JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2021, VOL. 00, NO. 0, 1–13: Theory and Methods
https://doi.org/10.1080/01621459.2021.1884561

Taylor & Francis
Taylor & Francis Group

Linear Hypothesis Testing in Linear Models With High-Dimensional Responses

Changcheng Li and Runze Li

Department of Statistics, Pennsylvania State University at University Park, University Park, PA

ABSTRACT

In this article, we propose a new projection test for linear hypotheses on regression coefficient matrices in linear models with high-dimensional responses. We systematically study the theoretical properties of the proposed test. We first derive the optimal projection matrix for any given projection dimension to achieve the best power and provide an upper bound for the optimal dimension of projection matrix. We further provide insights into how to construct the optimal projection matrix. One- and two-sample mean problems can be formulated as special cases of linear hypotheses studied in this article. We both theoretically and empirically demonstrate that the proposed test can outperform the existing ones for one- and two-sample mean problems. We conduct Monte Carlo simulation to examine the finite sample performance and illustrate the proposed test by a real data example.

ARTICLE HISTORY
Received December 2019
Accepted January 2021

KEYWORDS
Hotelling T^2 test; Multiple sample mean test; Projection test; Two-sample mean test

1. Introduction

High-dimensional data have been collected in a wide range of studies and applications from various fields such as genomics, finance, social science, and signal processing. Over the last two decades, one-sample and two-sample mean testing for high-dimensional data received considerable attentions in the literature. For fixed or low-dimensional data, the Hotelling T^2 test (Hotelling 1931) for the one-sample and two-sample mean problem is the uniformly most powerful test invariant under affine transformations. When the dimension p of observations is greater than the sample size n, the Hotelling T^2 test becomes inapplicable because of the singularity of the sample covariance matrix. To deal with the singularity, various tests for high-dimensional one- and two-sample mean testing have been proposed. Bai and Saranadasa (1996) first studied the impact of dimensionality on two-sample mean testing, and advocated sum-of-square-type statistics on the mean difference while ignoring correlation matrices. Their test avoids the problem caused by the rank deficiency of the sample correlation/covariance. Also see Chen and Qin (2010), Srivastava and Du (2008), and Zhang et al. (2020) for further study and extension. To deal with sparse alternatives, Cai, Liu, and Xia (2014) proposed a supremum-type statistic, and Chen, Li, and Zhong (2019) proposed a sum-of-square-type statistic together with a hard-threshold method. Furthermore, Fan, Liao, and Yao (2015) proposed a power enhancement principle in high-dimensional testing problems, especially for sparse alternatives. Also see Kock and Preinerstorfer (2019) for further discussion and extension. To achieve high power against various alternatives, Xu et al. (2016) proposed a sum-of-power-type statistic on the mean difference with an adaptive power index, and He et al. (2021) further proposed an adaptive testing procedure which combines p-values computed from the U-statistics of different orders. Regarding to covariance structures, Aoshima and Yata (2018) and Wang and Xu (2018) proposed tests specifically designed for strongly spiked covariance structures based on eigenvalue decomposition of the sample covariance. Xue and Yao (2020) proposed a distribution and correlation-free two-sample mean test, which is built upon a two-sample central limit theorem in high dimensions. To incorporate covariance structures into the test statistics, Lopes, Jacob, and Wainwright (2011), Thulin (2014), and Srivastava, Li, and Ruppert (2016) proposed various random projection methods. Although these tests are quite powerful under certain scenarios, they do not fully use the information of correlation among the variables. The proposals in Bai and Saranadasa (1996), Srivastava and Du (2008), and Chen and Qin (2010) ignore information of the correlation. The methods proposed in Cai, Liu, and Xia (2014), Xu et al. (2016), Aoshima and Yata (2018), Wang and Xu (2018), and Chen, Li, and Zhong (2019) can only utilize some particular kinds of covariance structures by data transformation methods or eigenvalue decomposition approaches, while the procedures developed in Lopes, Jacob, and Wainwright (2011), Thulin (2014), and Srivastava, Li, and Ruppert (2016) only use part of the correlation information preserved randomly.

This article aims to develop tests for the linear hypothesis in the following linear model.

$$Y = XB + E, \tag{1}$$

where Y is an $n \times p$ high-dimensional response matrix, that is, p is large, X is an $n \times d$ design matrix whose dimension d is fixed, B is a $d \times p$ regression coefficient matrix, and E is an $n \times p$ error matrix with mean zero and $\text{cov}(\text{vec}(E)) = I_n \otimes \Sigma$ for a positive definite matrix Σ. Here and hereafter $\text{vec}(A) = (a_1^T, \ldots, a_n^T)^T$ for $A = (a_1, \ldots, a_n)^T$. That is, we vectorize A by

CONTACT Runze Li rzli@psu.edu Department of Statistics, Pennsylvania State University at University Park, PA 16802-2111.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

its rows rather than its columns. We are interested in testing the following linear hypothesis with high-dimensional p.

$$H_0 : A_0 B = 0 \quad \text{versus} \quad H_1 : A_0 B \neq 0, \tag{2}$$

where A_0 is an $m \times d$ known constant matrix, m is fixed as well as d, and A_0 is of full row rank, which implies $m \leq d$. With properly specifying X and A_0, the one-sample and two-sample mean tests can be formulated as special cases of (2). Thus, the proposed procedure in this article is directly applicable to the one-sample and two-sample mean problems.

In this article, we propose a projection test for (2) under the setting (1). The projection test uses the correlation information to improve the power of the test. To construct the projection test, we first project the data into low dimension and then carry out the test on the projected data. We show that there exist low-dimensional projection matrices which use the information of covariance to make the power of projection tests optimal. We prove that the dimension of the optimal projections is not greater than m, and further derive the m optimal projection directions to achieve the best power. To carry out projection tests, we can use a sample-splitting procedure, in which we use the first part of the sample for estimation of projection matrices and the second part for testing. We further propose U-projection tests to improve the power of the test based on the sample-splitting procedure by constructing a U-type test statistic using all samples. We establish the theoretical properties of the U-projection test. For the one- and two-sample mean problems, we show that by using correlation information, the U-projection test can be much more powerful than some existing two-sample tests in the presence of high correlation like compound symmetry covariance structures theoretically. We also show that the U-projection test has the same asymptotic power with the test proposed in Chen and Qin (2010) in the presence of low correlation among variables. In addition to one-sample and two-sample mean problems, the proposed U-projection test can be directly applied to other testing problems such as multi-sample mean testing and predictor significance testing in high-dimensional settings, which are also important statistical problems. We examine the finite sample performance of the proposed U-projection test via Monte Carlo simulation studies, and illustrate the proposed methodology by an empirical analysis of a gene expression microarray dataset of murine heart under the conditions of cigarette smoke and obesity.

The remaining of the article is organized as follows. In Section 2, we propose the projection test with the optimal projection direction, and further propose the framework of the U-projection test and establish its asymptotic properties. In Section 3, we evaluate the performance of the proposed U-projection test on various simulated and real datasets. A brief discussion is given in Section 4, and technical proofs are given in the online appendix.

2. Projection Test on Linear Hypothesis

Consider model (1) and linear hypothesis (2). For $n > p$, we can carry out tests like the likelihood ratio test (LRT) under normality assumption on E. However, for $p \gg n$, the LRT cannot be applied directly. A natural way to accommodate this is to project Y into low dimensional space and then to carry on the test on the projected data.

2.1. Optimal Projection Direction

Let P be a $p \times r$ full column rank matrix with $r \ll p$ and $r < n$, consider the P-projected model:

$$Y^\star = XB^\star + E^\star, \tag{3}$$

where $Y^\star = YP$, $B^\star = BP$, and $E^\star = EP$, which is the $n \times r$ projected error matrix with mean zero and $\text{cov}(\text{vec}(E^\star)) = I_n \otimes (P^T \Sigma P)$.

The corresponding projected hypothesis becomes

$$H_{0P} : A_0 B^\star = 0 \quad \text{versus} \quad H_{1P} : A_0 B^\star \neq 0, \tag{4}$$

and the projected test statistic is

$$\Lambda = \frac{|G_P|}{|G_P + H_P|}, \tag{5}$$

where G_P is the residual sum of squares under H_{1P}, $G_P + H_P$ is the residual sum of squares under H_{0P}. More specifically, $G_P = P^T Y^T (I_n - P_{H_1}) YP$, $P_{H_1} = X(X^T X)^{-1} X^T$, and $G_P + H_P = P^T Y^T (I_n - P_{H_0}) YP$, $P_{H_0} = XA_1(A_1^T X^T X A_1)^{-1} A_1^T X^T$, and A_1 is a $d \times (d-m)$ matrix defined by $A_1 = (A_0^T)^\perp$, which means that (A_0^T, A_1) forms an orthogonal matrix. The test is the LRT under normality assumption and is equivalent to many useful test in various cases, like Hotelling T^2 in the one-sample mean testing.

The projected null hypothesis H_{0P} is rejected for small Λ, and H_0 is rejected if H_{0P} is rejected. In general, H_0 and H_{0P} are not equivalent. In this article, we will show that there exists an optimal projection direction P which makes H_{0P} equivalent to H_0 and also maximizes the power of the test (5). See Remark 1 for more details.

The problem how to construct an optimal direction can be divided into two subproblems: one is to find the dimension r of the optimal projection direction P, and the other is to find the optimal P of a particular dimension. These issues are addressed in Theorem 1. We assume the following conditions on the multivariate linear model (1) to derive the asymptotically optimal projection direction matrix.

Condition 1: There exists a positive constant C_1 such that $\|XX^T\|_\infty < C_1$, where $\|A\|_\infty = \max_{i,j} |a_{i,j}|$ for $A = (a_{i,j})_{i,j}$.
Condition 2: The limit $\lim_{n\to\infty} n^{-1}(X^T X) = M_X$ exists, where M_X is a nonsingular $d \times d$ matrix.
Condition 3: There exists a positive constant C_2 such that $\mathbb{E}(E_{ij}^4) < C_2$, $i = 1, \ldots, n$, $j = 1, \ldots, p$ for elements $E_{i,j}$ in the error matrix.

Conditions 1–3 are quite mild and are used to guarantee asymptotic normality of least squares estimate in linear models. For example, Conditions 1 and 2 hold in the one-sample mean testing problem automatically, and they also hold in the two-sample mean testing problem if $\frac{n_1}{n_1+n_2} \to \kappa \in (0, 1)$, where n_1 and n_2 are sample sizes of the first and the second samples, respectively. See Example 2 in this section.

Theorem 1. Under Conditions 1–3, the following statements are valid.

(A) Suppose that the error matrix E follows a matrix normal distribution.

(A1) Let $W = \Sigma^{-1/2}B^TX^T(I_n - P_{H_0})XB\Sigma^{-1/2}$, where $P_{H_0} = XA_1(A_1^TX^TXA_1)^{-1}A_1^TX^T$, and A_1 is a $d \times (d-m)$ matrix which is defined by $A_1 = (A_0^T)^{\perp}$. W is a $p \times p$ matrix of rank m. Suppose W admits the eigenvalue decomposition $W = \sum_{i=1}^{m} \lambda_{0i}R_iR_i^T$, where $0 < \lambda_{0m} \leq \cdots \leq \lambda_{01}$ are m nonzero eigenvalues, and R_i, $i = 1, \ldots, m$, are the corresponding orthogonal eigenvectors. Then for any given $r \leq m$, $P_0 = \Sigma^{-1/2}\left(R_1, \ldots, R_r\right)$ is the projection direction matrix which is optimal among all $p \times r$ matrices for the hypothesis testing problem (4).

(A2) The optimal dimension of projection matrix is less than or equal to m for the hypothesis testing problem (4).

(A3) Set $r = m$, $\Sigma^{-1}B^TA_0^T$ is the projection direction matrix which is optimal among all $p \times m$ matrices for the hypothesis testing problem (4).

(B) Without normality assumption, the statements in (A1)–(A3) are still valid in asymptotic sense by changing optimal projection to asymptotically optimal projection which maximizes the asymptotic local power.

Theorem 1(A1) gives the form of the optimal projection direction matrix of a given dimension. Theorem 1(A2) tells us there is no need to use projection direction matrices of more than m columns. Theorem 1(A3) gives a simple form of the optimal $p \times m$ projection direction matrix. When m is small as in the one-sample and two-sample mean testing problems ($m = 1$), we advocate to set $r = m$, and use the projection direction $\Sigma^{-1}B^TA_0^T$ provided by Theorem 1(A3). Theorem 1(B) confirms that even without the multinormality assumption, the results in Theorem 1(A) are still valid asymptotically.

Remark 1. From Theorem 1(A3), the optimal project direction $P = \Sigma^{-1}B^TA_0^T$ when $r = m$. Then the linear hypothesis in (4) becomes H_{0P} : $A_0B\Sigma^{-1}B^TA_0^T = 0$ versus H_{0P} : $A_0B\Sigma^{-1}B^TA_0^T \neq 0$. Since Σ^{-1} is supposed to be positive definite, the linear hypothesis in (4) with the optimal projection direction is equivalent to the original linear hypothesis (2).

In the following examples, we want to apply Theorem 1 in several specific types of hypothesis testing.

Example 1 (One-sample mean testing). For one-sample mean testing, suppose we have n samples Y from a p-dimensional multivariate distribution with mean μ and covariance Σ, and we want to test whether $\mu = 0$. It is easy to reformulate the problem as

$$Y = 1_nB + E, \tag{6}$$

where 1_n is a column vector full of 1 of length n, $B = \mu^T$, E is the $n \times p$ random error matrix, and the null hypothesis becomes H_0: $B = 0$. In this problem, $A_0 = 1$ is of rank 1. By Theorem 1, the optimal projection dimension is 1 and the optimal projection direction is $\Sigma^{-1}\mu$.

Example 2 (Multi-sample mean testing). For multi-sample mean testing problem, suppose for $k = 1, 2, \ldots, K$, we have n_k samples Y_k from a p-dimensional multivariate distribution F_k with mean μ_k and covariance Σ, and we want to test whether $\mu_1 = \mu_2 = \cdots = \mu_K$. The problem can be reformulated as

$$\begin{pmatrix} Y_1 \\ \cdots \\ Y_K \end{pmatrix} = \begin{pmatrix} 1_{n_1} & 0 & \cdots \\ \cdots & \cdots & \cdots \\ 0 & \cdots & 1_{n_K} \end{pmatrix} B + E, \tag{7}$$

where $B = (\mu_1, \ldots, \mu_K)^T$, and E is the $(\sum_{k=1}^{K} n_k) \times p$ random error matrix. We want to test H_0: $A_0B = 0$, where $A_0 = \{a_{ij}\}_{1\leq i\leq K-1, 1\leq j\leq K}$ with $a_{ii} = K-1$ and $a_{ij} = -1$, $i \neq j$. In this problem, A_0 is of rank $K-1$. By Theorem 1, the optimal projection dimension is less than or equal to $K-1$ and the optimal $p \times (K-1)$ projection direction matrix is $K\Sigma^{-1}(\mu_1 - \bar{\mu}, \mu_2 - \bar{\mu}, \ldots, \mu_{K-1} - \bar{\mu})$, where $\bar{\mu} = \frac{1}{K}\sum_{k=1}^{K} \mu_k$. In particular, when $K = 2$, the optimal projection dimension is 1 and the optimal projection direction is $\Sigma^{-1}(\mu_1 - \mu_2)$.

Example 3 (Testing of significance of predictors). Consider the linear model (1). Of interest is to test the significance of a certain predictor, $X^{(i)}$. That is, consider

$$H_0 : A_0B = 0 \qquad H_1 : A_0B \neq 0, \tag{8}$$

where $A_0 = e_i^T$ and e_i is a d-dimensional column vector with the ith element one and other elements zero. Note that in the normal case, testing whether the regression coefficients equal to zero or not is equivalent to testing of conditional independence, so the above hypothesis testing problem is also testing for the conditional independence of Y and $X^{(i)}$ given other $(d-1)$ predictors under normality assumption. From Theorem 1, the optimal projection dimension is 1 and the optimal projection direction is $\Sigma^{-1}B^Te_i$.

2.2. Estimation of Optimal Projection Direction

Theorem 1 provides us insights into how the optimal projection direction depends on Σ and B. We will develop an estimation procedure for the optimal direction. For hypothesis (2) with small m, as for the one-sample and two-sample mean testing problems, we advocate set $k = m$, and the optimal projection direction is $\Sigma^{-1}B^TA_0^T$ according to Theorem 1 (A2) and (A3). The optimal projection matrix can be estimated column by column as follows.

$$P_{0,k} = \Sigma^{-1}B^TA_{0,k}^T, \quad \text{for} \quad k = 1, \ldots, m, \tag{9}$$

where $A_{0,k}$ is the kth row of A_0, and $P_{0,k}$ is the kth column of P_0. So without loss of generality, in this subsection we assume $m = 1$ and treat A_0 and P_0 as vectors. Hence, we focus on estimating the optimal projection direction P_0:

$$P_0 = \Sigma^{-1}B^TA_0^T. \tag{10}$$

A natural approach for estimating the optimal projection direction is to use estimators $\hat{B}$ and $\hat{\Sigma}$ instead of B and Σ:

$$\hat{P} = \hat{\Sigma}^{-1}\hat{B}^TA_0^T. \tag{11}$$

However, $\hat{P}$ in (11) is not well defined when $\hat{\Sigma}$ is singular as in the case of sample covariance when p is larger than n. To deal with this problem, we can use a shrinkage method similar to ridge regression:

$$\hat{P}_{\text{ridge}} = (\lambda_0 I_p + \hat{\Sigma})^{-1}\hat{B}^T A_0^T, \tag{12}$$

where $\lambda_0 > 0$.

Projection test provides us with much flexibility in choosing $\hat{B}$ and $\hat{\Sigma}$. If we have some prior knowledge on B or Σ, we can use the prior knowledge to estimate them efficiently to construct a good estimate of the optimal projection direction and a powerful projection test. Here if we do not have any prior knowledge on the linear regression coefficient B, we use the least squares estimator $\hat{B}_{\text{LS}}$ as follows:

$$\hat{B}_{\text{LS}} = (X^T X)^{-1} X^T Y. \tag{13}$$

If we do not have any prior knowledge on Σ, we use the sample covariance of the residuals to estimate Σ given $\hat{B}$, so

$$\hat{\Sigma} = \frac{1}{n-d}(Y - X\hat{B})^T (Y - X\hat{B}), \tag{14}$$

where n is the sample size. This choice makes our conclusion robust to the assumptions on Σ. Moreover, we have

$$\hat{\Sigma}_{\text{LS}} = \frac{1}{n-d}(Y - X\hat{B}_{\text{LS}})^T (Y - X\hat{B}_{\text{LS}}). \tag{15}$$

Although it can bring more power to the test by having prior knowledge on B and Σ and using particular estimators for them, we propose using the general estimators of B and Σ in this article to reflect the general case. From $\hat{B}_{\text{LS}}$, $\hat{\Sigma}_{\text{LS}}$, we have

$$\hat{P}_{\text{LS}} = (\lambda_0 I_p + \hat{\Sigma}_{\text{LS}})^{-1}\hat{B}_{\text{LS}}^T A_0^T, \tag{16}$$

where $\lambda_0 > 0$ and n is the sample size. It is of interest to establish the theoretical properties of $\hat{P}_{\text{LS}}$. But this would require very technical treatments using high-dimensional random matrix theory, and is beyond the scope of this article. In general, $\hat{P}_{\text{LS}}$ might not be a consistent estimator for the optimal projection direction in high-dimensional settings. However, as shown in Theorem 2 in the next section, tests based on $\hat{P}_{\text{LS}}$ have the potential to bring power improvements under high correlation covariance structures.

2.3. Sample-Splitting Projection Test and U-Projection Test

In her unpublished dissertation, Huang (2015) introduced a sample-splitting method for estimating the optimal projection direction for *one-sample and two-sample mean testing*. The sample-splitting method randomly partitions the data into two parts, one part for the projection direction estimation, and the other part for the projection test. Since the data used for estimation of the projection direction is independent of the data for the projection test, the resulting projection test retains Type I error rate very well. The sample-splitting method may be used for our current setting, but there are several disadvantages of the sample-splitting projection test. (a) Even for a given sample, the sample-splitting projection test is random in nature because of the sample splitting procedure. (b) In sample splitting, we use part of the samples to estimate the projection direction and then use the projection test on the remaining samples. We have some power loss because of sample splitting. (c) The sample-splitting portion is hard to determine in general, because it requires the knowledge of the unknown true parameter. To deal with these issues, we propose the U-projection test. The key idea is to construct a test statistic similar to U-statistic with the kernel of the sample-splitting projection test statistic. We first construct the U-projection statistic in the one- and two-sample mean testing problems, and then we will construct the U-projection statistic in the general case similarly.

2.3.1. High-Dimensional One- and Two-Sample Mean Problems

Suppose there are n independent samples from a population F and Y is the $n \times p$ data matrix. We want to test $H_0 : \mu = 0_p$ versus $H_a : \mu \neq 0_p$. The sample-splitting test procedure randomly partitions the data into two parts Y_1 and Y_2, whose sample sizes are k and $n-k$, respectively. The first sample is used to obtain an estimator of the optimal projection direction $\hat{P}_{Y_1}$, which is a vector in the one-sample mean testing problem. We can then conduct a test for $H_0 : \mu = 0$ based on the projected second sample $Y_2^T \hat{P}_{Y_1}$. Note that t-test is equivalent to the likelihood ratio test for univariate normal mean test. Thus, one may directly apply the t-test for the projected second sample to test $H_0 : \mu = 0$.

The main drawbacks of the sample-splitting test are related to the random sample-splitting part. To overcome this issue, we can construct a statistic similar to U-statistic, which is the average of all possible sample-splitting test statistic with a fixed k. This results in the proposed U-projection statistic in the one-sample mean problem:

$$U_P = \frac{1}{\binom{n}{k}} \sum_{\gamma \in \Gamma} \bar{Y}_{-\gamma}^T \hat{P}_{Y_\gamma}, \tag{17}$$

where Γ is the collection of all size-k subsets of $\{1, 2, \ldots, n\}$, and Y_γ, $Y_{-\gamma}$ are subsamples of Y with and without index γ, respectively. In the construction of U_P, we use the projected sample mean instead of the LRT statistic (or equivalently t-statistic) based on the projected sample. This will be applied for two-sample mean testing problems below, and leads to a better performance than the projected LRT. See more details in the discussion after Equation (18).

From the construction of the sample-splitting test statistic and the U-projection statistic (17), it can be seen that the expectation of these two statistics should be same and that the variance of the U-projection statistic (17) should be smaller than or equal to that of the sample-splitting test statistic with the same k as the U-projection statistic U_P. In fact, $\text{var}(U_P) \leq \frac{1}{\binom{n}{k}} \sum_{\gamma \in \Gamma} \text{var}(\bar{Y}_{-\gamma}^T \hat{P}_{Y_\gamma}) = \frac{1}{\binom{n}{k}} \sum_{\gamma \in \Gamma} \text{var}(S_P) = \text{var}(S_P)$, where $S_P = \bar{Y}_{-\gamma}^T \hat{P}_{Y_\gamma}$, which is the sample splitting test statistic with the same k as U_P. By taking the average of sample-splitting test statistics, the U-projection statistic (17) solves the issues of the sample-splitting projection test, that is, its randomness and loss of power due to sample splitting. Thus, it is expected that the U-projection test is more powerful than the sample-splitting projection test.

Similarly, we have the U-projection statistic for two-sample mean testing problems. Suppose we have n_1 samples Y_1 from the first population, n_2 samples Y_2 from the second population, and

Table 1. Empirical Type I error rates and powers of two-sample mean tests with compound-symmetry covariances (in percentage).

θ	ρ	n	p	BS	CQ	SD	CLX	CLZ	XLWP	LJW	XY	HXWP	New
0	0.5	50	500	4.0	4.4	4.5	4.3	4.3	4.3	6.0	4.4	6.4	5.7
		100	500	4.9	4.8	4.6	5.0	5.4	5.3	6.9	4.3	6.6	4.3
		50	1000	5.6	5.6	5.7	4.7	5.7	5.2	4.4	3.8	6.0	3.5
		100	1000	3.5	3.7	3.9	4.4	3.8	4.1	6.3	4.9	6.6	3.9
0	0.8	50	500	5.2	5.1	5.1	5.6	4.7	5.2	4.7	5.2	6.1	7.0
		100	500	4.6	4.1	4.6	4.5	4.6	4.6	4.8	4.0	5.6	5.1
		50	1000	4.2	3.9	4.2	4.5	4.3	4.1	6.9	4.9	6.0	5.4
		100	1000	3.3	3.9	3.8	3.8	4.0	4.2	6.4	5.2	7.1	4.6
0.5	0.5	50	500	7.5	8.3	7.7	19.1	8.1	17.0	44.1	18.1	9.8	100.0
		100	500	5.4	5.7	5.8	16.1	5.8	14.3	66.7	19.1	11.1	99.8
		50	1000	8.1	7.3	7.7	19.3	7.4	16.7	43.4	18.0	9.1	100.0
		100	1000	7.4	7.1	7.4	20.8	7.4	17.1	61.0	20.7	12.6	100.0
0.5	0.8	50	500	6.4	6.8	6.6	20.7	7.0	18.2	87.6	22.7	12.9	100.0
		100	500	6.3	6.3	5.8	22.2	5.8	20.0	99.3	19.9	10.7	100.0
		50	1000	5.9	5.6	6.0	22.2	6.3	20.8	87.2	26.3	15.3	100.0
		100	1000	6.3	6.0	5.9	23.8	6.2	20.9	99.1	25.6	15.5	100.0

Table 2. Empirical Type I error rates and powers of two-sample mean tests with autoregressive covariances (in percentage).

θ	ρ	n	p	BS	CQ	SD	CLX	CLZ	XLWP	LJW	XY	HXWP	New
0	0.5	50	500	6.0	5.5	5.0	4.7	5.7	5.6	3.1	5.2	6.0	4.9
		100	500	4.1	4.4	4.2	4.2	5.2	4.5	4.6	4.3	6.7	5.8
		50	1000	4.2	4.5	4.4	4.2	4.4	4.6	5.0	5.6	6.3	3.9
		100	1000	4.8	4.7	5.2	4.6	5.3	4.8	4.1	4.9	6.8	4.2
0	0.8	50	500	5.8	5.4	5.5	5.1	5.0	4.6	5.3	5.2	6.6	4.8
		100	500	5.0	5.1	5.3	4.6	4.8	5.4	5.5	5.3	7.4	6.4
		50	1000	4.3	4.2	4.3	5.7	5.0	5.1	4.7	5.4	5.3	3.6
		100	1000	5.2	5.7	4.6	5.3	4.8	4.3	5.7	4.5	7.1	5.7
0.5	0.5	50	500	88.2	88.0	87.0	19.4	73.6	75.1	21.3	19.9	64.7	88.2
		100	500	87.4	87.4	87.4	21.5	75.3	73.2	32.0	22.1	66.5	88.1
		50	1000	99.5	99.6	99.2	22.1	95.7	97.3	19.8	22.8	91.4	98.4
		100	1000	99.2	99.3	99.4	23.0	95.9	96.6	26.7	26.7	93.0	98.4
0.5	0.8	50	500	51.5	52.9	52.5	21.6	39.6	34.2	26.4	25.1	30.9	94.0
		100	500	51.9	52.0	52.4	23.6	41.5	36.7	59.3	22.7	32.1	99.6
		50	1000	79.1	79.1	79.2	24.4	67.4	60.2	21.7	24.2	52.2	94.9
		100	1000	79.6	78.4	78.8	27.0	65.4	59.4	44.0	27.1	55.2	99.9

all samples are independent. Furthermore, suppose we choose k_i samples from Y_i for $i = 1, 2$ to estimate the projection direction, the two-sample U-projection statistic becomes

$$U_P = \frac{1}{\binom{n_1}{k_1}\binom{n_2}{k_2}} \sum_{\gamma_1 \in \Gamma_1} \sum_{\gamma_2 \in \Gamma_2} (\bar{Y}^T_{-\gamma_1,1} - \bar{Y}^T_{-\gamma_2,2}) \hat{P}_{Y_{\gamma_1,1}, Y_{\gamma_2,2}}, \quad (18)$$

where for $i = 1, 2$, Γ_i is the collections of all subsets of $\{1, \ldots, n_i\}$ with size k_i, $Y_{\gamma,i}$ and $Y_{-\gamma,i}$ are subsets of Y_i with and without index γ correspondingly, and $\hat{P}_{Y_{\gamma_1,1}, Y_{\gamma_2,2}}$ is the projection direction estimated using $Y_{\gamma_1,1}$ and $Y_{\gamma_2,2}$. Similar to U_P for one-sample mean problem, we use the projected sample mean in the construction of U_P instead of the t-test statistic based on the projected sample. From our simulation, we find this U_P has slightly higher power than the LRT based on the projected sample under most scenarios in Tables 1–3.

We next study the asymptotic properties of U-projection test statistic and its connection to other testing methods for the one- and two-sample mean testing problems. In the two-sample mean testing problem, suppose we use the projection direction

$$\hat{P}_{\mathrm{LS}} = (\lambda_0 I_p + \hat{\Sigma})^{-1} \hat{B}^T_{\mathrm{LS}} A^T_0 = (\lambda_0 I_p + \hat{\Sigma})^{-1} (\bar{Y}_1 - \bar{Y}_2), \quad (19)$$

the U-projection statistic becomes

$$U_{\mathrm{LS}} = \frac{1}{\binom{n_1}{k_1}\binom{n_2}{k_2}} \sum_{\gamma_1 \in \Gamma_1} \sum_{\gamma_2 \in \Gamma_2} (\bar{Y}^T_{-\gamma_1,1} - \bar{Y}^T_{-\gamma_2,2}) \times \left\{ \lambda_0 I_p + \frac{(k_1 - 1) S_{Y_{\gamma_1,1}} + (k_2 - 1) S_{Y_{\gamma_2,2}}}{k_1 + k_2 - 2} \right\}^{-1} \times (\bar{Y}_{\gamma_1,1} - \bar{Y}_{\gamma_2,2}), \quad (20)$$

where for $i = 1, 2$, k_i is the number of independent samples from Y_i to estimate the projection direction, Γ_i is collections of all subsets of $\{1, \ldots, n_i\}$ with size k_i, and $Y_{\gamma,i}$ and $Y_{-\gamma,i}$ are subsets of Y_i with and without index γ correspondingly. Under the special case that $\Sigma = I_p$, the optimal projection direction $\Sigma^{-1}(\mu_1 - \mu_2) = \mu_1 - \mu_2$, and it is natural to estimate the optimal projection direction by $\hat{P}_{\mathrm{N}} = \bar{Y}_1 - \bar{Y}_2$. Furthermore, with the projection direction P_{N}, the U-projection statistic becomes

$$U_{\mathrm{N}} = \frac{1}{\binom{n_1}{k_1}\binom{n_2}{k_2}} \sum_{\gamma_1 \in \Gamma_1} \sum_{\gamma_2 \in \Gamma_2} (\bar{Y}^T_{-\gamma_1,1} - \bar{Y}^T_{-\gamma_2,2})(\bar{Y}_{\gamma_1,1} - \bar{Y}_{\gamma_2,2}), \quad (21)$$

Table 3. Empirical powers of two-sample mean tests with sparse alternatives and $\rho = 0.8$ and $\theta = 1.0$ (in percentage).

Σ	s	n	p	BS	CQ	SD	CLX	CLZ	XLWP	LJW	XY	HXWP	New
CS	0.4	50	500	5.4	5.5	5.5	12.5	5.3	11.5	10.6	13.8	7.5	47.1
		100	500	5.8	5.8	5.5	14.2	6.0	13.0	14.6	13.5	9.0	39.6
		50	1000	5.4	4.6	4.4	12.3	4.7	11.8	9.6	12.0	7.3	42.8
		100	1000	5.1	5.2	5.3	11.9	5.0	10.6	11.1	11.8	8.1	44.5
	0.7	50	500	6.3	6.6	6.2	37.2	6.2	32.6	59.7	38.0	22.3	100.0
		100	500	4.3	4.6	4.6	35.0	3.9	32.6	87.0	38.4	24.6	100.0
		50	1000	4.4	5.0	4.8	37.3	5.0	33.7	49.0	41.3	25.5	100.0
		100	1000	5.7	5.3	6.0	37.8	5.7	34.8	76.2	39.8	26.0	100.0
AR	0.4	50	500	7.9	7.2	7.2	8.0	7.3	7.7	7.1	9.1	8.1	10.8
		100	500	7.1	6.6	6.2	8.9	6.2	7.4	8.7	8.5	8.5	13.5
		50	1000	5.9	6.8	6.0	7.8	5.5	7.0	5.2	8.9	8.9	8.1
		100	1000	5.5	4.8	5.2	7.7	5.6	7.0	6.3	7.7	8.5	7.0
	0.7	50	500	26.8	26.3	26.4	25.1	26.1	25.1	18.4	28.6	21.9	63.9
		100	500	28.8	27.7	27.4	28.4	26.6	27.4	34.5	27.0	24.2	91.2
		50	1000	31.2	31.0	30.6	24.8	27.4	27.8	12.5	27.5	26.1	48.8
		100	1000	33.1	34.3	33.4	28.0	31.7	29.4	20.2	28.1	28.8	71.4

which can be shown to be the same two-sample statistic T_{CQ} in Chen and Qin (2010). Hence, the test in Chen and Qin (2010) can be viewed as a U-projection test by pretending $\hat{\Sigma} = I_p$.

While Chen and Qin (2010) showed that T_{CQ} is quite powerful when the covariance is of low correlation structures as required by the condition that $\mathrm{tr}(\Sigma^4) = o(\mathrm{tr}^2(\Sigma^2))$, power of T_{CQ} can be diminished by the presence of high correlation in Σ. From Chen and Qin (2010), we know that $\frac{\mathrm{tr}(\Sigma^2)}{n^2} = O(\mathrm{var}(T_{\mathrm{CQ}}))$ and $\mathbb{E}(T_{\mathrm{CQ}}) = \|\mu_n\|^2$, where μ_n is the mean difference. Let $a_n \asymp b_n$ stand for that a and b have the order (i.e., $a_n/b_n = O(1)$ and $b_n/a_n = O(1)$). Under the high correlation case that $\lambda_{\max} \asymp p$, and local alternative hypothesis that $\|\mu_n\|^2 = o(p/n)$, it is easy to show that $\mathbb{E}(T_{\mathrm{CQ}}) = o(\sqrt{\mathrm{var}(T_{\mathrm{CQ}})})$, which suggests that T_{CQ} has no nontrivial power beyond the significance level under this high correlation scenario asymptotically. Unlike T_{CQ}, our proposed U_{LS} utilizes the covariance information, and Theorem 2 establishes the asymptotic normality of U_{LS} under certain high correlation covariance structures and shows that the test based on U_{LS} can be quite powerful under the high correlation covariance structure. Besides the advantage of U_{LS} in certain high correlation settings, Theorem 3 shows that the test U_{LS} and the test T_{CQ} in Chen and Qin (2010) have the same asymptotic power under low correlation covariance in high-dimensional settings.

For asymptotic normality of U_{LS}, we assume the following conditions on the error matrix E in the linear model (1):

Condition 4

$$E_i = \Gamma Z_i \quad \text{for } i = 1, \ldots, n, \tag{22}$$

where Γ is a $p \times t$ matrix with some $t \geq p$ such that $\Gamma\Gamma^T = \Sigma$, and $Z_i = (Z_{i,1}, \ldots, Z_{i,t})$ are t-variate independent and identically distributed random vectors satisfying $\mathbb{E}(Z_i) = 0$, $\mathrm{var}(Z_i) = I_t$, $\mathbb{E}(Z_{i,k}^3) = 0$, $\mathbb{E}(Z_{i,k}^6)$ is uniformly bounded, and

$$\mathbb{E}\left(Z_{i,l_1}^{\alpha_1} Z_{i,l_2}^{\alpha_2} \cdots Z_{i,l_s}^{\alpha_s}\right) = \mathbb{E}\left(Z_{i,l_1}^{\alpha_1}\right)\mathbb{E}\left(Z_{i,l_2}^{\alpha_2}\right)\cdots\mathbb{E}\left(Z_{i,l_s}^{\alpha_s}\right), \tag{23}$$

for a positive integer s such that $\sum_{l=1}^{s} \alpha_l \leq 8$ and $l_1 \neq l_2 \neq \cdots \neq l_s$.

Condition 5 $\|A_0 B\|_{\mathrm{F}}^2 = o(p/n)$.

Condition 4 is similar to conditions (3.1) and (3.2) in Chen and Qin (2010) under the linear model setting (1). Equation (22) says that the error matrix E can be expressed as a linear transformation of a random vector Z of length t (which is greater than or equal to p) with zero mean and unit variance, which allows a variety of structures for the covariance Σ. Equation (23) says that each vector Z_i has a kind of pseudo-independence among its elements, and the equation will be satisfied if elements of Z_i are independent. Condition 5 can be seen as the local alternative assumption under the linear model setting (1). In two-sample mean testing problems, it is easy to verify that Condition 5 is equivalent to $\|\mu_n\|_2^2 = o(p/(n_1 + n_2))$, where μ_n is the two-sample mean difference and n_i, $i = 1, 2$ are two sample sizes.

For high correlation covariance Σ with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p$, we assume the following condition:

Condition C1 There exists a uniformly bounded positive integer $q < p$ such that $\frac{\sqrt{n}\lambda_q}{\mathrm{tr}(\Sigma)} \to \infty$ and λ_{q+1} is uniformly bounded from above. The smallest eigenvalue λ_p is uniformly bounded from below.

The spiked covariance models studied in Johnstone (2001), Baik and Silverstein (2006), Wang and Fan (2017), and Donoho, Gavish, and Johnstone (2018) satisfy Condition C1, but Σ satisfying Condition C1 can be more general than the spiked models. For example, Condition C1 may be satisfied by the linear positive combination of compound-symmetry covariance and autoregressive covariance. Also note that the situation that $\lambda_{\max} \asymp \mathrm{tr}(\Sigma)$ which diminishes the power of T_{CQ} against local alternatives is also included in Condition C1.

Theorem 2. Suppose the covariance Σ satisfies Condition C1. Under high-dimensional setting $n_1 + n_2 = o(\mathrm{tr}(\Sigma))$, Conditions 3, 4, 5, and conditions that $k_i/n_i \to \gamma_i \in (0, 1)$ for $i = 1, 2$, $n_1/(n_1 + n_2) \to \kappa \in (0, 1)$, then U_{LS} has an asymptotically normal distribution with a uniformly bounded positive λ_0. More specifically, we have

$$\frac{\lambda_0(U_{\mathrm{LS}} - \mathbb{E}(U_{\mathrm{LS}}))}{\sigma_n} \xrightarrow{d} N(0, 1), \quad \text{and}$$
$$\lambda_0 \mathbb{E}(U_{\mathrm{LS}}) - \|W_{q+1}\mu_n\|_2^2 = o(\|W_{q+1}\mu_n\|_2^2), \tag{24}$$

where q is the number of divergent eigenvalues of Σ as specified in Condition C1, W_{q+1} is the projection matrix onto the linear span of eigenspaces of Σ corresponding to the smallest $p-q$ eigenvalues, $\lambda_{q+1}, \lambda_{q+2}, \dots, \lambda_p$, μ_n is the mean difference of the two populations, and $\sigma_n^2 = (\frac{2}{n_1^2} + \frac{2}{n_2^2} + \frac{4}{n_1 n_2}) \sum_{i=q+1}^{p} \lambda_i^2$.

The condition $k_i/n_i \to \gamma_i \in (0,1)$ for $i = 1,2$ means that enough samples need to be used for covariance estimation in the U-projection test. From the formula of σ_n^2 in Theorem 2, we can see that under local alternatives and high correlation covariances as in Condition C1, the divergent eigenvalues of Σ do not contribute to the asymptotic variance of U_{LS}, which is contrary to the case of T_{CQ} since $\text{var}(T_{\text{CQ}}) \asymp \text{tr}(\Sigma^2)/(n_1+n_2)^2$. It means that the high correlation structures specified in Condition C1 inflate the variance of T_{CQ} but not that of U_{LS}, which shows the advantage of U_{LS} under this high correlation scenario. From Theorem 2, we can calculate the asymptotic power of U_{LS} for local alternatives such that $\|\mu_n\|_2^2 = o(p/(n_1+n_2))$ by

$$\beta_n(\mu_n) = \Phi\left(-\xi_\alpha + \frac{\|W_{q+1}\mu_n\|_2^2}{\sigma_n}\right), \tag{25}$$

where α is the size of the test, and Φ is the standard normal distribution function. From Condition C1, we have $\sigma_n \asymp \frac{\sqrt{p}}{n_1+n_2}$ and it is easy to show that U_{LS} is consistent against local alternatives that satisfy $\frac{(n_1+n_2)\|W_{q+1}\mu_n\|_2^2}{\sqrt{p}} \to \infty$ under the conditions imposed by Theorem 2.

For low correlation covariance Σ, we assume the following condition:

Condition C2 $\text{tr}(\Sigma^4) = o(\text{tr}^2(\Sigma^2))$.

Condition C2 is the same condition on the correlation structure as in Chen and Qin (2010) when the two population covariances are the same. While Condition C2 is satisfied by low correlation covariance structures like autoregressive matrices where the maximum eigenvalue is bounded, the condition is not satisfied by high correlation covariance structures like compound-symmetry ones. Furthermore, Condition C2 cannot be satisfied by the high correlation covariance as specified by Condition C1 under the high-dimensional regime $n_1 + n_2 = O(\text{tr}(\Sigma))$. In fact, under Condition C1 and $n_1 + n_2 = O(\text{tr}(\Sigma))$, we have $\text{tr}(\Sigma^2) = O(p) + \sum_{i=1}^{q} \lambda_i^2 = O(\sum_{i=1}^{q} \lambda_i^2) = O(\sqrt{\sum_{i=1}^{q} \lambda_i^4}) = O(\sqrt{\text{tr}(\Sigma^4)})$, which is the opposite of Condition C2.

Theorem 3. Suppose the covariance Σ satisfies Condition C2. Under high-dimensional setting $n_1 + n_2 = O(\text{tr}(\Sigma))$, Conditions 3–5, and conditions that $k_i/n_i \to \gamma_i \in [0,1)$ for $i = 1,2$, $n_1/(n_1+n_2) \to \kappa \in (0,1)$, and

$$(k_1 + k_2) = o(\text{tr}(\Sigma)/\lambda_{\max}(\Sigma)^2), \tag{26}$$

where $\lambda_{\max}(\Sigma)$ is the largest eigenvalue of Σ, $\lambda_0 U_{\text{LS}}$ is asymptotically normally distributed and has the same asymptotic variance and similar expectation with T_{CQ},

$$\frac{\lambda_0(U_{\text{LS}} - \mathbb{E}U_{\text{LS}})}{\sqrt{\text{var}T_{\text{CQ}}}} \xrightarrow{d} N(0,1), \quad \text{and}$$
$$|\mathbb{E}(\lambda_0 U_{\text{LS}} - T_{\text{CQ}})| = o(\mathbb{E}T_{\text{CQ}}), \tag{27}$$

and U_{LS} and T_{CQ} have the same asymptotic power, $\beta_{U_{\text{LS}}}(\mu_1 - \mu_2) - \beta_{T_{\text{CQ}}}(\mu_1 - \mu_2) \to 0$.

The condition (26) restricts the product of $k_1 + k_2$ in the U-projection test and the largest eigenvalue of the covariance matrix. Under high correlation structures like the compound symmetry matrix, where the largest eigenvalue is of order $\text{tr}(\Sigma)$, there is no k_1 and k_2 satisfying the condition (26). Under low correlation covariance structures like autoregressive matrices, where the maximum eigenvalue is of order $O(1)$, the condition (26) can be satisfied by $k_1 + k_2 = o(\text{tr}(\Sigma))$, which is always true in the high-dimensional setting $n_1 + n_2 = o(\text{tr}(\Sigma))$.

From Theorem 3, it can be seen that under the low correlation case, the U-projection test cannot provide substantial improvement since its asymptotic power is the same with tests like Bai and Saranadasa (1996) and Chen and Qin (2010). To further appreciate Theorem 3, notice that for the low correlation structures, I_p can be a better estimation for Σ than the sample covariance. For example, in the high-dimensional setting $n = o(p)$, when $\Sigma = \{\rho^{|i-j|}\}_{1\le i,j\le p}$, $\rho \in (0,1)$, which is of autoregressive structures, then $\frac{\|\Sigma - I_p\|_{\text{F}}}{\sqrt{p}} = O(1)$ while $\frac{\|\Sigma - \hat{\Sigma}\|_{\text{F}}}{\sqrt{p}}$ goes to infinity, where $\|\cdot\|_{\text{F}}$ is the Frobenius norm. Hence, $\mu_1 - \mu_2$ can be more similar to the true optimal direction $\Sigma^{-1}(\mu_1 - \mu_2)$ than $(\lambda_0 I_p + \hat{\Sigma})^{-1}(\mu_1 - \mu_2)$ in direction, and T_{CQ} can use $P_{\text{N}} = \bar{Y}_1 - \bar{Y}_2$ as an estimator for the optimal projection direction $\Sigma^{-1}(\mu_1 - \mu_2)$ without much power loss.

2.3.2. U-Projection Test for General Cases

Now we can construct our U-projection statistic in general cases, which is an extension from the U-projection statistic (17) in the one-sample mean testing problem. Similar to the one-sample problem, it is helpful to consider the sample-splitting projection test statistic first. Consider the general multivariate linear model (1) and the hypothesis (2). From Theorem 1, $P_0 = \Sigma^{-1} B^T A_0^T$ is the asymptotic optimal projection matrix of dimension $p \times m$. The null hypothesis of the projection test with direction P_0 is $H_{0P}: A_0 B P_0 = A_0 B \Sigma^{-1} B^T A_0^T = 0$. Since $A_0 B \Sigma^{-1} B^T A_0^T$ is positive semidefinite, it is equivalent to the test $H_{0P}: \text{tr}(A_0 B P_0) = \text{tr}(A_0 B \Sigma^{-1} B^T A_0^T) = 0$. Following the sample-splitting test procedure, we split the data (X, Y) into two parts (X_1, Y_1) and (X_2, Y_2). The sample sizes of X_1, Y_1 are k, and the sample sizes of X_2, Y_2 are $n-k$. The first sample is used to obtain an estimator $\hat{P}_{(X_1,Y_1)}$ of the m-dimensional optimal projection direction P_0. The sample-splitting test statistic on the projected second sample is calculated as follows:

$$\text{tr}\left(A_0 \hat{B}_{(X_2,Y_2)} \hat{P}_{(X_1,Y_1)}\right) = \text{tr}\left(A_0 (X_2^T X_2)^{-1} X_2^T Y_2 \hat{P}_{(X_1,Y_1)}\right). \tag{28}$$

Note that for $\hat{P}_{(X_1,Y_1)}$ and $\hat{B}_{(X_2,Y_2)}$ to exist, X_1 and X_2 have to be of full column rank. Similar to the construction of U_P for the one-sample and two-sample mean problems, we do not directly use the likelihood ratio test based on the projected sample to construct U_P. Instead, we construct U_P via the estimate of $A_0 B P_0$ based on the projected sample.

From the construction of U-projection statistic (17) in the one sample problem, we can construct the statistic for U-

Table 4. Type I errors and empirical powers of U-projection test in multi-sample mean testing (in percentage).

			$\Sigma=\{\rho^{\vert i-j\vert}\}_{i,j}$				$\Sigma=(1-\rho)I_p+\rho 1_p 1_p^T$			
θ	ρ	n	p =125	250	500	1000	125	250	500	1000
0	0.5	50	6.0	6.3	4.4	4.8	4.6	5.0	5.3	4.9
		100	6.0	5.7	7.0	5.0	4.7	5.2	5.7	4.9
0	0.8	50	5.3	4.6	5.6	4.7	4.7	5.3	6.2	3.8
		100	3.7	5.2	4.6	4.4	5.0	5.3	4.9	5.0
0.5	0.5	50	90.4	97.2	99.8	100.0	98.9	100.0	100.0	100.0
		100	96.3	99.3	99.9	100.0	99.3	99.9	100.0	100.0
0.5	0.8	50	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
		100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 5. Type I errors and empirical powers of U-projection test in predictor significance testing (in percentage).

			$\Sigma=\{\rho^{\vert i-j\vert}\}_{i,j}$				$\Sigma=(1-\rho)I_p+\rho 1_p 1_p^T$			
θ	ρ	n	p =125	250	500	1000	125	250	500	1000
0	0.5	50	4.9	4.3	4.4	5.1	3.9	4.0	3.9	5.8
		100	4.7	4.7	5.0	4.8	5.7	5.4	4.7	5.1
0	0.8	50	5.3	5.4	5.3	4.5	4.5	4.9	5.2	4.9
		100	3.6	4.6	4.9	5.9	4.8	4.7	4.1	5.4
0.5	0.5	50	25.4	36.7	52.7	74.4	41.5	62.5	79.5	90.8
		100	33.7	45.2	64.3	88.3	46.6	72.4	94.8	99.6
0.5	0.8	50	13.2	14.4	13.6	17.3	40.6	61.5	79.4	88.6
		100	34.0	28.7	24.4	29.4	52.9	73.9	95.1	99.3

projection test in general cases similarly as follows:

$$U_P=\frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\operatorname{tr}\left(A_0(X_{-\gamma}^TX_{-\gamma})^{-1}X_{-\gamma}^TY_{-\gamma}\hat{P}_{(X_\gamma,Y_\gamma)}\right), \quad (29)$$

where $\Gamma=\{\gamma\mid\gamma\subset\{1,2,\dots,n\},|\gamma|=k,\operatorname{rank}(X_\gamma)>d,\operatorname{rank}(X_{-\gamma})\geq d\}$, and X_γ, Y_γ, $X_{-\gamma}$ $Y_{-\gamma}$ are subsamples of X and Y with and without index γ, respectively. From our simulation study, we find that the U_P defined in (29) performs slightly better than the likelihood ratio test based on projected sample under most scenarios in Tables 4 and 5.

The calculation of the U-projection statistic is illustrated in Algorithm 1.

Algorithm 1 General U-projection statistic

Input: $n\times d$ dimensional matrix X, $n\times p$ dimensional matrix Y, $m\times d$ dimensional matrix A_0, and $d<k\leq n-d$.
  Let $i=0$; $U=0$.
  for $\gamma\subset\{1,2,\dots,n\}$ and $|\gamma|=k$ **do**
    if X_γ are of rank greater than d and $X_{-\gamma}$ are of rank no less than d **then**
      Estimate the projection direction $\hat{P}_{(X_\gamma,Y_\gamma)}$ using X_γ and Y_γ.
      Calculate $U=U+\operatorname{tr}\left(A_0(X_{-\gamma}^TX_{-\gamma})^{-1}X_{-\gamma}^TY_{-\gamma}\hat{P}_{(X_\gamma,Y_\gamma)}\right)$.
      $i=i+1$.
    end if
  end for
  Calculate the U-projection statistic by $U_P=U/i$.

Note that in Algorithm 1 we loop over size-k subsets γ of $\{1,2,\dots,n\}$, so the algorithm will always end after finite iterations. We next study the asymptotic properties of U-projection test statistic in general cases.

Theorem 4. Consider $H_0: A_0B=0$ under model (1). Let

$$h_P(Z_{1,1},\dots,Z_{k+d,1};Z_{1,2},\dots,Z_{k+d,2})=\frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\operatorname{diag}\left(A_0(Z_{-\gamma,1}^TZ_{-\gamma,1})^{-1}Z_{-\gamma,1}^TZ_{-\gamma,2}\hat{P}_{(Z_{\gamma,1},Z_{\gamma,2})}\right), \quad (30)$$

where $-\gamma=\{1,\dots,k+d\}\backslash\gamma$, $\Gamma=\{\gamma,\gamma\subset\{1,2,\dots,k+d\},|\gamma|=k,\operatorname{rank}(Z_{\gamma,1})>d,\operatorname{rank}(Z_{-\gamma,1})=d\}$; $Z_{\gamma,i}$, $Z_{-\gamma,i}$ are samples of $Z_{\dots,i}$ with and without index γ for $i=1,2$; and $\hat{P}_{(Z_{\gamma,1},Z_{\gamma,2})}$ is the projection direction estimated with $(Z_{\gamma,1},Z_{\gamma,2})$. Then with the conditions that k is fixed and that $\mathbb{E}\operatorname{tr}\left(\operatorname{cov}(h_P(Z_{1,1},\dots,Z_{k+d,1};Z_{1,2},\dots,Z_{k+d,2}))\right)<C_0$ for some fixed $C_0>0$ and $(Z_{i,1},Z_{i,2})$, $i=1,\dots,k+d$ iid from the distribution of (X,Y), we have asymptotic normality of the general U-projection statistic (29):

$$\sqrt{n}(U_P-\mathbb{E}U_P)\xrightarrow{d}N(0,(k+d)^21_m^T\Xi_11_m), \quad (31)$$

where 1_m is a vector full of one with length m, and

$$\Xi_1=\operatorname{cov}(h_P(Z_{1,1},Z_{2,1},\dots,Z_{k+d,1};Z_{1,2},\dots,Z_{k+d,2}),\ h_P(Z_{1,1},Z'_{2,1},\dots,Z'_{k+d,1};Z_{1,2},Z'_{2,2},\dots,Z'_{k+d,2}) \quad (32)$$

for $(Z_{i,1},Z_{i,2})$, $i=1,\dots,k+d$ and $(Z'_{i,1},Z'_{i,2})$, $i=2,\dots,k+d$ iid from the distribution of (X,Y). Furthermore, under the null hypothesis, we have $\mathbb{E}U_P=0$.

2.4. p-Value From Randomization Test

Although the asymptotic distribution of U-projection statistic can be derived under certain conditions as in Theorems 2–4, it may not be very useful here because of several reasons: (a) The sample size n could be quite small which makes the asymptotic distribution far from the actual distribution. (b) The relationship between the asymptotic variance under the null and the covariance Σ is quite complex in general. For example, Theorem 2 and 3 provide different formulas to calculate the asymptotic variance under different structures of covariances. (c) The (asymptotic) distribution of the statistic also depends on the estimation method of the projection direction. If we use different estimation methods of the projection direction, the distribution will also change. Thus, the flexibility of choosing projection direction estimation method brings lots of difficulty to the distribution derivation of the U-projection statistic in (29). Table S.6 in the supplementary materials shows the empirical Type I error rates of the U-projection test for two-sample mean testing under two different covariance structures using asymptotic distributions derived from Theorems 2 and 3, and the details of the simulation settings are the same as described in Section 3.1. Due to the limitations of the asymptotic distributions, it is more appropriate to use the randomization method to calculate the p-value in the U-projection test. The randomization method relies on few assumptions on the distribution, and it also provides us the flexibility in choosing the kernel size and in choosing the estimation method of the projection direction. Algorithm 2 summarizes the p-value calculation procedure by randomization methods.

Algorithm 2 p-value calculated from randomization

Input: $n \times d$ dimensional matrix X, $n \times p$ dimensional matrix Y, $m \times d$ dimensional matrix A_0, $d < k \leq n - d$, and randomization times N.
Calculate the U-projection statistic U_0 on the original dataset X and Y.
Let $i = 0$.
for $j = 1 : N$ **do**
  Do randomization, get X' and Y'.
  Calculate the U-projection statistic U' on the randomized dataset X' and Y'.
  if $U' \geq U$ **then**
    $i = i + 1$.
  end if
end for
Calculate the p-value from $p = i/N$.

The randomization method to generate distribution under the null hypothesis depends on the hypothesis testing problem. For example, under the null hypothesis in the one-sample mean testing problem, $-Y_i$ has the same mean as the original sample Y_i. Thus, we can flip Y_i to $-Y_i$ randomly to generate randomly distributed dataset following the null distribution. In the two-sample or multi-sample mean testing problem, a permutation on the group can be used as the randomization method. In the predictor significance testing, a random shuffle on the index of response variable Y can be used as the randomization method.

3. Numerical Studies

In this section, we assess the finite sample performance of the proposed U-projection test via Monte Carlo simulation studies. In Section 3.1, we conduct numerical comparisons between the proposed test and existing ones for the two-sample mean problem. We examine the Type I error rate and power for multi-sample mean problem and test of significance of predictor in Sections 3.2 and 3.3, respectively.

To apply the U-projection test, we need to choose k_1, k_2, and λ_0 to use. We study how to select these parameters in the supplementary materials. In all the simulations and real data example, we take $k_i = \lfloor 0.9n_i \rfloor$ for $i = 1, 2$ for two-sample mean testing problems and $k = \lfloor 0.9n \rfloor$ for all other testing problems, where $\lfloor x \rfloor$ is the largest integer smaller than or equal to x. Furthermore, we set $\lambda_0 = \frac{1}{\sqrt{n_1+n_2-2}}$ in the two sample mean testing simulations and $\lambda_0 = \frac{1}{\sqrt{n-d}}$ in other testing problems where n is the sample size and d is the number of columns of X.

3.1. Two-Sample Mean Testing

We compare the performance of the proposed U-projection test with that of existing tests for the two-sample mean problem: Bai and Saranadasa (1996), Srivastava and Du (2008), Chen and Qin (2010), Lopes, Jacob, and Wainwright (2011), Cai, Liu, and Xia (2014), Xu et al. (2016), Chen, Li, and Zhong (2019), He et al. (2021), and Xue and Yao (2020), which are abbreviated as BS, SD, CQ, LJW, CLX, XLWP, CLZ, HXWP, and XY, respectively. Note that the test XLWP proposed by Xu et al. (2016) and the test HXWP proposed by He et al. (2021) are adaptive testing procedures which combine p-values of testing statistics of different orders. Following Xu et al. (2016) and He et al. (2021), we use statistics of orders $(1, 2, \ldots, 6, \infty)$ and the minimum p-value combination method for the XLWP and HXWP tests in our numerical comparisons. To this end, we generate random matrices Y_k, $k = 1, 2$, from $N_{n_k p}(1_{n_k}\mu_k^T, I_{n_k} \otimes \Sigma)$ and consider test $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$. In this simulation, we set $n = n_1 = n_2 = 50, 100$ and the dimension $p = 500, 1000$. We consider two covariance structures for Σ: (1) compound-symmetry structure $\Sigma_1 = (1 - \rho)I_p + \rho 1_p 1_p^T$ with $\rho = 0.2, 0.5, 0.8$ and (2) autoregressive structure Σ_2 with (i, j)-element being $\rho^{|i-j|}$ and $\rho = 0.2, 0.5, 0.8$. To make the problem challenging, we focus on local alternatives. Also, we will use both dense and sparse local alternatives to check the performance of the proposed U-projection test under various types of alternatives. Note that the samples in this simulation are generated from multivariate normal distributions, and we have also carried out simulations under non-Gaussian settings for both dense and sparse local alternatives. The corresponding non-Gaussian simulation results are put into Tables S.7–S.14 in the supplementary materials to save space. The simulation results under the non-Gaussian setting have similar patterns to those under the Gaussian setting.

Dense signal: For the dense type of local alternatives, we generate μ_k from a multivariate normal distribution in each simulation. That is,

$$\mu_k \sim N_p(0_p, (\theta^2/n)I_p) \tag{33}$$

for $k = 1, 2$. We set $\theta = 0$ to examine Type I error and $\theta = 0.5$ and 1.0 to examine empirical power.

Type I error rates and empirical powers for compound-symmetry covariance and autoregressive covariance based on 1000 replications are depicted in Tables 1 and 2, respectively. To save space, we present results with $\theta = 0, 0.5$ and $\rho = 0.5, 0.8$ in these two tables. Results for other scenarios are reported in Tables S.1 and S.2 in the supplementary materials of this article. Since the proposed U-projection test is performed using a randomization method, other existing tests are also performed using a randomization method for a fair comparison. The top two panels of Table 1 depict empirical Type I error rates at level 0.05 for the compound-symmetry covariance, and imply that with permutation test approaches, the proposed U-projection test and the other existing tests retain Type I error rates reasonably well across different sample sizes and dimensions under the compound-symmetry covariance, including the BS test, the CQ test, the SD test, the CLX test, the CLZ test, the LJW (random projection) test, the XLWP test, the XY test, and the HXWP test. Under the condition of same covariance, the BS test and the CQ test are actually the same test, and they had similar simulation results. From the bottom two panels of Table 1, it can be seen that under the compound-symmetry covariance, the U-projection test is much more powerful than the tests including BS, CQ, SD, CLX, CLZ, XLWP, LJW, XY, and HXWP. It illustrates that the U-projection test utilizes the correlation information, and can have good powers under high correlation cases. In Table 1, we can also notice that the power of the LJW (random projection) test is between the power of the other existing tests and that of the U-projection test. It implies that the LJW test only utilizes part of the correlation information by the random projection. Although with a much larger projection dimension, which is the integer part of $(n_1 + n_2 - 2)/2$ as suggested by Lopes, Jacob, and Wainwright (2011), the LJW test is still not as powerful as the U-projection test, which projects the high-dimensional data onto only one dimension for the two-sample mean testing problem. The top two panels of Table 2 depict empirical Type I error rates at level 0.05 for the autoregressive covariance. It implies that the proposed U-projection test and the other existing tests retain Type I error rates reasonably well across different sample sizes and dimensions under the autoregressive covariance. From the bottom two panels of Table 2, it can be seen that the tests of BS, CQ, SD, CLZ, XLWP, HXWP, and the proposed U-projection test all have relatively good powers when $\rho = 0.5$ except the CLX, the LJW, and the XY tests. The CLX test and the XY test only pick up the most significant difference between means and loses information in the dense alternative setting in the simulations here. The LJW test also losses information due to the random projection. Hence, it is expected that the CLX, the XY, and the LJW tests are not as powerful as other tests. From Table 2, we can also notice that the powers of tests of BS, CQ, SD, CLZ, XLWP, and HXWP shrink significantly (although not at the same level) as the correlation goes from low to high, for example, from $\rho = 0.5$ to 0.8; while the proposed U-projection test has consistently good powers for both low level and high level of correlation. Also notice that when $\rho = 0.5$, the power of U-projection test is comparable with the power of the BS test and the CQ test, which is expected from results in Theorem 3. In summary, the proposed U-projection test can retain Type I error rates well, is quite robust for different levels of correlation, has comparable powers in low correlation cases as other tests like BS and CQ, and can be much more powerful in high correlation cases.

Sparse signal: For the sparse type of local alternatives, we generate μ_k, $k = 1, 2$, from a two-step procedure as follows.

1. We first generate the index A_k of the nonzero elements of μ_k as a simple random sample of size $\lfloor p^s \rfloor$ without replacement from $\{1, 2, \ldots, p\}$, where $s \in (0, 1]$ controls the sparseness of the signal. In our simulations, we set $s = 0.4, 0.7$ for very sparse and moderately sparse signals, respectively.
2. After generation of A_k, we generate μ_k by $\mu_{k,-A_k} = 0$ and

$$\mu_{k,A_k} \sim N_p(0_{\lfloor p^s \rfloor}, (\theta^2/n)I_{\lfloor p^s \rfloor}), \tag{34}$$

where μ_{k,A_k} and $\mu_{k,-A_k}$ are the vectors formed by elements in μ_k with and without index A_k, respectively. Also, we set $\theta = 1$ to examine empirical power.

Empirical powers for compound-symmetry covariance and autoregressive covariance with sparse alternatives based on 1000 replications are depicted in Table 3. To save space, we present results with $\rho = 0.8$ in the table. Results for other scenarios are reported in Table S.3 in the supplementary materials of this article. From the comparison of Table 3 and Tables 1 and 2, we can notice that the powers of CLX, CLZ, XLWP, XY, HXWP improve relatively compared to BS and CQ since the former methods are derived with special considerations for sparse alternatives. However, from Table 3, it can be seen that the U-projection test is still more powerful than the tests including BS, CQ, SD, CLX, CLZ, XLWP, LJW, XY, and HXWP under high correlation covariance structures for the sparse local alternatives.

3.2. Multi-Sample Mean Testing

We now examine the proposed U-projection test for the multi-sample mean problem. We generate independent simulated data Y_{kj} from $N_p(\mu_k, \Sigma)$, for $k = 1, \ldots, K$, and $j = 1, \ldots, n_k$. The multi-sample mean problem is to test $H_0 : \mu_1 = \cdots = \mu_K$. Simulation results under non-Gaussian settings are summarized in Tables S.14 and S.15 in the supplementary materials to save space. From Tables S.14 and S.15, we can see that the simulation results under non-Gaussian settings and Gaussian settings are similar. In this simulation, we take $K = 4$ and set $n = n_1 = n_2 = n_3 = n_4$. We set $n = 50, 100$ and the dimension $p = 125, 250, 500, 1000$. As in Section 3.1, we consider both autoregressive and compound symmetric covariance structure for Σ, and we generate μ_k from a multivariate normal distribution in each simulation:

$$\mu_k \sim N_p(0_p, (\theta^2/n)I_p) \tag{35}$$

for $k = 1, 2, 3, 4$. We take $\theta = 0, 0.5$ and 1.0 to examine Type I error and empirical power. The simulation results based on 1000 simulations are summarized in Table 4 and Table S.4 in the supplementary materials. The top two panels in Table 4 depict the empirical Type I error rates for the proposed method at level 0.05 and implies that the proposed U-projection test retains Type I error rates reasonably well across different sample sizes, dimensions, and structures of covariance matrices. The empirical powers are reported in the bottom two panels in Table 4.

Comparing the panels of the autoregressive and compound-symmetry covariance matrices, we can see that the proposed test is powerful for different structures of covariance matrices. It can also be seen that the U-projection test is more powerful as the correlation ρ increases, which illustrates the utilization of the correlation information of the U-projection test. In summary, the proposed test can retain the Type I error rates well, and it is powerful under the alternative hypothesis, especially in the high correlation cases.

3.3. Testing of Significance of Predictors

In this section, we examine the performance of the proposed U-projection test for testing the significance of predictors in the linear model with high-dimensional responses:

$$Y = B_0 + \sum_{k=1}^{K} X_k \beta_k^T + E, \tag{36}$$

where Y is the $n \times p$ response matrix, B_0 is the $n \times p$ intercept matrix, $X_k, k = 1, \ldots, K$ are K predictors, each of which is a column vector of length n, $\beta_k, k = 1, \ldots, K$, are K coefficients, each of which is a column vector of length p, and E is the random error matrix and follows a multivariate normal distribution $\text{vec}(E) \sim N_{n\times p}(0, I_n \otimes \Sigma)$. Suppose we want to test the significance of the predictor X_{i_0}, then we have H_0: $\beta_{i_0} = 0_p$, and we choose $i_0 = 1$ in the simulations, that is, we want to test the significance of the predictor X_1. Simulation results under non-Gaussian settings are summarized in Tables S.17 and S.18 in the supplementary materials to save space. The simulation results under the non-Gaussian and Gaussian settings are similar. In the simulations, we choose the sample size $n = 50, 100$, the dimension $p = 125, 250, 500, 1000$, and the number of predictors $K = 5$. And we set Σ as autoregressive structure $\Sigma = \{\rho^{|i-j|}\}_{i,j}$ with $\rho = 0.2, 0.5, 0.8$ or compound-symmetry structure $\Sigma = (1-\rho)I_p + \rho 1_p 1_p^T$ with $\rho = 0.2, 0.5, 0.8$. In each simulation, we generate X from a multivariate normal distribution $N_{n\times d}(0_{nd}, I_n \otimes \Sigma_X)$, where Σ_X is assumed to follow similar structures as Σ, that is, autoregressive structures or compound-symmetry structures with $\rho = 0.2, 0.5, 0.8$. We set $B_0 = \mathbf{0}$ and generate $\beta_k, k = 2, 3, 4, 5$, iid from normal distribution $N(0_p, \frac{1}{n} I_p)$, and we generate β_1 from distribution $N(0_p, \frac{\theta^2}{n} I_p)$. Type I error rates are examined when $\theta = 0$ and empirical power rates are checked when $\theta \neq 0$. In the simulations, we set $\theta = 0, 0.5, 1.0$. The simulation results are summarized in Table 5 and Table S.5 in the supplementary materials. The top two panels with $\theta = 0$ in Table 5 summarize Type I error rates for the U-projection test, and imply that the proposed U-projection test retains Type I error rates reasonably well across all different scenarios. The empirical powers are listed in the bottom two panels in Table 5. We can notice that the power of the test increases when the sample size n increases from 50 to 100. From the comparison of the panels of autoregressive Σ and compound symmetry Σ, the proposed U-projection test is powerful for different structures of covariance matrices. Also note that since Σ_X has the same structure as Σ in the simulations, the larger ρ is, the more correlated the predictors are, and the more difficult the testing problem is, which is reflected in the fact that the empirical power of the U-projection test decreases with increase in the correlation within the same n, p, and structure of covariance matrices. In summary, in the hypothesis testing problem in the multivariate linear regression model, the proposed U-projection test can retain the Type I error rates well and is powerful under the alternative hypothesis for different sample sizes, dimensions, and structure of covariance matrices.

3.4. Real Data Analysis

We now apply the tests on a gene expression microarray dataset of murine heart under the conditions of cigarette smoke and obesity. The dataset was analyzed by Tilton et al. (2013) and is publicly available from Gene Expression Omnibus (GEO) under the serial number GSE47022. The dataset consists of 22,690 gene expression levels of 45 mouse subjects, including 23 regular weight (RW) mice and 22 diet-induced obese (DIO) mice. The mice in each group were divided into three subgroups with size 6–8 and were exposed to either filtered air (sham controls, SC), mainstream (MS) cigarette smoke or sidestream (SS) cigarette smoke. The MS group mimicked the smoker while the SS group mimicked environmental exposure through second-hand smoke.

Using the dataset, we want to investigate three problems: the first one is a two-sample testing problem between RW and DIO mice, the second one is a three-sample testing problem between mice that received filtered air, MS and SS cigarette smoke, and the third one is to analyze linear regression models with main effects of weight and cigarette smoke and also the interaction between weight and cigarette smoke. While the BS, CQ, SD, CLX, CLZ, XLWP, LJW, XY, HXWP, and U-projection tests are all applicable on the two-sample testing problem, the U-projection test can also deal with the three-sample testing problem and the testing problem in linear regression models.

Instead of functioning individually, genes work in groups to perform various biological functions. Gene Ontology system assigns genes into different gene-sets (also called GO terms) depending on their functional characteristics, see Ashburner et al. (2000) for more details. There are 9440 GO terms in total for genes from the original dataset, and we assign the 22,690 genes into the GO terms. We further remove the GO terms with less than 100 genes to accommodate high dimensionality setting, and there are still 246 GO terms left. On each GO set, we will carry on the tests as aforementioned. Since there are 246 GO sets, we will use the false discovery rate (FDR) controlling procedure (Benjamini and Hochberg 1995) to find out the significant GO terms. Note that since the correlation between GO sets are unknown and some GO sets have overlapping genes, we will use the general FDR procedure proposed in Benjamini and Yekutieli (2001) to account for unknown dependencies between GO sets.

In the U-projection test, we use $k = \lfloor 0.9n \rfloor$, where $\lfloor x \rfloor$ is the largest integer smaller than or equal to x. For choosing λ_0, we use gene expressions which are not included in the 246 GO sets to simulate the situations for the 246 GO sets. There are 4,996 genes not contained in the 246 GO sets. We divide them randomly into 50 groups, and each group has approximately 100 genes. We run U-projection tests with λ_0 from 0.0005 to 0.5 on the 50 groups and we choose λ_0 with the lowest average p-value, which is approximately 0.002.

Table 6. Number of significant GO terms in RW versus DIO mice using different tests with different false discovery rates q.

q	BS	CQ	SD	CLX	CLZ	XLWP	LJW	XY	HXWP	New
0.001	0	0	220	224	235	232	61	220	223	245
0.002	142	144	231	236	240	233	78	236	223	246
0.005	163	166	239	241	244	239	96	237	227	246
0.01	177	181	242	242	245	239	117	239	227	246
0.02	191	195	243	243	245	242	149	239	230	246
0.05	209	210	245	243	245	246	183	240	238	246

Table 7. The number of significant GO terms in linear models of DIO, MS, SS, DIO × MS, DIO × SS with different false discovery rates q.

q	DIO	MS	SS	DIO × MS	DIO × SS
0.001	229	60	240	16	235
0.002	229	60	240	16	235
0.005	229	60	240	16	235
0.01	241	60	243	16	242
0.02	244	94	244	16	245
0.05	246	136	246	16	246

In the two-sample mean testing problem, we apply the BS, CQ, SD, CLX, CLZ, XLWP, LJW, XY, HXWP, and U-projection tests to check the equality of gene expression means in RW and DIO mice. The numbers of significant GO sets of each test using FDR procedure with different false discovery rates are listed in Table 6, from which it can be seen that the factor of obesity is quite significant in most of the GO terms in heart cells. The BS and CQ tests perform similarly. And the BS, CQ, and LJW tests are not as powerful as other tests. The performances of SD, CLX, CLZ, XLWP, XY, and HXWP tests are similar, and they all declare most of the GO terms to be significant. Across different false discovery rates, it can be seen that the proposed U-projection test is one of the most powerful tests in this example: the U-projection test only drops one GO term at a very small false discovery rate $q = 0.001$. The numerical result supports our findings that the proposed U-projection test retains high power by using the information provided by covariance.

In the multi-sample mean testing problem, we test the equality of gene expression means of SC, MS and SS mice. If the U-projection test is carried out on all the gene expressions, then a p-value less than 0.0001 is obtained, and we can see there is a significant difference between gene expressions of the three populations. However, if the U-projection test is carried on each GO set and an FDR procedure is used, then there is no significant GO set found even when FDR rate is set to 0.05. It suggests that even the three groups are significantly different overall, but the difference is not significant when coming to individual GO sets, which also follows the analysis in Tilton et al. (2013) that the factor of cigarette smoke is not as important as the weight factor in the heat cells. Also, it can be seen in the next problem of testing on linear regression coefficients that the cigarette factor is significant on many GO terms when the weight factor is also considered.

In the testing for the linear regression model, we build a model of the obesity factor DIO, the cigarette smoke factors MS and SS, the interaction of DIO × MS, and the interaction of DIO × SS. The numbers of significant GO sets of each factor using FDR procedure with different false discovery rates are listed in Table 7. From the table, it can be seen that the factor DIO, SS and the interaction between DIO and SS are significant in most of the GO terms (246 in total) in heart cells; and the factor MS and the interaction between DIO and MS are also significant in some of the GO terms, but not as many as DIO, SS, and DIO × SS, which is also in agreement with the findings from Tilton et al. (2013).

In summary, it can be seen that the proposed U-projection test retains high power by using the covariance information and can deal with important high-dimensional testing problems including one- and two-sample mean testing.

4. Discussions

In this article, we proposed a new testing procedure for the linear model with a high-dimensional response vector. The proposed testing procedure can be directly applied for one-sample, two-sample, or multi-sample mean problems. For one-sample and two-sample mean testing problems, our theoretical analysis and numerical results show that under the low correlation condition, the power of the proposed test is similar to some of the existing tests, while the proposed test can have substantial power improvements in the presence of high correlation among the variables.

In this article, we focused on the asymptotic regime where p and n go to infinity while d and m are fixed. An interesting possible future extension is the more general asymptotic regime with d and m also diverging. Theorem 1 established an upper bound for the optimal dimension to be m, and we advocated to use the optimal projection matrix of dimension m when m is small. However, if m also goes to infinity, then it may be beneficial to further reduce the dimension of the projection matrix, and it will be quite important to understand the asymptotic behavior of the proposed U-projection test with diverging m. A related insightful study is He et al. (2019), which considered the LRT for the general linear hypothesis under the general asymptotic regime with $n, p, d, m \to \infty$. More specifically, He et al. (2019) studied the asymptotic boundary where the classical LRT fails and developed the limiting distribution of the LRT for a general asymptotic regime, which can provide a great insight for the asymptotic behavior of the U-projection test with diverging m.

Supplementary Materials

The online appendix consists of proofs of theorems in the paper. The supplementary materials consist of additional simulation results and technical details for proofs in the online appendix.

Acknowledgments

The authors are indebted to the referees, the associate editor, and the co-editor for their valuable comments, which have significantly improved the article. The content is solely the responsibility of the authors and does not necessarily represent the official views of NSF and NIH.

Funding

This research was supported by NSF grants DMS 1820702, DMS 1953196, DMS 2015539 and NIH grants NIAID 5-R01-AI136664, R01CA229542, and R01 ES019672.

References

Aoshima, M., and Yata, K. (2018), "Two-Sample Tests for High-Dimension, Strongly Spiked Eigenvalue Models," *Statistica Sinica*, 28, 43–62. [1]

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., and Harris, M. A. (2000), "Gene Ontology: Tool for the Unification of Biology," *Nature Genetics*, 25, 25–29. [11]

Bai, Z., and Saranadasa, H. (1996), "Effect of High Dimension: By an Example of a Two Sample Problem," *Statistica Sinica*, 6, 311–329. [1,7,9]

Baik, J., and Silverstein, J. W. (2006), "Eigenvalues of Large Sample Covariance Matrices of Spiked Population Models," *Journal of Multivariate Analysis*, 97, 1382–1408. [6]

Benjamini, Y., and Hochberg, Y. (1995), "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society*, Series B, 57, 289–300. [11]

Benjamini, Y., and Yekutieli, D. (2001), "The Control of the False Discovery Rate in Multiple Testing Under Dependency," *The Annals of Statistics*, 29, 1165–1188. [11]

Cai, T., Liu, W., and Xia, Y. (2014), "Two-Sample Test of High Dimensional Means Under Dependence," *Journal of the Royal Statistical Society*, Series B, 76, 349–372. [1,9]

Chen, S. X., Li, J., and Zhong, P.-S. (2019), "Two-Sample and ANOVA Tests for High Dimensional Means," *The Annals of Statistics*, 47, 1443–1474. [1,9]

Chen, S. X., and Qin, Y.-L. (2010), "A Two-Sample Test for High-Dimensional Data With Applications to Gene-Set Testing," *The Annals of Statistics*, 38, 808–835. [1,2,6,7,9]

Donoho, D. L., Gavish, M., and Johnstone, I. M. (2018), "Optimal Shrinkage of Eigenvalues in the Spiked Covariance Model," *The Annals of Statistics*, 46, 1742. [6]

Fan, J., Liao, Y., and Yao, J. (2015), "Power Enhancement in High-Dimensional Cross-Sectional Tests," *Econometrica*, 83, 1497–1541. [1]

He, Y., Jiang, T., Wen, J., and Xu, G. (2019), "Likelihood Ratio Test in Multivariate Linear Regression: From Low to High Dimension," *Statistica Sinica* (accepted). [12]

He, Y., Xu, G., Wu, C., and Pan, W. (2021), "Asymptotically Independent U-Statistics in High-Dimensional Testing," *The Annals of Statistics*, 49, 154–181. [1,9]

Hotelling, H. (1931), "The Generalization of Student's Ratio," *The Annals of Mathematical Statistics*, 2, 360–378. [1]

Huang, Y. (2015), "Projection Test for High-Dimensional Mean Vectors With Optimal Direction," Ph.D. dissertation, Department of Statistics, The Pennsylvania State University at University Park. [4]

Johnstone, I. M. (2001), "On the Distribution of the Largest Eigenvalue in Principal Components Analysis," *The Annals of Statistics*, 29, 295–327. [6]

Kock, A. B., and Preinerstorfer, D. (2019), "Power in High-Dimensional Testing Problems," *Econometrica*, 87, 1055–1069. [1]

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*, pp. 1206–1214. [1,9,10]

Srivastava, M. S., and Du, M. (2008), "A Test for the Mean Vector With Fewer Observations Than the Dimension," *Journal of Multivariate Analysis*, 99, 386–402. [1,9]

Srivastava, R., Li, P., and Ruppert, D. (2016), "RAPTT: An Exact Two-Sample Test in High Dimensions Using Random Projections," *Journal of Computational and Graphical Statistics*, 25, 954–970. [1]

Thulin, M. (2014), "A High-Dimensional Two-Sample Test for the Mean Using Random Subspaces," *Computational Statistics & Data Analysis*, 74, 26–38. [1]

Tilton, S. C., Karin, N. J., Webb-Robertson, B.-J. M., Waters, K. M., Mikheev, V., Lee, K. M., Corley, R. A., Pounds, J. G., and Bigelow, D. J. (2013), "Impaired Transcriptional Response of the Murine Heart to Cigarette Smoke in the Setting of High Fat Diet and Obesity," *Chemical Research in Toxicology*, 26, 1034–1042. [11,12]

Wang, R., and Xu, X. (2018), "On Two-Sample Mean Tests Under Spiked Covariances," *Journal of Multivariate Analysis*, 167, 225–249. [1]

Wang, W., and Fan, J. (2017), "Asymptotics of Empirical Eigenstructure for High Dimensional Spiked Covariance," *The Annals of Statistics*, 45, 1342. [6]

Xu, G., Lin, L., Wei, P., and Pan, W. (2016), "An Adaptive Two-Sample Test for High-Dimensional Means," *Biometrika*, 103, 609–624. [1,9]

Xue, K., and Yao, F. (2020), "Distribution and Correlation-Free Two-Sample Test of High-Dimensional Means," *The Annals of Statistics*, 48, 1304–1328. [1,9]

Zhang, J.-T., Guo, J., Zhou, B., and Cheng, M.-Y. (2020), "A Simple Two-Sample Test in High Dimensions Based on l^2-Norm," *Journal of the American Statistical Association*, 115, 1011–1027. [1]