2)}} \to \mathcal{N}(0,1),$$

which completes the proof. □

3.6.10 Proof of Theorem 3.8

Proof. Let the definition of $\mathbf{D}$, $\mathbf{R}_{k_2}$ are the same as that in Section 3.6.9. Let $\boldsymbol{\Sigma}_{\mathbf{w}} = \mathbf{D}^\top \boldsymbol{\Sigma} \mathbf{D}$ and $\boldsymbol{\Sigma}_1 = \mathbf{R}_{k_2}^\top \boldsymbol{\Sigma} \mathbf{R}_{k_2}$, which are the covariance matrix of $\mathbf{w}$ and $\mathbf{u}_{k2}$, respectively. Define $\mathbf{U} = \mathbf{X} \mathbf{R}_{k_2} \boldsymbol{\Sigma}_1^{-1/2}$ and denote its i-th row by $\mathbf{u}_i$. The assumption of normality leads to a decomposition, $\mathbf{x}_i^\top \boldsymbol{\theta} = \mathbf{u}_i^\top \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\xi} + r_i$, where $\boldsymbol{\xi} = \boldsymbol{\Sigma}_1^{-1} \mathbf{R}_{k_2}^\top \boldsymbol{\Sigma} \boldsymbol{\theta}$ and r_i is independent of $\mathbf{u}_i$. Therefore, the model can be written as

$$y_i = f(\mathbf{u}_i^\top \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\xi} + r_i, \epsilon_i), \tag{3.50}$$

where r_i and ϵ_i are independent of $\mathbf{u}_i$. Consider $\mathbf{w}_i$. It can be denoted as $\mathbf{w}_i = \mathbf{D}^\top \mathbf{x}_i$ and satisfies $\mathbf{u}_i^\top \boldsymbol{\Sigma}_1^{1/2} = (\mathbf{w}_i^\top, \mathbf{x}_{2i}^\top \mathbf{S}_{\mathbf{B}} \mathbf{P}_{k_2})$. Let $\boldsymbol{\delta} = \boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \mathbf{D}^\top \boldsymbol{\Sigma} \mathbf{R}_{k_2} \boldsymbol{\Sigma}_1^{-1} \mathbf{R}_{k_2}^\top \boldsymbol{\Sigma} \boldsymbol{\theta}$. Then,

$$\mathbf{u}_i^\top \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\xi} = \mathbf{w}_i^\top \boldsymbol{\delta} + q_i,$$

where q_i is independent of $\mathbf{w}_i$. Let $\boldsymbol{\eta} = \mathbf{R}_{k_2}\boldsymbol{\xi} = (\boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)^\top$, where $\boldsymbol{\eta}_1 \in \mathbb{R}^{p_1}$ and $\boldsymbol{\eta}_2 \in \mathbb{R}^{p_2}$. Define $\nu^2 = Var(q_1)$. The calculation yields

$$\nu^2 = \boldsymbol{\eta}_2^\top \left[\boldsymbol{\Sigma}_{22} - (\boldsymbol{\Sigma}_{21}\ \boldsymbol{\Sigma}_{22})\, \mathbf{D}(\mathbf{D}^\top \boldsymbol{\Sigma} \mathbf{D})^{-1} \mathbf{D}^\top \begin{pmatrix} \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{22} \end{pmatrix} \right] \boldsymbol{\eta}_2.$$

Let $c_{0,k_2} = \boldsymbol{\xi}^\top \boldsymbol{\Sigma}_1^{1/2} E(\mathbf{u}_i y_i)/\boldsymbol{\xi}^\top \boldsymbol{\Sigma}_1 \boldsymbol{\xi}$, where $\mathbf{P}_{k_2}$ is treated as fixed in the expectation. Under the condition $c_{0,k_2}^2 \nu^2 = o(1)$, the method in Section 3.6.5 is applicable. Therefore, we obtain

$$\begin{aligned} \Psi_n^{RP}(\boldsymbol{\theta}_2; \mathbf{B}, \mathbf{P}_{k_2}) &= P(\frac{T_{n,k_2} - 1}{\sqrt{2(1-\rho_1)/n\rho_2(1-\rho_1-\rho_2)}} > z_\alpha) \\ &= \Phi(-z_\alpha + \sqrt{\frac{n(1-\rho_1)(1-\rho_1-\rho_2)}{2\rho_2}} \frac{c_{0,k_2}^2 \nu^2}{\sigma^2}) + o(1), \end{aligned}$$

where $\sigma^2 = Var(y) - c_{0,k_2}^2 \boldsymbol{\xi}^\top \boldsymbol{\Sigma}_1 \boldsymbol{\xi}$. The proof is completed. □

3.6.11 Proof of Auxiliary Lemmas

Proof of Lemma 3.5. Let $\mathbf{S}_{n-1} = \frac{1}{n-1}\sum_{i=1}^{n-1} \mathbf{x}_i \mathbf{x}_i^\top$, where $\mathbf{x}_i$ denotes the i-th row of $\mathbf{X}$. According to Bai and Yin (1993, Theorem 2), it shows that the extreme eigenvalues of $\mathbf{S}_{n-1}$ satisfy $\lambda_{\max}(\mathbf{S}_{n-1}) \to (1+\sqrt{\zeta})^2$ and $\lambda_{\min}(\mathbf{S}_{n-1}) \to (1-\sqrt{\zeta})^2, a.s.$. Based on the result in Marčenko and Pastur (1967), $\frac{1}{p} tr(\mathbf{S}_{n-1}^{-1}) \to \frac{1}{1-\zeta}$. Then, we obtain

$$E(\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1} \mathbf{x}_n | \mathbf{S}_{n-1}) = \frac{p}{n-1} \frac{tr(\mathbf{S}_{n-1}^{-1})}{p} \to \frac{\zeta}{1-\zeta} \quad \text{and}$$

$$Var(\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1} \mathbf{x}_n | \mathbf{S}_{n-1}) = \frac{2}{(n-1)^2} tr\left(\mathbf{S}_{n-1}^{-2}\right) \le \frac{2p}{(n-1)^2} \frac{1}{\lambda_{\min}^2(\mathbf{S}_{n-1})} \to 0.$$

Therefore,

$$E\left\{ \left(\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1} \mathbf{x}_n - \frac{\zeta}{1-\zeta} \right)^2 \right\} \to 0.$$

From Woodbury formula,

$$\mathbf{x}_n^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_n = \frac{\mathbf{x}_n^\top(\sum_{j\neq n}\mathbf{x}_j\mathbf{x}_j^\top)^{-1}\mathbf{x}_n}{1+\mathbf{x}_n^\top(\sum_{j\neq n}\mathbf{x}_j\mathbf{x}_j^\top)^{-1}\mathbf{x}_n}$$

$$= \frac{\frac{1}{n-1}\mathbf{x}_n^\top\mathbf{S}_{n-1}^{-1}\mathbf{x}_n}{1+\frac{1}{n-1}\mathbf{x}_n^\top\mathbf{S}_{n-1}^{-1}\mathbf{x}_n}.$$

Let $f(x) = \frac{x}{1+x}$. It satisfies $f'(x) \leq 1$, for $x \geq 0$. Based on the mean value theorem, we get

$$\left|\mathbf{x}_n^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_n - \zeta\right| \leq \left|\frac{1}{n-1}\mathbf{x}_n^\top\mathbf{S}_{n-1}^{-1}\mathbf{x}_n - \frac{\zeta}{1-\zeta}\right|,$$

which implies

$$E\left\{(h_{nn}-\zeta)^2\right\} \leq E\left\{\left(\frac{1}{n-1}\mathbf{x}_n^\top\mathbf{S}_{n-1}^{-1}\mathbf{x}_n - \frac{\zeta}{1-\zeta}\right)^2\right\} \to 0. \tag{3.51}$$

For $i = 1, \ldots, n-1$, we follow a similar method and derive

$$\begin{aligned} E\left\{(h_{ii}-\zeta)^2\right\} &\leq E\left\{\left(\frac{1}{n-1}\mathbf{x}_i^\top(\frac{1}{n-1}\sum_{j\neq i}\mathbf{x}_j\mathbf{x}_j^\top)^{-1}\mathbf{x}_i - \frac{\zeta}{1-\zeta}\right)^2\right\} \\ &= E\left\{\left(\frac{1}{n-1}\mathbf{x}_n^\top\mathbf{S}_{n-1}^{-1}\mathbf{x}_n - \frac{\zeta}{1-\zeta}\right)^2\right\} \to 0. \end{aligned} \tag{3.52}$$

Therefore,

$$\max_{i=1,\ldots,n} E\left\{(h_{ii}-\zeta)^2\right\} \to 0,$$

which completes the proof.

□

Proof of Lemma 3.7. According to Woodbury formula, we have

$$\mathbf{H} = \mathbf{H_w} + (\mathbf{I}-\mathbf{H_w})\mathbf{G}[\mathbf{G}^\top(\mathbf{I}-\mathbf{H_w})\mathbf{G}]^{-1}\mathbf{G}^\top(\mathbf{I}-\mathbf{H_w})$$

$$= \mathbf{H_w} + (\mathbf{I}-\mathbf{H_w})[\mathbf{H_G} + \frac{\mathbf{H_G}\mathbf{H_w}\mathbf{H_G}}{1-tr(\mathbf{H_G}\mathbf{H_w})}](\mathbf{I}-\mathbf{H_w}).$$

This leads to

$$\mathbf{e}^\top \mathbf{H}\mathbf{e} = \mathbf{e}^\top [\mathbf{H_w} + (\mathbf{I} - \mathbf{H_w})(\mathbf{H_G} + \frac{\mathbf{H_G H_w H_G}}{1 - tr(\mathbf{H_G H_w})})(\mathbf{I} - \mathbf{H_w})]\mathbf{e}$$

$$= \mathbf{e}^\top \mathbf{H_G}\mathbf{e} + Re,$$

where Re is defined as

$$Re = \frac{r_1^2}{r_4} + \frac{r_3^2}{r_4(1-g)} - \frac{2r_1r_3}{r_4} - \frac{2r_1r_2r_3}{r_4^2(1-g)} + \frac{r_1^2 r_2}{r_4^2} + \frac{r_1^2 r_2^2}{r_4^3(1-g)},$$

with $r_1 = \mathbf{e}^\top \mathbf{w}$, $r_2 = \mathbf{w}^\top \mathbf{H_G} \mathbf{w}$, $r_3 = \mathbf{e}^\top \mathbf{H_G} \mathbf{w}$, $r_4 = \mathbf{w}^\top \mathbf{w}$, and $g = tr(\mathbf{H_G H_w})$.

For r_1, Lemma 3.4 shows that $E(r_1) = 0, Var(r_1) = nE(e_1^2 w_1^2)$. Then, $r_1 = o_p(n^{\gamma_1 + 0.5})$, for any $\gamma_1 > 0$, based on Markov's inequality. For r_2, the independence between $\mathbf{w}$ and $\mathbf{G}$ implies that $E(r_2) = p - 1$ and $Var(r_2) = 2(p-1)$. Then, $\frac{r_2}{n} = \zeta + o_p(1)$ by Markov's inequality. For r_3, Lemmas 3.4 and 3.5 show that

$$E(r_3) = 0 \text{ and } E(r_3^2) \leq Cn(1 + o(1)) + (p-1)E(e_1^2)$$

for a constant C. Then, $r_3 = o_p(n^{\gamma_2 + 0.5})$, for any $\gamma_2 > 0$, based on Markov's inequality. For r_4, the strong law of large numbers shows that $\frac{r_4}{n} \to 1$, a.s.. From $g = r_2/r_4$, we obtain $g = \zeta + o_p(1)$. Consequently, for any $\delta > 0$, $Re = o_p(n^\delta)$, which completes the proof. □

Chapter 4

Concluding Remarks

In this thesis, the tests are studied in ultrahigh-dimensional settings, where p can be much greater than n. We derive theoretical and numerical results in the linear regression model and SIM, respectively. Although the linear regression model is included in SIM, the proposed testing approaches have different intuitions and theoretical analysis methods.

In Chapter 2, we investigate the proposed test in the linear regression model. The proposed test statistic could be naturally considered as an extension of the classical F-statistic, which is undefinable under the $p > n$ settings. The extension is mainly summarized from two aspects. First, with the proposed test, there is little explicit restrictions on the relationship between n and p, for which the test could be implemented in ultrahigh-dimensional cases and is more adaptable to modern data analysis. Second, the proposed test needs mild assumption on the setting, where a general class of distributions of the covariate and the error are included. The advantage of the wide application range is contributed by the implement of random projection. Motivated by the result that randomly projected data is asymptotically close to the normal distribution, we give rigorous theoretical analysis and demonstrate that the performance of the proposed test statistic is similar to the case where the covariate follows the normal distribution. In addition, we provide a sufficient condition where the asymptotic testing power could be deterministic and reach the optimal value, even with randomness form random projection. Another significant problem is the choice of ρ, which is the proportion of the dimension of the random projection data over the dimension p of the covariate. We provide a detailed discussion and give suggestion on its selection under different situations. Through simulation study, our proposed test is shown to have a well control of the type I error and a powerful performance on the empirical power. It is also demonstrated to have certain advantages in sparse or highly correlated cases.

In Chapter 3, we investigate the proposed test in a more general class of models, known as SIM. In the high-dimensional hypothesis testing problem, the idea of studying a nonlinear model in a linear way is proposed for the first time. One of the significant benefits of this method is its ability to complete statistical test on the vector of regression coefficients $\boldsymbol{\theta}$ without estimating the link function. Specifically, we transform SIM into a linear form and find that the linear regression coefficients is $\boldsymbol{\theta}$ up to a scalar $c_0 \neq 0$. This makes it possible to design simple and effective methods for testing the significance of $\boldsymbol{\theta}$. In the high-dimensional regime where $p/n \to \zeta \in (0, 1)$, we provide a detailed analysis of asymptotic null distribution and asymptotic local power function of the F-statistic and demonstrate the effect of ζ on the power of the test. In an ultrahigh-dimensional setting, theoretical and numerical studies demonstrate that the proposed test has good power over a wide range of alternatives and possesses certain advantages in sparse cases. In addition, the strong testing power of the proposed method can be guaranteed with a sufficient condition. The usage of random projection is significant in our proposed test. It reduces the dimension of the data while preserving the main information. Besides, the property of randomly projected data suggests that the theoretical analysis might be available for general distributions. The selection of ρ is analyzed in several different settings. In practice, a value around 0.5 is recommended.

Overall, our proposed tests provide effective approaches to hypothesis testing in high-dimensional linear regression model and SIM. The proposed test statistics have the advantages of simple structure and straightforward application. And they are computationally simple to implement. Therefore, it is an useful contribution to the literature on statistical inference in high-dimensional models.

For future research, several directions can be considered. First, the developed method could be investigated in more complicated models. This method could be extended to the multiple-index model. In addition, there are more challenges and interests to studying other types of structural nonparametric regression models, such as the additive models (Friedman and Stuetzle , 1981) and varying-coefficient models (Hastie and Tibshirani , 1993; Fan and Zhang , 2008). These models have more flexibility and arise in many practical applications. In addition, applications of the theoretical analysis of the proposed tests could be further studied, such as multiple testing problems. Second, it is worthy investigating the possibility of applying random projection to some other statistical inference problems. Recent literature have already contributed to some areas, such as hypothesis testing, nonparametric regression estimation (Yang, Pilanci, and Wainwright , 2017) and classification (Cannings and Samworth , 2017). It is demonstrated that random projection could perform successfully to simplify the computation, while preserving certain statistical performance. Third, the theoretical development of projected data is of significance. Many basic statistical methods assume that the conditional mean is linear and the conditional variance is constant, such as sufficient dimension reduction methods. Since the normal distribution is the only distribution satisfying the both properties, the assumptions seem to be relatively rigorous. However, this concern can be alleviated to a certain extent through projection in high-dimensional settings. Specifically, from Hall and Li (1993), Leeb (2013) and Steinberger and Leeb (2018), the both properties hold in an approximate sense by a large class of distributions when conditional on lower-dimensional projections. Motivated by this result, other significant properties of the normal distribution might be considered to be satisfied by the projected data, such as Stein's lemma and the property that uncorrelateness and independence are equivalent. Considering another series of literature, known as projection pursuit, methods were proposed to detect interesting low-dimensional structures in high-dimensional data. For example, Diaconis and Freedman (1984) studied the behavior of the empirical distributions of univariate projections of the data and showed that almost all of them were asymptotically close to the normal. Bickel, Kur, and Nadler (2018) considered the problem in high-dimensional settings. They demonstrated that there were projections whose corresponding empirical distribution can approximate any arbitrary distribution, and this was shown to be significant in many statistical methods, such as non-Gaussian component analysis methods. Other applications of these results are still needed to be explored. Fourth, considering the statistical inference methods developed based on sparse structures, the criterion for when to implement these methods needs to be further considered. For example, in a practical setting, it is not clear whether the maximal-absolute-error-type testing methods can be applied, and it is uncertain how the performance can be expected when the sparsity condition is violated. This kind of question is common in real data analysis, since the level of sparsity is usually unknown in advance. In addition, there is little reason to believe that the sparsity condition holds without any inspection measures available. Recently, related problems have gain much interest. For example, in Cai and Guo (2017) and Javanmard and Montanari (2018), the question whether the strong sparsity conditions were needed was investigated. In Carpentier and Verzelen (2021), the problem of testing sparsity of the regression coefficients was considered. And the statistical inference when sparsity condition might be absent was studied in Zhu and Bradic (2018) and Bradic, Fan, and Zhu (2018).

Bibliography

Bryc, W. (1995). *The Normal Distribution: Characterizations with Applications. Lecture Notes in Statistics* **100**. Springer, New York.

Bechar, I. (2009). A Bernstein-type inequality for stochastic processes of quadratic forms of Gaussian variables. *arXiv preprint arXiv:0909.3595*.

Bradic, J., Fan, J. and Zhu, Y. (2018). Testability of high-dimensional linear models with non-sparse structures. *arXiv preprint arXiv:1802.09117*.

Bhansali, R. J., Giraitis, L. and Kokoszka, P. S. (2007). Convergence of quadratic forms with nonvanishing diagonal. *Statistics & Probability Letters* **77** 726–734.

Bickel, P. J., Kur, G. and Nadler, B. (2018). Projection pursuit in high dimensions. *Proceedings of the National Academy of Sciences of the United States of America* **115** 9151–9156.

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

Bai, Z. D. and Yin, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *The Annals of Probability* **21** 1275–1294.

Bühlmann, P. (2013). Statistical significance in high-dimensional linear models. *Bernoulli* **19** 1212–1242.

Chikuse, Y. (2003). *Statistics on special manifolds. Lecture Notes in Statistics* **174**. Springer-Verlag, New York.

Cai, T. T. and Guo, Z. (2017). Confidence intervals for high-dimensional linear regression: minimax rates and adaptivity. *The Annals of Statistics* **45** 615–646.

Cannings, T. I. and Samworth, R. J. (2017). Random-projection ensemble classification. *Journal of the Royal Statistical Society. Series B. Statistical Methodology* **79** 959–1035.

Carpentier, A. and Verzelen, N. (2021). Optimal sparsity testing in linear regression model. *Bernoulli* **27** 727–750.

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.

Durrett, R. (2019). *Probability—theory and examples. Cambridge Series in Statistical and Probabilistic Mathematics* **49**. Cambridge University Press, Cambridge.

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

Davidson, K. R. and Szarek, S. J. (2001). Local operator theory, random matrices and Banach spaces. In *Handbook of the geometry of Banach spaces, Vol. I*, 317–366. North-Holland, Amsterdam.

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* **74** 37–65.

Friedman, J. H. and Stuetzle, W. (1981). Projection pursuit regression. *Journal of the American Statistical Association* **76** 817–823.

Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric multivariate and related distributions. Monographs on Statistics and Applied Probability* **36**. Chapman and Hall, Ltd., London.

Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and its Interface* **1** 179–195.

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

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

Gamez-Pozo, A., Carrión, R. M. P., Manso, L., Crespo, C., Mendiola, C., Lopez-Vacas, R., Berges-Soria, J., López, I. Á, Margeli, M., Calero, J. L. B. 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. (2017). A Statistically and Numerically Efficient Independence Test based on Random Projections and Distance Covariance. *arXiv preprint arXiv: 1701.06054*.

Hall, P. and Li, K.C. (1993). On almost linearity of low-dimensional projections from high-dimensional data. *The Annals of Statistics* **21** 867–889.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B. Methodological* **55** 757–796.

Javanmard, A. and Montanari, A. (2018). Debiasing the Lasso: optimal sample size for Gaussian designs. *The Annals of Statistics* **46** 2593–2622.

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

Leeb, H. (2013). On the conditional distributions of low-dimensional projections from high-dimensional data. *The Annals of Statistics* **41** 464–483.

Li, K.C. and Duan, N.H. (1989). Regression analysis under link violation. *The Annals of Statistics* **17** 1009–1052.

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.

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. (2020). Global and simultaneous hypothesis testing for high-dimensional logistic regression models. *Journal of the American Statistical Association* **0** 1–15.

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik* **1** 457.

Neykov, M., Liu, J. S. and Cai, T. X. (2016). L1-Regularized Least Squares for Support Recovery of High Dimensional Single Index Models with Gaussian Designs. *The Journal of Machine Learning Research* **17** 2976–3012.

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., Gómez, P., Semiglazov, V., Eiermann, W., Tjulandin, S. 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.

Hall, P. and Li, K. C. (1993). On almost linearity of low-dimensional projections from high-dimensional data. *The Annals of Statistics* **21** 867–889.

Parker, J. S., Mullins, M., Cheang, M. C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., Quackenbush, J. F., Stijleman, I. J., Palazzo, J., Marron, J. S., Nobel, A. B., Mardis, E., Nielsen, T. O., Ellis, M. J., Perou, C. M., & Bernard, P. S. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of clinical oncology* **27** 1160–1167.

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

Steinberger, L. and Leeb, H. (2018). On conditional moments of high-dimensional random vectors given lower-dimensional projections. *Bernoulli* **24** 565–591.

Thrampoulidis, C., Abbasi, E. and Hassibi, B. (2015). LASSO with non-linear measurements is equivalent to one with linear measurements. In *Advances in Neural Information Processing Systems* 3420–3428.

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* **3**. 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.

Yang, Y., Pilanci, M. and Wainwright, M. J. (2017). Randomized sketches for kernels: fast and optimal nonparametric regression. *The Annals of Statistics* **45** 991–1023.

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

Zhang, X. Y. and Cheng, G. (2017). Simultaneous Inference for High-Dimensional Linear Models. *Journal of the American Statistical Association* **112** 757–768.

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

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