Journal of Multivariate Analysis 109 (2012) 61–72

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On model-free conditional coordinate tests for regressions

Zhou Yu *, Lixing Zhu, Xuerong Meggie Wen

School of Finance and Statistics, East China Normal University, China
Department of Mathematics, Hong Kong Baptist University, Hong Kong
Department of Mathematics and Statistics, Missouri University of Science and Technology, United States

ARTICLE INFO

Article history:
Received 13 April 2011
Available online 26 February 2012

AMS subject classifications:
62H15
62H25
62H86

Keywords:
Conditional coordinate test
Sufficient dimension reduction
Sliced inverse regression

ABSTRACT

Existing model-free tests of the conditional coordinate hypothesis in sufficient dimension reduction (Cook (1998) [3]) focused mainly on the first-order estimation methods such as the sliced inverse regression estimation (Li (1991) [14]). Such testing procedures based on quadratic inference functions are difficult to be extended to second-order sufficient dimension reduction methods such as the sliced average variance estimation (Cook and Weisberg (1991) [9]). In this article, we develop two new model-free tests of the conditional predictor hypothesis. Moreover, our proposed test statistics can be adapted to commonly used sufficient dimension reduction methods of eigendecomposition type. We derive the asymptotic null distributions of the two test statistics and conduct simulation studies to examine the performances of the tests.

1. Introduction

For parametric regressions, hypothesis testing for predictor contributions to the response is a well developed research area. For instance, in the linear models, t test is often applied to check the contribution of every predictor. However, for semiparametric models, this topic has not yet received enough attention because how to construct a test therein is a challenge. To attack this problem, Cook [4] investigated this issue in a dimension reduction framework.

For a typical regression problem with a univariate random response Y and a vector of random predictors $\mathbf{X} = (X_1, \ldots, X_p)^T \in \mathbb{R}^p$, the goal is to understand how the conditional distribution $Y|\mathbf{X}$ depends on the value of $\mathbf{X}$. The spirit of sufficient dimension reduction [14,3] is to reduce the dimension of $\mathbf{X}$ without loss of information on the regression and without requiring a pre-specified parametric model. Assuming the following semiparametric regression model: $Y = g(\boldsymbol{\beta}_1^T\mathbf{X}, \boldsymbol{\beta}_2^T\mathbf{X}, \ldots, \boldsymbol{\beta}_d^T\mathbf{X}, \epsilon)$, where $g(\cdot)$ is an unknown function and ϵ is an unknown random error independent of $\mathbf{X}$, we can see that the conditional distribution of $Y|(\boldsymbol{\beta}_1^T\mathbf{X}, \ldots, \boldsymbol{\beta}_d^T\mathbf{X})$ is the same as that of $Y|\mathbf{X}$ for all values of $\mathbf{X}$. Hence, these $\boldsymbol{\beta}$'s provide a parsimonious characterization of the conditional distribution of $Y|\mathbf{X}$. We call them the *effective (sufficient) directions* [14,3]. When d is small which is often the case in real applications, the original regression problem (data) can be effectively reduced by projecting $\mathbf{X}$ along these effective directions.

More formally, we search for subspaces $\mathcal{S} \subseteq \mathbb{R}^p$ such that $Y \perp\!\!\!\perp \mathbf{X}|P_{\mathcal{S}}\mathbf{X}$ where $\perp\!\!\!\perp$ indicates independence, and $P_{(\cdot)}$ stands for a projection operator with respect to the standard inner product. The intersection of all such $\mathcal{S}$ is defined as the *central subspace*, denoted as $\mathcal{S}_{Y|X}$ [3], which almost always exists in practice under mild conditions [25]. We assume the existence of the central subspaces throughout this article. Sufficient dimension reduction is concerned with making inferences for the central subspace. $d = \dim(\mathcal{S}_{Y|X})$ is called the structural dimension of the regression. Unlike other nonparametric approaches, sufficient dimension reduction can often avoid the curse of dimensionality. Many sufficient dimension reduction methods enjoy $\sqrt{n}$ convergence rates since they exploit the global features of the dependence of Y on $\mathbf{X}$.

* Corresponding author at: East China Normal University, School of Finance and Statistics, Shanghai, China.
E-mail address: yz19830224@gmail.com (Z. Yu).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2012.02.004

Sufficient dimension reduction has been a promising field during the past decades. It has attracted considerable interests, and many methods have been developed. Among them, sliced inverse regression (SIR; [14]), sliced average variance estimation (SAVE; [9]), minimum average variance estimation [22], inverse regression estimation [7] and directional regression (DR; [17]) are the most widely investigated methods in the literature. All these aforementioned methods except Xia et al. [22] mainly focus on the estimation of the central subspace.

Other than estimating the central subspace, it is also of interest to evaluate the predictor effects in a model free setting. Cook [4] considered two types of hypotheses to test the significance of subsets of predictors under the framework of sufficient dimension reduction. The first type is the *Marginal Coordinate Hypothesis*: $\mathcal{P}_{\mathcal{H}}\mathcal{S}_{Y|X} = \mathcal{O}_p$ versus $\mathcal{P}_{\mathcal{H}}\mathcal{S}_{Y|X} \neq \mathcal{O}_p$. The second type is called the *Conditional Coordinate Hypothesis*:

$$\mathcal{P}_{\mathcal{H}}\mathcal{S}_{Y|X} = \mathcal{O}_p \quad \text{versus} \quad \mathcal{P}_{\mathcal{H}}\mathcal{S}_{Y|X} \neq \mathcal{O}_p \quad \text{given } d; \tag{1.1}$$

where $\mathcal{H}$ is an r-dimensional ($r \leq p-d$) user-selected subspace of the predictor space and $\mathcal{O}_p$ indicates the origin in $\mathbb{R}^p$. For example, suppose that $\mathbf{X}^T = (\mathbf{X}_1^T, \mathbf{X}_2^T)$, where $\mathbf{X}_1 \in \mathbb{R}^r$ and $\mathbf{X}_2 \in \mathbb{R}^{p-r}$, and we would like to test if $\mathbf{X}_1$ makes any contributions to the regression $Y|\mathbf{X}$, we then consider these two types of hypotheses tests with $\mathcal{H} = \text{Span}((I_r, 0)^T)$. Although in general, $\mathcal{H}$ need not correspond to a subset of predictors (coordinates).

Hence, both the marginal coordinate hypothesis and the conditional coordinate hypothesis can be used to test the contributions of selected predictors without requiring a pre-specified model about the original regression $Y|\mathbf{X}$. When d, the structural dimension of the regression, is specified as a modeling device, or inferences on d result in a clear estimate, a conditional coordinate hypothesis test will be the natural choice. Otherwise, a marginal coordinate hypothesis would be tested. We would expect that the conditional coordinate hypothesis will provide us with greater power when a correct d is given prior to testing predictors. On the other hand, when d is misspecified, a conditional coordinate hypothesis test might lead to misleading results, while the marginal coordinate hypothesis test should be considered. Although simulation studies provided in Section 4 suggest that the misspecification of d need not be a worrisome issue in practice.

Based on a nonlinear least squares formulation of the sliced inverse regression estimation, Cook [4] constructed asymptotic tests for the marginal and conditional coordinate hypotheses. Cook and Ni [7] showed how to test marginal (conditional) coordinate hypotheses using various quadratic inference functions, which stimulated the tests of conditional independence hypotheses based on the minimum discrepancy approach [7] and the covariance inverse regression estimation [8].

All the aforementioned tests are based on the first moment of the inverse regression of $\mathbf{X}|Y$ that are called the first-order sufficient dimension reduction methods. Note that these tests for the predictor contributions might be invalid when the response surface is symmetric about the origin since these first-order sufficient dimension reduction methods themselves would fail in such cases. Therefore, it is of great interest to consider coordinate tests using the second-order sufficient dimension reduction methods which involve both the first and second moments of the inverse regression of $\mathbf{X}|Y$. However, the commonly used second-order sufficient dimension reduction methods such as the sliced average variance estimation [9], and the directional regression [17], are very different from those first-order methods which could be derived from quadratic inference functions. Hence, the asymptotic tests developed by Cook and Ni [7] are not directly applicable. Shao et al. [21] provided a marginal coordinate test based on the sliced average variance estimation. But to the best of our knowledge, there are no methods available in the literature for testing of the conditional coordinate hypotheses of (1.1) with second-order dimension reduction methods. To address this issue, we in this article present two new tests of conditional coordinate hypotheses which could be adapted to essentially all existing sufficient dimension reduction methods of the eigendecomposition type, including both the sliced inverse regression estimation and the sliced average variance estimation methods.

The rest of the paper is organized as follows. Section 2 revisits several moment based sufficient dimension reduction methods. In Section 3, we construct two new tests and present their asymptotic null distributions. Sections 4 and 5 are concerned with simulation studies and a real data application. We conclude with a brief discussion in Section 6. For easy of exposition, the proofs of the asymptotic results are deferred to the Appendix A.

2. Sufficient dimension reduction methods revisited

Let $\boldsymbol{\mu} = E(\mathbf{X})$, $\boldsymbol{\Sigma} = \text{Var}(\mathbf{X})$, and $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ be the standardized predictor. Many moment based sufficient dimension reduction methods can be formulated as the following eigendecomposition problem:

$$\mathcal{M}\boldsymbol{\eta}_i = \lambda_i\boldsymbol{\eta}_i, \quad i = 1, \ldots, p, \tag{2.2}$$

where $\mathcal{M}$ is the $\mathbf{Z}$ scale method-specific candidate matrix. Under certain conditions imposed only on the marginal distribution of the predictor, the eigenvectors $(\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_d)$ corresponding to the nonzero eigenvalues $\boldsymbol{\lambda}_1 \geq \cdots \geq \boldsymbol{\lambda}_d$ form a basis of the $\mathbf{Z}$ scale central subspace $\mathcal{S}_{Y|Z}$. Then by the invariance property $\mathcal{S}_{Y|X} = \boldsymbol{\Sigma}^{-1/2}\mathcal{S}_{Y|Z}$ as described by Cook [3], $\boldsymbol{\beta} = (\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\eta}_1, \ldots, \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\eta}_d)$ forms a basis of $\mathcal{S}_{Y|X}$.

As most of commonly used sufficient dimension reduction methods that target $\mathcal{S}_{Y|Z}$ are of candidate matrices satisfying the above eigendecomposition, we only list some as follows:

Sliced Inverse Regression: $\mathcal{M} = \text{Var}\{E(\mathbf{Z}|Y)\}$;

Sliced Average Variance Estimation: $\mathcal{M} = E\{I_p - \text{Var}(\mathbf{Z}|Y)\}^2$;

Directional Regression: $\mathcal{M} = 2E\{E^2(\mathbf{Z}\mathbf{Z}^T|Y)\} + 2E^2\{E(\mathbf{Z}|Y)E(\mathbf{Z}^T|Y)\} + 2E\{E(\mathbf{Z}^T|Y)E(\mathbf{Z}|Y)\}E\{E(\mathbf{Z}|Y)E(\mathbf{Z}^T|Y)\} - 2I_p$.

As we discussed in Section 1, SIR is a first-order estimation method, while SAVE and DR are second-order methods. Li and Wang [17] showed that both SAVE and DR are exhaustive under certain conditions posed on the marginal distribution of $\mathbf{X}$; that is, the column spaces of the candidate matrices based on SAVE and DR are equal to $\mathcal{S}_{Y|Z}$. However, it is known that SIR is not exhaustive when the relation between Y and $\mathbf{X}$ contains a U-shaped trend. Thus in addition to SIR, it is necessary to develop conditional coordinate tests based on second-order sufficient dimension reduction methods. In the following, we adopt SIR, SAVE and DR to illustrate the idea of our proposed tests and evaluate their performances in simulation studies.

3. Two general tests of conditional coordinate hypotheses

3.1. The first test statistic

Let $\mathcal{W}$ be a user-selected $p \times r$ matrix basis of $\mathcal{H}$. The null hypotheses of (1.1) implies that $\mathcal{W}^T\boldsymbol{\beta}_i = \mathcal{W}^T\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\eta}_i = 0$ for $i = 1, \ldots, d$. It is then natural to study the asymptotic null distribution of $\mathcal{W}^T\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{\eta}}_i$ and further construct test statistic, where $\widehat{\boldsymbol{\Sigma}}$ and $\widehat{\boldsymbol{\eta}}_i$ are the sample estimators of $\boldsymbol{\Sigma}$ and $\boldsymbol{\eta}_i$ respectively. Zhu and Fang [26], Chen and Li [2] and Fung et al. [11] derived the asymptotic distributions of the ith $(1 \le i \le d)$ estimated direction $\widehat{\boldsymbol{\eta}}_i$ by different inverse regression methods. However, their theoretical results require that the nonzero eigenvalues of $\mathcal{M}$ are distinct, which is a rather stringent condition. When this condition is violated, their asymptotic results for the $\widehat{\boldsymbol{\eta}}_i$'s will not be valid.

To address this issue, we consider $P = \sum_{i=1}^{d}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T$ rather than $\boldsymbol{\eta}_i$. Let $Q = \sum_{i=d+1}^{p}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T$. Note that Q is the unique eigenprojection corresponding to zero eigenvalue with multiplicity $p-d$. Therefore $P = I_p - Q$ is always identifiable, while $\boldsymbol{\eta}_i$ is not identifiable if the corresponding eigenvalue, λ_i, is a multiple eigenvalue. We then take $H = \boldsymbol{\Sigma}^{-1/2}\mathcal{W}(\mathcal{W}^T\boldsymbol{\Sigma}^{-1}\mathcal{W})^{-1/2}$ to be an orthonormal basis of $\boldsymbol{\Sigma}^{-1/2}\mathcal{H}$ and let $L_1 = \mathrm{trace}(H^TPH)$. The next proposition connects the conditional coordinates hypothesis with L_1.

Proposition 1. *Assume that* $\mathrm{Span}(\mathcal{M}) = \mathcal{S}_{Y|Z}$ *and* $d = \dim(\mathcal{S}_{Y|X})$ *is known, then* $\mathcal{P}_{\mathcal{H}}\mathcal{S}_{Y|X} = \mathcal{O}_p$ *if and only if* $L_1 = 0$.

Proposition 1 inspires us to construct the following test statistic:

$$T_1 = n\widehat{L}_1 = n\mathrm{trace}(\widehat{H}^T\widehat{P}\widehat{H}),$$

where $\widehat{P}$ is the estimator of P and $\widehat{H} = \widehat{\boldsymbol{\Sigma}}^{-1/2}\mathcal{W}(\mathcal{W}^T\widehat{\boldsymbol{\Sigma}}^{-1}\mathcal{W})^{-1/2}$. As we know, the sliced inverse regression may miss directions in $\mathcal{S}_{Y|Z}$ when the regression function $E(Y|\mathbf{X})$ is curved with little linear trend, hence the condition $\mathrm{Span}(\mathcal{M}) = \mathcal{S}_{Y|Z}$ may not hold for estimation methods by the sliced inverse regression. Therefore, in addition to the methods developed based on sliced inverse regression, we also consider the asymptotic null distribution of T_1 with sliced average variance estimation and directional regression which can better or even exhaustively estimate the central subspace. Let $\widehat{\mathcal{M}}$ be the sample level candidate matrix corresponding to one of the aforementioned three dimension reduction methods. The following lemma provides us with the expansions of $\widehat{\boldsymbol{\Sigma}}^{-1/2}$ and $\widehat{\mathcal{M}}$.

Lemma 1. *Assume that the data* $(\mathbf{X}_i, Y_i)$, *for* $i = 1, \ldots, n$, *are a random sample from* $(\mathbf{X}, Y)$ *with finite fourth order moments. Then we have the following expansions:*

$$\widehat{\boldsymbol{\Sigma}}^{-1/2} = \boldsymbol{\Sigma}^{-1/2} + E_n\{\boldsymbol{\Sigma}^{*-1/2}(\mathbf{X}, Y)\} + o_p(n^{-1/2}),$$
$$\widehat{\mathcal{M}} = \mathcal{M} + E_n\{\mathcal{M}^*(\mathbf{X}, Y)\} + o_p(n^{-1/2}),$$

where $E_n\{.\}$ *indicates the sample average* $n^{-1}\sum_{i=1}^{n}\{.\}$, *and the explicit formula of* $\boldsymbol{\Sigma}^{*-1/2}(\mathbf{X}, Y)$ *and* $\mathcal{M}^*(\mathbf{X}, Y)$ *(for SIR, SAVE and DR) are given in Appendix B.*

The next theorem gives the asymptotic distribution of T_1 under null hypothesis (1.1).

Theorem 1. *Assume the conditions of Proposition 1 and Lemma 1 hold. Then under null hypotheses (1.1)*

$$T_1 \longrightarrow \sum_{i=1}^{dr}\omega_i\chi_i^2(1),$$

where $d = \dim(\mathcal{S}_{Y|X})$, $r = \dim(\mathcal{H})$, $\omega_1 \ge \cdots \ge \omega_{dr}$ *are the eigenvalues of* $\Omega = E\{\mathrm{vec}(A)\mathrm{vec}(A)^T\}$ *with*

$$A = (\mathcal{W}^T\boldsymbol{\Sigma}^{-1}W)^{-1/2}\mathcal{W}^T\left\{\boldsymbol{\Sigma}^{*-1/2}(\mathbf{X}, Y)P + \boldsymbol{\Sigma}^{-1/2}\mathcal{M}^*(\mathbf{X}, Y)\sum_{i=1}^{d}\lambda_i^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\right\},$$

and $\mathrm{vec}(\cdot)$ *is the operator which stacks the columns of a matrix to form a vector.*

By substituting sample estimates for the unknown quantities, we can get a consistent estimates of Ω, denoted by $\widehat{\Omega}$. Moreover, the weights ω_i's can be consistently estimated as the sample eigenvalues $\hat{\omega}_i$'s of $\widehat{\Omega}$. Then a p-value for the conditional coordinate hypothesis can be constructed by comparing the observed value of the test statistic to the percentage points of $\sum_{i=1}^{dr}\hat{\omega}_i\chi_i^2(1)$, which can be approximately obtained by Monte Carlo simulations. We can also approximate the tail probabilities by using the modified test statistics proposed by Bentler and Xie [1].

3.2. The second test statistic

Naik and Tsai [19] suggested a constrained sufficient dimension regression approach for incorporating the prior information. If we regard the null hypothesis as the prior information, we can follow Naik and Tsai [19] to impose corresponding constraints when applying any sufficient dimension reduction methods. Then we can compare the sufficient dimension reduction estimators under the null hypothesis (1.1) and under the full model to construct our test statistic.

Let $P_{\mathcal{W}} = HH^T$, $Q_{\mathcal{W}} = I_p - P_{\mathcal{W}}$ and $\mathcal{M}_c = Q_{\mathcal{W}}\mathcal{M}Q_{\mathcal{W}}$. Then $\mathcal{M}_c$ can be regarded as the candidate matrix under the null hypothesis (1.1). Denote the ith eigenvalue and its corresponding eigenvector of $\mathcal{M}_c$ by ρ_i and $\boldsymbol{\xi}_i$ respectively, $i = 1, \ldots, p$. Let $P_c = \sum_{i=1}^{d} \boldsymbol{\xi}_i\boldsymbol{\xi}_i^T$ and $L_2 = \|P - P_c\|^2$. Similar to Proposition 1, the next proposition relates the null hypothesis (1.1) to L_2.

Proposition 2. *Assume the same conditions of Proposition 1 hold, then $\mathcal{P}_{\mathcal{H}}\mathcal{S}_{Y|X} = \mathcal{O}_p$ if and only if $L_2 = 0$.*

Let $\widehat{P}_{\mathcal{W}} = \widehat{H}\widehat{H}^T$, $\widehat{Q}_{\mathcal{W}} = I_p - \widehat{P}_{\mathcal{W}}$ and $\widehat{\mathcal{M}}_c = \widehat{Q}_{\mathcal{W}}\widehat{\mathcal{M}}_c\widehat{Q}_{\mathcal{W}}$ be the sample estimators of $P_{\mathcal{W}}$, $Q_{\mathcal{W}}$ and $\mathcal{M}_c$ respectively. Let $\widehat{\rho}_i$'s and $\widehat{\boldsymbol{\xi}}_i$'s be the sample eigenvalues and sample eigenvectors of $\widehat{M}_c$. Denote the sample estimator of P_c by $\widehat{P}_c = \sum_{i=1}^{d} \widehat{\boldsymbol{\xi}}_i\widehat{\boldsymbol{\xi}}_i^T$. Then a test statistic can be constructed as the difference between $\widehat{P}$ and $\widehat{P}_c$, that is, $T_2 = n\widehat{L}_2 = n\|\widehat{P} - \widehat{P}_c\|^2$. The next theorem gives the limiting distribution of T_2 under null hypothesis.

Theorem 2. *Assume the same conditions of Theorem 1 hold, then under null hypotheses (1.1),*

$$T_2 \longrightarrow \sum_{i=1}^{dr} \delta_i \chi_i^2(1),$$

where $d = \dim(\mathcal{S}_{Y|X})$, $r = \dim(\mathcal{H})$, and $\delta_1 \geq \cdots \geq \delta_{dr}$ are the eigenvalues of $\Delta = E\{\mathrm{vec}(B + B^T)\mathrm{vec}(B + B^T)^T\}$ with

$$B = \left\{P_{\mathcal{W}}\mathcal{M}^*(\mathbf{X}, Y) + \boldsymbol{\Sigma}^{-1/2}\mathcal{W}(\mathcal{W}^T\boldsymbol{\Sigma}^{-1}\mathcal{W})^{-1}\mathcal{W}^T\boldsymbol{\Sigma}^{*-1/2}(\mathbf{X}, Y)\mathcal{M}\right\}\sum_{i=1}^{d}\lambda_i^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T.$$

Δ can be consistently estimated using the obvious sample analogues. Then the nonzero eigenvalues δ_i's of Δ can also be consistently estimated, denoted by $\hat{\delta}_i$'s.

4. Simulation studies

4.1. Study 1: tests with the sliced inverse regression estimation

In this section, we conduct a simulation study to examine the asymptotic results of proposed tests based on the sliced inverse regression. We also compare our tests with the sliced inverse regression based general marginal coordinate test and general conditional coordinate test developed by Cook [4]. We consider the following two models: (I) $Y = 1/X_1 + 0.2\epsilon$, (II) $Y = \exp(X_1)\mathrm{sgn}(X_2) + 0.2\epsilon$. The predictor vector $\mathbf{X} = (X_1, \ldots, X_p)$ follows a multivariate normal distribution with mean 0, and the correlation between X_i and X_j is $0.5^{|i-j|}$, The error ϵ is standard normal and is independent of $\mathbf{X}$. The predictor dimension p is taken as 4 and 8. Model I is one dimensional, and Model II is of two dimensional structure. We use $h = 5$ slices and summarize the results over 1000 replications for each simulation study.

Table B.1 contains the empirical levels of the tests under the null hypothesis that $\mathcal{S}_{Y|X} \subseteq \mathcal{V}$ versus the alternative hypothesis $\mathcal{S}_{Y|X} \nsubseteq \mathcal{V}$ based on Models I and II, where $\mathcal{V} = \mathrm{Span}(e_1, e_2)$ and e_i is a canonical basis vector with its ith element being 1 and all other elements being 0. It becomes apparent that the significance levels are well attained in most cases for every test if n is large. However, when p is large and n is relatively small, our tests perform better overall. For example, when $p = 8$, $n = 200$, and the nominal level is $\alpha = 0.05$, the actual levels of our tests are 0.055 and 0.052, comparing to 0.079 and 0.081 from those tests developed by Cook [4] for Model II.

Table B.2 provides the estimated power based on Models I and II for testing the hypothesis $\mathcal{S}_{Y|X} \subseteq \mathcal{V}$ versus the alternative hypothesis $\mathcal{S}_{Y|X} \nsubseteq \mathcal{V}$, where $\mathcal{V} = \mathrm{Span}(e_3, e_4)$. The power was computed at 5% nominal level. We can see from Table B.2 that the power for all the four tests approached 100%.

Therefore, our limited simulations suggest that our tests perform at least as well as the tests developed by Cook [4] based on the sliced inverse regression estimation.

4.2. Study 2: tests with second-order methods

In this section, we conduct a simulation study to check the validity of our asymptotic tests with second-order sufficient dimension reduction methods such as the sliced average variance estimation and the directional regression. We also include the general marginal coordinate test with the sliced average variance estimation developed by Shao et al. [21] for a comparison. Consider the following two models: (III) $Y = \log|X_1| + 0.2\epsilon$, (IV) $Y = 0.4X_1^2 + 3\sin(X_2/4) + 0.2\epsilon$. The predictor

X and the error ϵ are generated in the same way as in the previous simulation study. For Model III, $d = 1$; whereas for Model IV, $d = 2$. In models III and IV, at least one component is symmetric about 0, so the condition $\mathrm{Span}(\mathcal{M}) = \mathcal{S}_{Y|Z}$ does not hold if a first-order sufficient dimension reduction method such as the sliced inverse regression estimation is used.

Similar to Tables B.1 and B.2, the simulation results of estimated nominal levels and estimated power are summarized in Tables B.3 and B.4 respectively. The estimated levels of our tests seem a little far from the nominal levels when $n = 100$ and $p = 8$, but the agreement between the nominal levels and the estimated levels seems good for our tests when the sample size is relatively large. For example, for model III, with $n = 200$, $p = 4$, the estimated levels of our tests are: 0.053 and 0.049 for T_1 and T_2 respectively, which are pretty close to the nominal level $\alpha = 0.05$. It turns out that the marginal coordinate test tends to underestimate the nominal levels in most scenarios and has a lower power when the sample size is relatively small. Also, it is known that directional regression is more accurate than or competitive with any other second-order sufficient dimension methods. As expected, while enjoying similar performances in estimating the nominal levels, the conditional coordinate tests based on the directional regression have greater power than those based on the sliced average variance estimation. Moreover, because our second test statistic, T_2, involves more plug-in estimates than that of our first test statistic, T_1, it is understandable that T_1 provides greater power than T_2, when the sample size is small.

4.3. Choice of d

As discussed in [4], misspecification of d may lead to conclusions different from those based on the true value. In this section, we conduct a simulation study based on Model IV to investigate the estimated 5% nominal level and estimated power with different choices of d, in which the null hypothesis and alternative hypothesis are the same as in the previous section. Simulation results are summarized in Table B.5. For Model IV, the structural dimension is 2. Table B.5 suggests that the empirical significance levels are not very far from 5% even d is misspecified. With d underspecified as 1, the power still approaches to 100%. However, the power deteriorates as d is overspecified since X_3 and X_4 may be regarded as contributing predictors with $d = 3$ or $d = 4$. This is not a troublesome issue since the conclusions drawn from conditional coordinate test with an overspecified d give an upper bound on the set of relevant predictors. Our limited experiences from this study are consistent with Cook [4]'s findings that misspecification of d need not be a worrisome issue. In practice, we can always first estimate d through marginal dimension test and then use the estimated d in conditional coordinate test.

5. Swiss banknote data

The Swiss banknote data [10]consists of 200 observations. The response variable is a note's authenticity, $Y = 0$ for genuine notes and $Y = 1$ for counterfeit notes. There are six predictors measuring the size of a note in millimeters: Length at center (*Length*), Left-edge length (*Left*), Right-edge length (*Right*), length of Bottom edge (*Bottom*), length of Top edge (*Top*), and Diagonal length (*Diagonal*).

Based on the marginal coordinate test with the sliced average variance estimation, Shao et al. [21] concluded that *Length, Top, Left* and *Right* are irrelevant predictors and could be removed from the regression without much loss of information. In addition, Cook and Lee [5], Li [16] and Shao et al. [21] suggested that the structural dimension of this data is two. Here we apply our proposed two tests of the conditional coordinate hypotheses with $d = 2$ to analyze this data. With the additional information on the structural dimension, we expect potential gains from the conditional coordinate test. To approximate the p-values, we follow Bentler and Xie [1] to adopt the adjusted test statistics $\widetilde{T}_1 = (\sum \hat{\omega}_i/d_1)^{-1}T_1$ and $\widetilde{T}_2 = (\sum \hat{\delta}_i/d_2)^{-1}T_2$, where d_1 and d_2 are the nearest integers to $(\sum \hat{\omega}_i)^2/(\sum \hat{\omega}_i^2)$ and $(\sum \hat{\delta}_i)^2/(\sum \hat{\delta}_i^2)$ respectively. Then $\widetilde{T}_i$ is approximately distributed as a chi-squared variate with degrees of freedom of d_i, $i = 1, 2$.

Table B.6 presents the backward elimination procedure based on the two adjusted test statistics. Different from the conclusions drawn from the marginal coordinate test, Table B.6 suggests that at the 5% level that only *Left* and *Right* are uninformative predictors. To be a little less conservative, we can even conclude that *Length* is also an uninformative predictor at the 1% level.

We further conduct a series of tests for the joint effects of the predictors as presented in Table B.7. It is clear that we should not reject the hypothesis that Y is independent of *Length*, *Left* and *Right* given the remaining three predictors; the p-value of this hypothesis was 0.119 for $\widetilde{T}_1$, and 0.120 for $\widetilde{T}_2$. Li [16] applied a sparse sliced average variance estimation method to analyze this dataset and estimated the two directions as $(0 \times Length + 0 \times Left + 0 \times Right + 0.785 \times Bottom + 0.619 \times Top + 0 \times Diagonal)$ and $(0 \times Length + 0 \times Left + 0 \times Right + 0.400 \times Bottom + 0 \times Top + 0.917 \times Diagonal)$. With the additional information of $d = 2$, both the sparse estimation and the test for joint effects of the predictors agree that *Top*, *Bottom* and *Diagonal* are significant predictors, while *Top* is regarded as an irrelevant predictor by the marginal coordinate test.

6. Concluding remarks

We proposed two unified tests for testing the conditional coordinate hypotheses based on the first-order or second-order sufficient dimension reduction methods. The asymptotic properties were also investigated. Moreover, our tests can also be adapted to some other sufficient dimension reduction methods, as long as they admit the eigendecomposition formulation (2.2). For example, we can similarly develop conditional coordinate test for the central subspace with the inverse third order

moments method [23], the contour regression [18], the Fourier method [27], the discretization–expectation estimation method [29] and the cumulative slicing estimation method [28]. In analogy, we may also develop conditional coordinate test for the central mean subspace [6] with the principal Hessian directions [15], the iterative Hessian directions [6] and the marginal fourth order moments method [24].

Cook [4] also discussed another type of hypothesis: marginal dimension hypothesis given a coordinate constraint, which has not been studied systematically in the literature. We expect our present work would be helpful in developing a unified test of such a hypothesis with the commonly used sufficient dimension reduction methods.

Another interesting issue is to study the global behavior of the asymptotic power functions. Under local alternative hypotheses, T_1 or T_2 is expected to converge to a linear combination of non-central chi-square variables with non-centrality parameters that depend on the alternative hypothesis. Then it would be of great interest to study the power of the two test statistics against any alternative especially a local sequence of alternatives. Whether the coordinate tests in sufficient dimension reduction are consistent in power against any alternative remains an open question. Research along this direction deserves further study.

Appendix A

Proof of Proposition 1. If $\mathrm{Span}(\mathcal{M}) = \mathcal{S}_{Y|Z}$ and d is known, we have $\mathrm{Span}(\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_d) = \mathcal{S}_{Y|Z}$. Moreover, $\mathcal{P}_{\mathcal{H}}\mathcal{S}_{Y|X} = \mathcal{O}_p$ if and only if $\mathcal{W}^T\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\eta}_i = 0$ for $i = 1, \ldots, d$. Then if $\mathcal{P}_{\mathcal{H}}\mathcal{S}_{Y|X} = \mathcal{O}_p$ holds, we can derive that $\mathcal{W}^T\boldsymbol{\Sigma}^{-1/2}P = 0$ and hence $L_1 = 0$.

On the other side, observe that $L_1 = \|H^TP\|^2$, then $L_1 = 0$ implies that $H^TP = (\mathcal{W}^T\boldsymbol{\Sigma}^{-1}\mathcal{W})^{-1/2}\mathcal{W}^T\boldsymbol{\Sigma}^{-1/2}P = 0$. Because $\mathcal{W}^T\boldsymbol{\Sigma}^{-1}\mathcal{W}$ is invertible, we then have $\mathcal{W}^T\boldsymbol{\Sigma}^{-1/2}P = 0$, and hence $\mathcal{P}_{\mathcal{H}}\mathcal{S}_{Y|X} = \mathcal{O}_p$. □

Proof of Theorem 1. Let $Q(\mathbf{X}, Y)^* = -\sum_{i=1}^d[\lambda_i^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\{E_n\mathcal{M}^*(\mathbf{X}, Y)\}Q + Q\{E_n\mathcal{M}^*(\mathbf{X}, Y)\}\lambda_i^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T]$. From the perturbation theory [13] and Theorem 1 in [20], $\widehat{Q}$ can be expanded as $\widehat{Q} = Q + E_n\{Q^*(\mathbf{X}, Y)\} + o_p(n^{-1/2})$. It then follows that

$$\begin{aligned}
&(\mathcal{W}^T\widehat{\boldsymbol{\Sigma}}^{-1}W)^{-1/2}\mathcal{W}^T\widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{P}\\
&= (\mathcal{W}^T\Sigma^{-1}W)^{-1/2}\mathcal{W}^T\left[\boldsymbol{\Sigma}^{-1/2} + E_n\left\{\boldsymbol{\Sigma}^{*-1/2}(\mathbf{X}, Y)\right\}\right][P - E_n\{Q^*(\mathbf{X}, Y)\}] + o_p(n^{-1/2})\\
&= E_n\left[(\mathcal{W}^T\Sigma^{-1}W)^{-1/2}\mathcal{W}^T\left\{\boldsymbol{\Sigma}^{*-1/2}(\mathbf{X}, Y)P + \boldsymbol{\Sigma}^{-1/2}\mathcal{M}^*(\mathbf{X}, Y)\sum_{i=1}^d\lambda_i^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\right\}\right] + o_p(n^{-1/2}).
\end{aligned}$$

The last equality holds since $\mathcal{W}^T\boldsymbol{\Sigma}^{-1/2}P = 0$ and $\mathcal{W}^T\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\eta}_i = 0$ for $i = 1, \ldots, d$. The conclusion is then obvious. □

Proof of Proposition 2. Under the null hypotheses, we have $Q_W\boldsymbol{\eta}_i = \boldsymbol{\eta}_i$ and $\mathcal{M}_c\boldsymbol{\eta}_i = \lambda_i\boldsymbol{\eta}_i$, which indicates that $P = P_c$ and $L_2 = 0$. If $L_2 = 0$, we see that $\mathrm{Span}(P) \subseteq \mathrm{Span}(Q_W)$ and hence $\mathrm{Span}(P_W) \subseteq \mathrm{Span}(Q)$. Then Proposition 2 of Cook [4] suggests that $\mathcal{P}_{\mathcal{H}}\mathcal{S}_{Y|X} = \mathcal{O}_p$. □

Proof of Theorem 2. First we can expand $\widehat{\mathcal{M}}_c$ as follows:

$$\begin{aligned}
\widehat{\mathcal{M}}_c = \mathcal{M}_c &+ E_n[Q_W\mathcal{M}^*(\mathbf{X}, Y)Q_W - \boldsymbol{\Sigma}^{-1/2}\mathcal{W}(\mathcal{W}^T\boldsymbol{\Sigma}^{-1}\mathcal{W})^{-1}\mathcal{W}^T\boldsymbol{\Sigma}^{*-1/2}(\mathbf{X}, Y)\mathcal{M}Q_W\\
&- Q_WM\boldsymbol{\Sigma}^{*-1/2}(\mathbf{X}, Y)\mathcal{W}(\mathcal{W}^T\boldsymbol{\Sigma}^{-1}\mathcal{W})^{-1}\mathcal{W}^T\boldsymbol{\Sigma}^{-1/2}] + o_p(n^{-1/2}).
\end{aligned}$$

Under the null hypothesis, we have $P = P_c$, $Q_WP = P$ and $\mathcal{W}^T\boldsymbol{\Sigma}^{-1/2}P = 0$. Then

$$\begin{aligned}
\widehat{P} - \widehat{P}_c &= E_n\left[Q\{\mathcal{M}^*(\mathbf{X}, Y) - \mathcal{M}_c^*(\mathbf{X}, Y)\}\sum_{i=1}^d\lambda_i^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T + \sum_{i=1}^d\lambda_i^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\{\mathcal{M}^*(\mathbf{X}, Y) - \mathcal{M}_c^*(\mathbf{X}, Y)\}Q\right] + o_p(n^{-1/2})\\
&= E_n\left[\left\{P_W\mathcal{M}^*(\mathbf{X}, Y) + \boldsymbol{\Sigma}^{-1/2}\mathcal{W}(\mathcal{W}^T\boldsymbol{\Sigma}^{-1}\mathcal{W})^{-1}\mathcal{W}^T\boldsymbol{\Sigma}^{*-1/2}(\mathbf{X}, Y)\mathcal{M}\right\}\sum_{i=1}^d\lambda_i^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\right.\\
&\quad\left. + \sum_{i=1}^d\lambda_i^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\left\{\mathcal{M}^*(\mathbf{X}, Y)P_W + M\boldsymbol{\Sigma}^{*-1/2}(\mathbf{X}, Y)\mathcal{W}(\mathcal{W}^T\boldsymbol{\Sigma}^{-1}\mathcal{W})^{-1}\mathcal{W}^T\boldsymbol{\Sigma}^{-1/2}\right\}\right] + o_p(n^{-1/2}).
\end{aligned}$$

The conclusion is then straightforward. □

Appendix B

In this section, we give the detailed proof of Lemma 1, especially for the asymptotic expansion of $\widehat{\mathcal{M}}$ with sliced inverse regression, sliced average variance estimation and directional regression respectively.

B.1. Useful lemmas

We first deal with the asymptotic expansions of $\widehat{\boldsymbol{\Sigma}}$, $\widehat{\boldsymbol{\Sigma}}^{-1}$ and $\widehat{\boldsymbol{\Sigma}}^{-1/2}$.

Table B.1
Estimated test levels, as percentages, based on Models I, and II.

Sample size	Model I with $p = 4$				Model I with $p = 8$			
	Test[a]	Nominal level (%)			Test	Nominal level (%)		
		1	5	10		1	5	10
$n = 100$	MCT(SIR)	1.00	6.20	11.1	MCT(SIR)	3.10	11.4	20.2
	CCT(SIR)	2.00	6.90	13.7	CCT(SIR)	2.60	11.6	18.9
	T_1(SIR)	1.20	6.30	14.9	T_1(SIR)	1.80	7.80	14.1
	T_2(SIR)	0.60	5.90	13.5	T_2(SIR)	1.50	7.00	13.0
$n = 200$	MCT(SIR)	0.70	5.20	10.4	MCT(SIR)	1.60	7.80	14.5
	CCT(SIR)	1.60	6.80	13.0	CCT(SIR)	2.20	8.10	14.1
	T_1(SIR)	1.30	6.60	12.0	T_1(SIR)	0.90	6.00	11.7
	T_2(SIR)	1.10	6.10	11.7	T_2(SIR)	1.00	5.40	10.9
$n = 400$	MCT(SIR)	1.60	5.50	10.8	MCT(SIR)	1.10	6.10	11.4
	CCT(SIR)	1.50	6.10	11.4	CCT(SIR)	2.10	6.60	11.4
	T_1(SIR)	0.80	5.80	11.7	T_1(SIR)	0.80	5.90	11.9
	T_2(SIR)	0.70	5.60	11.2	T_2(SIR)	0.80	5.80	11.6
Sample size	Model II with $p = 4$				Model II with $p = 8$			
	Test	Nominal level (%)			Test	Nominal level (%)		
		1	5	10		1	5	10
$n = 100$	MCT(SIR)	1.70	7.80	13.6	MCT(SIR)	2.40	8.40	13.8
	CCT(SIR)	2.20	8.60	14.4	CCT(SIR)	2.30	7.70	13.0
	T_1(SIR)	1.30	6.00	12.0	T_1(SIR)	1.10	6.20	12.2
	T_2(SIR)	1.00	5.30	11.5	T_2(SIR)	0.80	5.60	11.2
$n = 200$	MCT(SIR)	1.10	6.40	11.7	MCT(SIR)	1.70	7.90	14.7
	CCT(SIR)	1.70	6.30	11.2	CCT(SIR)	2.30	8.10	14.2
	T_1(SIR)	1.60	6.30	11.4	T_1(SIR)	1.30	5.50	10.0
	T_2(SIR)	1.60	5.80	11.3	T_2(SIR)	1.10	5.20	9.80
$n = 400$	MCT(SIR)	0.60	5.60	11.3	MCT(SIR)	1.80	6.20	10.6
	CCT(SIR)	0.90	6.50	12.5	CCT(SIR)	1.60	7.50	12.6
	T_1(SIR)	0.90	6.10	11.0	T_1(SIR)	1.30	5.40	10.1
	T_2(SIR)	0.80	6.00	10.8	T_2(SIR)	1.20	5.10	9.90

[a] SIR: sliced inverse regression. MCT(SIR) and CCT(SIR) refer to the sliced inverse regression based general marginal coordinate test and general conditional coordinate test developed by Cook [4].

Table B.2
Estimated power (%) at 5% nominal level based on Models I, and II.

Test	Sample size			Test	Sample size		
	$n = 50$	$n = 100$	$n = 200$		$n = 50$	$n = 100$	$n = 200$
Model I with $p = 4$				Model I with $p = 8$			
MCT(SIR)	98.90	100.0	100.0	MCT(SIR)	94.20	100.0	100.0
CCT(SIR)	99.70	100.0	100.0	CCT(SIR)	98.40	100.0	100.0
T_1(SIR)	100.0	100.0	100.0	T_1(SIR)	97.80	100.0	100.0
T_2(SIR)	98.70	100.0	100.0	T_2(SIR)	92.80	100.0	100.0
Model II with $p = 5$				Model II with $p = 10$			
MCT(SIR)	100.0	100.0	100.0	MCT(SIR)	100.0	100.0	100.0
CCT(SIR)	100.0	100.0	100.0	CCT(SIR)	100.0	100.0	100.0
T_1(SIR)	100.0	100.0	100.0	T_1(SIR)	100.0	100.0	100.0
T_2(SIR)	100.0	100.0	100.0	T_2(SIR)	99.90	100.0	100.0

Lemma 2. *Let* $\mathbf{\Sigma}^* = (\mathbf{X} - \mu)(\mathbf{X} - \mu)^T - \Sigma$. *Let* P_1 *be an orthogonal matrix such that* $\mathbf{\Sigma} = P_1 C P_1^T$, *where* $C = \mathrm{diag}(c_1, \ldots, c_p)$ *is a diagonal matrix. Let* C_1 *be a square matrix having* $\frac{(c_i^{-\frac{1}{2}} - c_j^{-\frac{1}{2}})}{c_i - c_j}$ *as the* (i, j) *entry for* $i \neq j$ *and* $-\frac{1}{2}c_i^{-\frac{3}{2}}$ *as the* (i, i) *entry. Define* $\mathbf{\Sigma}^{*-1} = -\mathbf{\Sigma}^{-1}\mathbf{\Sigma}^*\mathbf{\Sigma}^{-1}$ *and* $\mathbf{\Sigma}^{*-1/2} = P_1\big(C_1 \odot (P_1^T \mathbf{\Sigma}^* P_1)\big)P_1^T$, *where* $\odot$ *denotes the Hadamard product. Then* $\widehat{\Sigma}$, $\widehat{\mathbf{\Sigma}}^{-1}$ *and* $\widehat{\mathbf{\Sigma}}^{-1/2}$ *can be expanded asymptotically as follows:*

$$\widehat{\mathbf{\Sigma}} = \mathbf{\Sigma} + E_n\{\mathbf{\Sigma}^*\} + o_p(n^{-\frac{1}{2}}),$$
$$\widehat{\mathbf{\Sigma}}^{-1} = \mathbf{\Sigma}^{-1} + E_n\{\mathbf{\Sigma}^{*-1}\} + o_p(n^{-\frac{1}{2}}),$$
$$\widehat{\mathbf{\Sigma}}^{-1/2} = \mathbf{\Sigma}^{-1/2} + E_n\{\mathbf{\Sigma}^{*-1/2}\} + o_p(n^{-\frac{1}{2}}).$$

Table B.3
Estimated test levels, as percentages, based on Models III, and VI.

Sample size	Model III with $p = 4$				Model III with $p = 8$			
	Test[a]	Nominal level (%)			Test	Nominal level (%)		
		1	5	10		1	5	10
$n = 100$	MCT(SAVE)	0.50	3.60	7.90	MCT(SAVE)	0.20	2.80	6.00
	T_1(SAVE)	1.60	6.10	12.5	T_1(SAVE)	1.30	7.00	12.6
	T_2(SAVE)	1.60	6.00	12.4	T_2(SAVE)	1.00	7.20	13.0
	T_1(DR)	1.70	7.60	13.6	T_1(DR)	2.40	7.90	13.2
	T_2(DR)	1.60	7.60	13.8	T_2(DR)	2.40	8.10	13.9
$n = 200$	MCT(SAVE)	1.10	4.20	9.00	MCT(SAVE)	0.80	3.90	8.60
	T_1(SAVE)	1.00	5.30	11.5	T_1(SAVE)	0.80	6.70	10.7
	T_2(SAVE)	0.80	4.90	11.3	T_2(SAVE)	0.90	6.40	10.5
	T_1(DR)	1.30	6.30	11.8	T_1(DR)	1.40	6.90	12.7
	T_2(DR)	1.10	6.20	11.5	T_2(DR)	1.20	7.10	12.6
$n = 400$	MCT(SAVE)	1.00	5.90	9.20	MCT(SAVE)	0.70	3.70	8.80
	T_1(SAVE)	0.90	4.00	9.10	T_1(SAVE)	0.80	4.90	10.2
	T_2(SAVE)	0.90	4.00	9.40	T_2(SAVE)	0.90	5.00	10.1
	T_1(DR)	0.80	4.00	10.4	T_1(DR)	1.00	5.30	10.5
	T_2(DR)	0.90	4.00	10.3	T_2(DR)	1.00	5.00	10.5
Sample size	Model IV with $p = 4$				Model IV with $p = 8$			
	Test	Nominal level (%)			Test	Nominal level (%)		
		1	5	10		1	5	10
$n = 100$	MCT(SAVE)	0.50	3.90	8.50	MCT(SAVE)	0.10	1.50	4.40
	T_1(SAVE)	2.20	7.90	17.4	T_1(SAVE)	0.60	8.10	22.6
	T_2(SAVE)	0.60	6.50	14.3	T_2(SAVE)	0.30	4.30	15.0
	T_1(DR)	2.30	7.70	15.9	T_1(DR)	3.00	9.80	16.9
	T_2(DR)	1.10	7.40	13.0	T_2(DR)	1.40	6.40	11.9
$n = 200$	MCT(SAVE)	0.50	4.40	9.10	MCT(SAVE)	0.50	2.20	6.60
	T_1(SAVE)	0.60	4.30	8.80	T_1(SAVE)	1.30	6.00	11.2
	T_2(SAVE)	0.60	4.50	8.00	T_2(SAVE)	0.60	5.10	9.90
	T_1(DR)	1.40	5.40	10.6	T_1(DR)	1.10	5.40	10.3
	T_2(DR)	0.90	4.40	9.60	T_2(DR)	0.70	4.70	8.60
$n = 400$	MCT(SAVE)	0.40	4.30	9.90	MCT(SAVE)	0.50	4.90	8.90
	T_1(SAVE)	1.10	4.40	8.60	T_1(SAVE)	0.70	4.50	7.80
	T_2(SAVE)	0.50	4.00	8.20	T_2(SAVE)	0.50	4.20	7.70
	T_1(DR)	0.70	6.20	11.8	T_1(DR)	0.80	4.70	9.50
	T_2(DR)	0.70	5.50	11.0	T_2(DR)	0.80	5.10	9.10

[a] SAVE: sliced average variance estimate; DR: directional regression; MCT(SAVE) refers to the general marginal coordinate test with sliced average variance estimation developed by Shaoe et al. [21].

Table B.4
Estimated powers (%) at 5% nominal level based on Models III, and VI.

Test	Sample size			Test	Sample size		
	$n = 50$	$n = 100$	$n = 200$		$n = 50$	$n = 100$	$n = 200$
Model III with $p = 4$				Model III with $p = 8$			
MCT(SAVE)	86.00	99.80	100.0	MCT(SAVE)	31.80	82.20	100.0
T_1(SAVE)	98.80	100.0	100.0	T_1(SAVE)	79.00	99.70	100.0
T_2(SAVE)	94.50	100.0	100.0	T_2(SAVE)	64.40	99.40	100.0
T_1(DR)	100.0	100.0	100.0	T_1(DR)	80.80	100.0	100.0
T_2(DR)	99.70	100.0	100.0	T_2(DR)	66.90	100.0	100.0
Model IV with $p = 4$				Model IV with $p = 8$			
MCT(SAVE)	50.80	96.70	100.0	MCT(SAVE)	7.10	24.80	89.50
T_1(SAVE)	77.40	97.10	100.0	T_1(SAVE)	19.00	77.50	99.50
T_2(SAVE)	53.70	90.70	99.80	T_2(SAVE)	10.40	57.40	98.20
T_1(DR)	84.70	99.30	100.0	T_1(DR)	44.40	93.20	99.90
T_2(DR)	68.30	94.10	100.0	T_2(DR)	28.20	81.30	99.50

Proof of Lemma 2. The asymptotic expansion of $\widehat{\mathbf{\Sigma}}$ is a classic result. The asymptotic expansions of $\widehat{\mathbf{\Sigma}}^{-1}$ and $\widehat{\mathbf{\Sigma}}^{-1/2}$ can be derived by standard procedure of Von Mises expansion in combination with Theorem 6.6.30 in [12]. □

As with the usual protocol in sufficient dimension reduction, we make a partition of the range of Y as $\{J_1, \ldots, J_h\}$. Let $p_k = E\{I(Y \in J_k)\}$, $U_k = E\{(\mathbf{X} - \mu)I(Y \in J_k)\}$ and $V_k = E\{(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T I(Y \in J_k)\}$. Denote $\hat{p}_k = E_n\{I(Y \in J_k)\}$, $\hat{U}_k =$

Table B.5
Estimated 5% nominal levels and estimated powers (%) at 5% nominal level for different choices of d based on Model VI with $n = 200$ and $p = 8$.

Estimated 5% nominal levels				Estimated powers			
Test	$d = 1$	$d = 3$	$d = 4$	Test	$d = 1$	$d = 3$	$d = 4$
T_1(SAVE)	7.60	3.80	1.20	T_1(SAVE)	98.80	45.60	11.60
T_2(SAVE)	3.10	12.4	18.8	T_2(SAVE)	100.0	92.40	65.40
T_1(DR)	0.40	4.40	1.80	T_1(DR)	100.0	46.80	8.60
T_2(DR)	0.50	11.3	15.6	T_2(DR)	100.0	87.40	64.20

Table B.6
Results from 5% backward elimination procedures based on conditional coordinate test with sliced average variance estimation for the Swiss banknote data.

Predictor	$\widetilde{T}_1$(SAVE)	df	p-value	$\widetilde{T}_2$(SAVE)	df	p-value
(a) Conditional coordinate test results with all six predictors. ($d = 2$)						
Length	2.007	1	0.157	2.040	1	0.153
Left	0.187	1	0.665	0.186	1	0.666
Right	0.676	1	0.411	0.686	1	0.408
Top	34.49	1	0.000	24.95	1	0.000
Bottom	14.32	1	0.000	13.28	1	0.000
Diagonal	25.50	1	0.001	17.50	1	0.000
(b) Conditional coordinate test results with five predictors. ($d = 2$)						
Length	3.148	1	0.076	3.067	1	0.080
Right	1.180	1	0.277	1.170	1	0.280
Top	45.52	1	0.000	25.59	1	0.000
Bottom	18.09	1	0.000	16.28	1	0.000
Diagonal	35.00	1	0.000	20.83	1	0.000
(c) Conditional coordinate test results with four predictors. ($d = 2$)						
Length	5.033	1	0.025	4.834	1	0.028
Top	46.06	1	0.000	26.19	1	0.000
Bottom	17.08	1	0.000	15.44	1	0.000
Diagonal	33.33	1	0.000	19.53	1	0.000

Table B.7
Conditional coordinate test results for the Swiss banknote data with different null hypotheses.

$\widetilde{T}_1$(SAVE)	df	p-value	$\widetilde{T}_2$(SAVE)	df	p-value
(a) $H_0 : Y \perp\!\!\!\perp$ (*Length,Top, Left, Right*) given (*Bottom, Diagonal*) ($d = 2$)					
54.72	3	0.000	27.10	2	0.000
(b) $H_0 : Y \perp\!\!\!\perp$ (*Length,Left,Right*) given (*Top, Bottom, Diagonal*) ($d = 2$)					
5.858	3	0.119	5.822	3	0.120

$E_n\{(\mathbf{X}-\hat{\mu})I(Y \in J_k)\}$ and $\widehat{V}_k = E_n\{(\mathbf{X}-\hat{\mu})(\mathbf{X}-\hat{\mu})^T I(Y \in J_k)\}$ be the corresponding sample estimators. The following lemma is useful for deriving the asymptotic expansion of $\widehat{\mathcal{M}}$.

Lemma 3. *Let* $\mu^* = \mathbf{X} - \mu$, $p_k^* = I(Y \in J_k) - p_k$, $U_k^* = (\mathbf{X} - \mu)I(Y \in J_k) - U_k + p_k(\mathbf{X} - \mu)$, $V_k^* = (\mathbf{X} - \mu)(\mathbf{X} - \mu)^T I(Y \in J_k) - V_k - U_k^*\mu^T - U_k(\mu^*)^T$. *We have the following expansions:*

$$\hat{\mu} = \mu + E_n(\mu^*) + o_p(n^{-1/2}); \qquad \hat{p}_k = p_k + E_n(p_k^*) + o_p(n^{-1/2});$$
$$\widehat{U}_k = U_k + E_n(U_k^*) + o_p(n^{-1/2}); \qquad \widehat{V}_k = V_k + E_n(V_k^*) + o_p(n^{-1/2});$$
$$\hat{p}_k^{-1} = p_k^{-1} - E_n(p_k^2 p_k^*) + o_p(n^{-1/2}); \qquad \hat{p}_k^{-2} = p_k^{-2} - E_n(2p_k^3 p_k^*) + o_p(n^{-1/2});$$
$$\hat{p}_k^{-3} = p_k^{-3} - E_n(3p_k^4 p_k^*) + o_p(n^{-1/2}).$$

Proof of Lemma 3. These expansion can be derived by Von Mises expansion techniques, see [17]. We omit the details here. □

B.2. Asymptotic expansion of $\widehat{\mathcal{M}}_{\mathrm{SIR}}$

In this section, we consider the asymptotic expansion of the estimated candidate matrix of sliced inverse regression [14]. Define $\Lambda_{\mathrm{SIR}} = \sum_{l=1}^{h} p_l E(\mathbf{X} - \mu | Y \in J_l)\{E(\mathbf{X} - \mu | Y \in J_l)\}^T$. It is easy to see that $\Lambda_{\mathrm{SIR}} = \sum_{l=1}^{h} p_l^{-1} U_l U_l^T$ and $\mathcal{M}_{\mathrm{SIR}} =$

$\mathbf{\Sigma}^{-1/2}\Lambda_{\mathrm{SIR}}\mathbf{\Sigma}^{-1/2}$. Then their sample estimators are $\widehat{\Lambda}_{\mathrm{SIR}} = \sum_{l=1}^{h}\hat{p}_l^{-1}\widehat{U}_l\widehat{U}_l^T$ and $\widehat{\mathcal{M}}_{\mathrm{SIR}} = \widehat{\mathbf{\Sigma}}^{-1/2}\widehat{\Lambda}_{\mathrm{SIR}}\widehat{\mathbf{\Sigma}}^{-1/2}$. We in the following give the explicit expansion forms of $\widehat{\Lambda}_{\mathrm{SIR}}$ and $\widehat{\mathcal{M}}_{\mathrm{SIR}}$.

Lemma 4. *Let* $\Lambda^*_{\mathrm{SIR}} = \sum_{l=1}^{h}(-\frac{p_l^* U_l U_l^T}{p_l^2} + \frac{U_l^* U_l^T}{p_l} + \frac{U_l U_l^{*T}}{p_l})$, *then we have the expansion* $\widehat{\Lambda}_{\mathrm{SIR}} = \Lambda_{\mathrm{SIR}} + E_n(\Lambda^*_{\mathrm{SIR}}) + o_p(n^{-1/2})$.

Proof of Lemma 4. The conclusion can be derived by using the expansions of $\hat{p}_l^{-1}$, $\widehat{U}_l$ and $\hat{\mu}$ given in Lemma 3. The details are omitted here. □

Theorem 3. $\widehat{\mathcal{M}}_{\mathrm{SIR}}$ *can be expanded asymptotically as* $\widehat{\mathcal{M}}_{\mathrm{SIR}} = \widehat{\mathcal{M}}_{\mathrm{SIR}} + E_n(\mathcal{M}^*_{\mathrm{SIR}}) + o_p(n^{-1/2})$, *where* $\mathcal{M}^*_{\mathrm{SIR}} = \mathbf{\Sigma}^{*-1/2}\Lambda_{\mathrm{SIR}}\mathbf{\Sigma}^{-1/2} + \mathbf{\Sigma}^{-1/2}\Lambda^*_{\mathrm{SIR}}\mathbf{\Sigma}^{-1/2} + \mathbf{\Sigma}^{-1/2}\Lambda_{\mathrm{SIR}}\mathbf{\Sigma}^{*-1/2}$.

Proof of Theorem 3. With the expansion of $\widehat{\mathbf{\Sigma}}^{-1/2}$ given in Lemma 2, the conclusion can be easily derived by invoking Lemma 4. □

If sliced inverse regression is used, A and B can be simplified under the null hypothesis as stated in the following corollary.

Corollary 1. *Let* $U_k^{**} = (\mathbf{X}-\mu)I(Y\in J_k) + p_k(\mathbf{X}-\mu)$, $V_k^{**} = (\mathbf{X}-\mu)(\mathbf{X}-\mu)^T I(Y\in J_k) - p_k\mathbf{\Sigma} - U_k^{**}\mu^T$, $\Lambda^*_{\mathrm{SIR}} = \sum_{l=1}^{h}\frac{U_l^{**}U_l^T}{p_l}$ *and* $\mathcal{M}^{**}_{\mathrm{SIR}} = \mathbf{\Sigma}^{*-1/2}\Lambda_{\mathrm{SIR}}\mathbf{\Sigma}^{-1/2} + \mathbf{\Sigma}^{-1/2}\Lambda^{**}_{\mathrm{SIR}}\mathbf{\Sigma}^{-1/2}$. *Then if sliced inverse regression is used, A and B in Theorems 1 and 2 are defined as:*

$$A = (\mathcal{W}^T\Sigma^{-1}\mathcal{W})^{-1/2}\mathcal{W}^T\left\{\mathbf{\Sigma}^{*-1/2}P + \mathbf{\Sigma}^{-1/2}\mathcal{M}^{**}_{\mathrm{SIR}}\sum_{i=1}^{d}\lambda_i^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\right\},$$

$$B = \left\{P_{\mathcal{W}}\mathcal{M}^{**}_{\mathrm{SIR}} + \mathbf{\Sigma}^{-1/2}\mathcal{W}(\mathcal{W}^T\mathbf{\Sigma}^{-1}\mathcal{W})^{-1}\mathcal{W}^T\mathbf{\Sigma}^{*-1/2}\mathcal{M}\right\}\sum_{i=1}^{d}\lambda_i^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T.$$

Proof of Corollary 1. Under the null hypothesis, we have $\mathcal{W}^T\mathbf{\Sigma}^{-1}U_l/p_l = 0$ and $\mathcal{W}^T\mathbf{\Sigma}^{-1}\mathcal{M}_{\mathrm{SIR}}\mathbf{\Sigma}^{-1/2} = 0$. Then we can derive the simplified forms of A and B as given in the above. □

B.3. Asymptotic expansion of $\widehat{\mathcal{M}}_{\mathrm{SAVE}}$

In this section, we consider the asymptotic expansion of the estimated candidate matrix of sliced average variance estimation [9]. Let $\Lambda_{\mathrm{SAVE}} = E[\mathbf{\Sigma} - \mathrm{Var}\{\mathbf{X}|\delta(Y)\}]\mathbf{\Sigma}^{-1}[\mathbf{\Sigma} - \mathrm{Var}\{\mathbf{X}|\delta(Y)\}]$, where $\delta(Y) = \sum_{l=1}^{h} lI(Y\in J_l)$. Then $\mathcal{M}_{\mathrm{SAVE}} = \sum_{l=1}^{h}[I_p - \mathrm{Var}\{Z|\delta(Y)\}]^2 = \mathbf{\Sigma}^{-1/2}\Lambda_{\mathrm{SAVE}}\mathbf{\Sigma}^{-1/2}$. The following lemma expresses Λ_{SAVE} in terms of μ, Σ, p_l, U_l and V_l.

Lemma 5. $\Lambda_{\mathrm{SAVE}} = 2\Lambda_{\mathrm{SIR}} - \mathbf{\Sigma} + \Gamma$, *where* $\Gamma = \sum_{l=1}^{h}(\Gamma_l^1 - \Gamma_l^2 - \Gamma_l^3 + \Gamma_l^4)$ *with* $\Gamma_l^1 = \frac{V_l\mathbf{\Sigma}^{-1}V_l}{p_l}$, $\Gamma_l^2 = \frac{V_l\mathbf{\Sigma}^{-1}U_lU_l^T}{p_l^2}$, $\Gamma_l^3 = \frac{U_lU_l^T\mathbf{\Sigma}^{-1}V_l}{p_l^2}$ *and* $\Gamma_l^4 = \frac{U_lU_l^T\mathbf{\Sigma}^{-1}U_lU_l^T}{p_l^3}$.

Proof of Lemma 5. By some algebra calculations, we can derive that

$$\Lambda_{\mathrm{SAVE}} = \mathbf{\Sigma} - 2\mathrm{E}[\mathrm{Var}\{\mathbf{X}|\delta(Y)\}] + E[\mathrm{Var}\{\mathbf{X}|\delta(Y)\}\mathbf{\Sigma}^{-1}\mathrm{Var}\{\mathbf{X}|\delta(Y)\}].$$

The EV-VE formula gives that $\Lambda_{\mathrm{SAVE}} = 2\Lambda_{\mathrm{SIR}} - \mathbf{\Sigma} + E[\mathrm{Var}\{\mathbf{X}|\delta(Y)\}\mathbf{\Sigma}^{-1}\mathrm{Var}\{\mathbf{X}|\delta(Y)\}]$. Moreover, we can check that $E[\mathrm{Var}\{\mathbf{X}|\delta(Y)\}\mathbf{\Sigma}^{-1}\mathrm{Var}\{\mathbf{X}|\delta(Y)\}]$ is equal to Γ as defined in this lemma. □

Let $\widehat{\Gamma}_l^1$, $\widehat{\Gamma}_l^1$, $\widehat{\Gamma}_l^1$, $\widehat{\Gamma}_l^4$ and $\widehat{\Lambda}_{\mathrm{SAVE}}$ be the sample estimators of Γ_l^1, Γ_l^2, Γ_l^3, Γ_l^4 and Λ_{SAVE} respectively. Define

$$\left(\Gamma_l^1\right)^* = -\frac{p_l^* V_l\mathbf{\Sigma}^{-1}V_l}{p_l^2} + \frac{V_l^*\mathbf{\Sigma}^{-1}V_l}{p_l} + \frac{V_l\mathbf{\Sigma}^{*-1}V_l}{p_l} + \frac{V_l\mathbf{\Sigma}^{-1}V_l^*}{p_l},$$

$$\left(\Gamma_l^2\right)^* = -2\frac{p_l^* V_l\mathbf{\Sigma}^{-1}U_lU_l^T}{p_l^3} + \frac{V_l^*\mathbf{\Sigma}^{-1}U_lU_l^T}{p_l^2} + \frac{V_l\mathbf{\Sigma}^{*-1}U_lU_l^T}{p_l^2} + \frac{V_l\mathbf{\Sigma}^{-1}U_l^*U_l^T}{p_l^2} + \frac{V_l\mathbf{\Sigma}^{-1}U_lU_l^{*T}}{p_l^2},$$

$$\left(\Gamma_l^3\right)^* = -2\frac{p_l^* U_lU_l^T\mathbf{\Sigma}^{-1}V_l}{p_l^3} + \frac{U_l^*U_l^T\mathbf{\Sigma}^{-1}V_l}{p_l^2} + \frac{U_lU_l^{*T}\mathbf{\Sigma}^{-1}V_l}{p_l^2} + \frac{U_lU_l^T\mathbf{\Sigma}^{*-1}V_l}{p_l^l} + \frac{U_lU_l^T\mathbf{\Sigma}^{-1}V_l^*}{p_l^2},$$

$$\begin{aligned}\left(\Gamma_l^4\right)^* = {} & -3\frac{p_l^* U_lU_l^T\mathbf{\Sigma}^{-1}U_lU_l^T}{p_l^4} + \frac{U_l^*U_l^T\mathbf{\Sigma}^{-1}U_lU_l^T}{p_l^3} + \frac{U_lU_l^{*T}\mathbf{\Sigma}^{-1}U_lU_l^T}{p_l^3} + \frac{U_lU_l^T\mathbf{\Sigma}^{*-1}U_lU_l^T}{p_l^3} \\ & + \frac{U_lU_l^T\mathbf{\Sigma}^{-1}U_l^*U_l^T}{p_l^3} + \frac{U_lU_l^T\mathbf{\Sigma}^{-1}U_lU_l^{*T}}{p_l^3}.\end{aligned}$$

The following lemma and theorem confirm the expansion forms of $\widehat{\Lambda}_{\mathrm{SAVE}}$ and $\widehat{\mathcal{M}}_{\mathrm{SAVE}}$.

Lemma 6. *Let* $\Gamma^* = \sum_{l=1}^{h}\{(\Gamma_l^1)^* - (\Gamma_l^2)^* - (\Gamma_l^3)^* + (\Gamma_l^4)^*\}$ *and* $\Lambda^*_{\mathrm{SAVE}} = 2\Lambda^*_{\mathrm{SIR}} - \boldsymbol{\Sigma}^* + \Gamma^*$. *Then we have* $\widehat{\Lambda}_{\mathrm{SAVE}} = \Lambda_{\mathrm{SAVE}} + \Lambda^*_{\mathrm{SAVE}} + o_p(n^{-1/2})$.

Proof of Lemma 6. The conclusion can be derived by Lemmas 2–5. Details are omitted. □

Theorem 4. $\widehat{\mathcal{M}}_{\mathrm{SAVE}}$ *can be expanded asymptotically as*

$$\widehat{\mathcal{M}}_{\mathrm{SAVE}} = \widehat{\mathcal{M}}_{\mathrm{SAVE}} + E_n(\mathcal{M}^*_{\mathrm{SAVE}}) + o_p(n^{-1/2}),$$

where $\mathcal{M}^*_{\mathrm{SAVE}} = \boldsymbol{\Sigma}^{*-1/2}\Lambda_{\mathrm{SAVE}}\boldsymbol{\Sigma}^{-1/2} + \boldsymbol{\Sigma}^{-1/2}\Lambda^*_{\mathrm{SAVE}}\boldsymbol{\Sigma}^{-1/2} + \boldsymbol{\Sigma}^{-1/2}\Lambda_{\mathrm{SAVE}}\boldsymbol{\Sigma}^{*-1/2}$.

Proof of Theorem 4. With the expansion of $\widehat{\boldsymbol{\Sigma}}^{-1/2}$ given in Lemma 2, the conclusion can be easily derived by invoking Lemma 6. □

Similar to Corollary 1, we simplify the expressions of A and B under null hypothesis when slice average variance estimation is used.

Corollary 2. *Define* $\left(\Gamma_l^1\right)^{**} = -p_l^*\boldsymbol{\Sigma} + \frac{V_l^{**}\boldsymbol{\Sigma}^{-1}V_l}{p_l} + \boldsymbol{\Sigma}\boldsymbol{\Sigma}^{*-1}V_l + V_l^{**}$, $\left(\Gamma_l^2\right)^{**} = \frac{V_l^{**}\boldsymbol{\Sigma}^{-1}U_lU_l^T}{p_l^2} + \frac{\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{*-1}U_lU_l^T}{p_l} + \frac{U_l^{**}U_l^T}{p_l}$, $\left(\Gamma_l^3\right)^{**} = \frac{U_l^{**}U_l^T\boldsymbol{\Sigma}^{-1}V_l}{p_l^2}$, $\left(\Gamma_l^4\right)^{**} = \frac{U_l^{**}U_l^T\boldsymbol{\Sigma}^{-1}U_lU_l^T}{p_l^3}$, $\Gamma^{**} = \sum_{h=1}^{H}\{(\Gamma_l^1)^{**} - (\Gamma_l^2)^{**} - (\Gamma_l^3)^{**} + (\Gamma_l^4)^{**}\}$, $\Lambda^{**}_{\mathrm{SAVE}} = 2\Lambda^{**}_{\mathrm{SIR}} - \boldsymbol{\Sigma}^* + \Gamma^{**}$ *and* $\mathcal{M}^{**}_{\mathrm{SAVE}} = \boldsymbol{\Sigma}^{*-1/2}\Lambda_{\mathrm{SAVE}}\boldsymbol{\Sigma}^{-1/2} + \boldsymbol{\Sigma}^{-1/2}\Lambda^{**}_{\mathrm{SAVE}}\boldsymbol{\Sigma}^{-1/2}$. *Then if sliced average variance estimation is used, A and B in Theorems 1 and 2 are defined as:*

$$A = (\mathcal{W}^T\Sigma^{-1}\mathcal{W})^{-1/2}\mathcal{W}^T\left\{\boldsymbol{\Sigma}^{*-1/2}P + \boldsymbol{\Sigma}^{-1/2}\mathcal{M}^{**}_{\mathrm{SAVE}}\sum_{i=1}^{d}\lambda_i^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\right\},$$

$$B = \left\{P_{\mathcal{W}}\mathcal{M}^{**}_{\mathrm{SAVE}} + \boldsymbol{\Sigma}^{-1/2}\mathcal{W}(\mathcal{W}^T\boldsymbol{\Sigma}^{-1}\mathcal{W})^{-1}\mathcal{W}^T\boldsymbol{\Sigma}^{*-1/2}\mathcal{M}\right\}\sum_{i=1}^{d}\lambda_i^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T.$$

Proof of Corollary 2. Under the null hypothesis, we have $\mathcal{W}^T\boldsymbol{\Sigma}^{-1}U_l/p_l = 0$, $\mathcal{W}^T\boldsymbol{\Sigma}^{-1}(I_p - V_l/p_l)\boldsymbol{\Sigma}^{-1/2} = 0$ and $\mathcal{W}^T\boldsymbol{\Sigma}^{-1}\mathcal{M}_{\mathrm{SAVE}}\boldsymbol{\Sigma}^{-1/2} = 0$. Then the conclusion can be derived by some algebra calculations. □

B.4. Asymptotic expansion of $\widehat{\mathcal{M}}_{\mathrm{DR}}$

In this section, we consider the asymptotic expansion of the estimated candidate matrix of directional regression [17]. The candidate matrix of directional regression is

$$\begin{aligned}\mathcal{M}_{\mathrm{DR}} = {} & 2E[E^2\{\mathbf{Z}\mathbf{Z}^T|\delta(Y)\}] + 2E^2[E\{\mathbf{Z}|\delta(Y)\}E\{\mathbf{Z}^T|\delta(Y)\}] + 2E[E\{\mathbf{Z}^T|\delta(Y)\}E\{\mathbf{Z}|\delta(Y)\}] \\ & \times E[E\{\mathbf{Z}|\delta(Y)\}E\{\mathbf{Z}^T|\delta(Y)\}] - 2I_p.\end{aligned}$$

We first rewrite $\mathcal{M}_{\mathrm{DR}}$ as given in the following lemma.

Lemma 7. *Let* $\Phi_1 = \sum_{l=1}^{h}\frac{1}{p_l}U_l^T\boldsymbol{\Sigma}^{-1}U_l$, *then* $\mathcal{M}_{\mathrm{DR}}$ *can be reformulated as* $\mathcal{M}_{\mathrm{DR}} = \boldsymbol{\Sigma}^{-1/2}\Lambda_{\mathrm{DR}}\boldsymbol{\Sigma}^{-1/2}$, *where*

$$\Lambda_{\mathrm{DR}} = 2\sum_{l=1}^{h}\Gamma_l^1 + 2(\Lambda_{\mathrm{SIR}})^2 + 2\Phi_1\Lambda_{\mathrm{SIR}} - 2I_p.$$

Proof of Lemma 7. The conclusion can be derived by algebra calculations. We omit the details here. □

Let $\widehat{\Lambda}_{\mathrm{DR}}$ and $\widehat{\mathcal{M}}_{\mathrm{DR}}$ be the sample estimators of Λ_{DR} and $\mathcal{M}_{\mathrm{DR}}$ respectively.

Lemma 8. *Let* $\Phi_1^* = \sum_{l=1}^{h}(-\frac{p_l^*U_l^T\boldsymbol{\Sigma}^{-1}U_l}{p_l^2} + \frac{U_l^{*T}\boldsymbol{\Sigma}^{-1}U_l}{p_l} + \frac{U_l^T\boldsymbol{\Sigma}^{*-1}U_l}{p_l} + \frac{U_l^T\boldsymbol{\Sigma}^{-1}U_l^*}{p_l})$ *and* $\Lambda^*_{\mathrm{DR}} = 2\sum_{l=1}^{h}(\Gamma_l^1)^* + 2\Lambda_{\mathrm{SIR}}\Lambda^*_{\mathrm{SIR}} + 2\Lambda^*_{\mathrm{SIR}}\Lambda_{\mathrm{SIR}} + 2\Phi_1^*\Lambda_{\mathrm{SIR}} + 2\Phi_1\Lambda^*_{\mathrm{SIR}}$. *Then we have the expansion* $\widehat{\Lambda}_{\mathrm{DR}} = \Lambda_{\mathrm{DR}} + \Lambda^*_{\mathrm{DR}} + o_p(n^{-1/2})$.

Proof of Lemma 8. The conclusion can be derived by Lemmas 2–4 and 7. Details are omitted. □

Theorem 5. $\widehat{\mathcal{M}}_{\mathrm{DR}}$ *can be expanded asymptotically as*

$$\widehat{\mathcal{M}}_{\mathrm{DR}} = \widehat{\mathcal{M}}_{\mathrm{DR}} + E_n(\mathcal{M}^*_{\mathrm{DR}}) + o_p(n^{-1/2}),$$

where $\mathcal{M}^*_{\mathrm{DR}} = \boldsymbol{\Sigma}^{*-1/2}\Lambda_{\mathrm{DR}}\boldsymbol{\Sigma}^{-1/2} + \boldsymbol{\Sigma}^{-1/2}\Lambda^*_{\mathrm{DR}}\boldsymbol{\Sigma}^{-1/2} + \boldsymbol{\Sigma}^{-1/2}\Lambda_{\mathrm{DR}}\boldsymbol{\Sigma}^{*-1/2}$.

Proof of Theorem 5. With the expansion of $\widehat{\mathbf{\Sigma}}^{-1/2}$ given in Lemma 2, the conclusion can be easily derived by invoking Lemma 8. □

Similar to Corollaries 1 and 2, we simplify the expressions of A and B under null hypothesis when directional regression is used.

Corollary 3. *Let* $\Lambda_{\mathrm{DR}}^{**} = 2\sum_{l=1}^{h}(\Gamma_l^1)^{**} + 2\Lambda_{\mathrm{SIR}}^{**}\Lambda_{\mathrm{SIR}} + 2\Phi_1\Lambda_{\mathrm{SIR}}^{**}$ *and* $\mathcal{M}_{\mathrm{DR}}^{**} = \mathbf{\Sigma}^{*-1/2}\Lambda_{\mathrm{DR}}\mathbf{\Sigma}^{-1/2} + \mathbf{\Sigma}^{-1/2}\Lambda_{\mathrm{DR}}^{**}\mathbf{\Sigma}^{-1/2}$. *Then if directional regression is used, A and B in Theorems 1 and 2 are defined as:*

$$A = (\mathcal{W}^T\Sigma^{-1}\mathcal{W})^{-1/2}\mathcal{W}^T\left\{\mathbf{\Sigma}^{*-1/2}P + \mathbf{\Sigma}^{-1/2}\mathcal{M}_{\mathrm{DR}}^{**}\sum_{i=1}^{d}\lambda_i^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\right\},$$

$$B = \left\{P_{\mathcal{W}}\mathcal{M}_{\mathrm{DR}}^{**} + \mathbf{\Sigma}^{-1/2}\mathcal{W}(\mathcal{W}^T\mathbf{\Sigma}^{-1}\mathcal{W})^{-1}\mathcal{W}^T\mathbf{\Sigma}^{*-1/2}\mathcal{M}\right\}\sum_{i=1}^{d}\lambda_i^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T.$$

Proof of Corollary 3. The proof is similar to that for Corollaries 1 and 2. Details are omitted. □

References

[1] P. Bentler, J. Xie, Corrections to test statistics in principal Hessian directions, Statist. Probab. Lett. 47 (2000) 381–389.
[2] C.-H. Chen, K.C. Li, Can SIR be as popular as multiple linear regression? Statist. Sinica 8 (1998) 289–316.
[3] R.D. Cook, Regression Graphics, Wiley, New York, 1998.
[4] R.D. Cook, Testing predictor contributions in sufficient dimension reduction, Ann. Statist. 32 (2004) 1062–1092.
[5] R.D. Cook, H. Lee, Dimension reduction in binary response regression, J. Amer. Statist. Assoc. 94 (1999) 1187–1200.
[6] R.D. Cook, B. Li, Dimension reduction for conditional mean in regression, Ann. Statist. 30 (2002) 455–474.
[7] R.D. Cook, L. Ni, Sufficient dimension reduction via inverse regression: a minimum discrepancy approach, J. Amer. Statist. Assoc. 100 (2005) 410–428.
[8] R.D. Cook, L. Ni, Using intraslice covariances for improved estimation of the central subspace in regression, Biometrika 93 (2006) 65–74.
[9] R.D. Cook, S. Weisberg, Discussion of "sliced inverse regression for dimension reduction", J. Amer. Statist. Assoc. 86 (1991) 316–342.
[10] B. Flury, H. Riedwyl, Multivariate Statistics, A Practical Approach, John Wiley and Sons, New York, 1988.
[11] K.F. Fung, X. He, L. Liu, P. Shi, Dimension reduction based on canonical correlation, Statist. Sinica 12 (2002) 1093–1113.
[12] R.A. Horn, C.R. Johnson, Topics in Matrix Analysis, Cambridge University Press, 1991.
[13] K. Kato, A Short Introduction to Perturbation Theory for Linear Operators, Springer-Verlag, New York, 1982.
[14] K.C. Li, Sliced inverse regression for dimension reduction (with discussion), J. Amer. Statist. Assoc. 86 (1991) 316–342.
[15] K.C. Li, On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma, J. Amer. Statist. Assoc. 87 (1992) 1025–1039.
[16] L. Li, Sparse sufficient dimension reduction, Biometrika 94 (2007) 603–613.
[17] B. Li, S. Wang, On directional regression for dimension reduction, J. Amer. Statist. Assoc. 102 (2007) 997–1008.
[18] B. Li, H. Zha, C. Chiaromonte, Contour regression: a general approach to dimension reduction, Ann. Statist. 33 (2005) 1580–1616.
[19] P.A. Naik, C.L. Tsai, Constrained inverse regression for incorporating prior information, J. Amer. Statist. Assoc. 100 (2005) 204–211.
[20] J.R. Schott, Asymptotics of eigenprojections of correlation matrices with some applications in principal components analysis, Biometrika 84 (1997) 327–337.
[21] Y. Shao, R.D. Cook, S. Weisberg, Marginal tests with sliced average variance estimation, Biometrika 94 (2007) 285–296.
[22] Y. Xia, H. Tong, W.K. Li, L.X. Zhu, An adaptive estimation of optimal regression subspace, J. R. Stat. Soc. Ser. B 64 (2002) 363–410.
[23] X. Yin, R.D. Cook, Estimating central subspace via inverse third moment, Biometrika 90 (2003) 113–125.
[24] X. Yin, R.D. Cook, Dimension reduction via marginal fourth moments in regression, J. Comput. Graph. Statist. 13 (2004) 554–570.
[25] X. Yin, B. Li, R.D. Cook, Successive direction extraction for estimating the central subspace in a multiple-index regression, J. Multivariate Anal. 99 (2008) 1733–1757.
[26] L.-X. Zhu, K.-T. Fang, Asymptotics for kernel estimate of sliced inverse regression, Ann. Statist. 3 (1996) 1053–1068.
[27] Y. Zhu, P. Zeng, Fourier methods for estimating the central subspace and the central mean subspace in regression, J. Amer. Statist. Assoc. 101 (2006) 1638–1651.
[28] L.P. Zhu, L.-X. Zhu, Z.H. Feng, Dimension reduction in regressions via average partial mean estimation, J. Amer. Statist. Assoc. 105 (2010) 1455–1466.
[29] L.P. Zhu, L.-X. Zhu, L. Ferré, T. Wang, Sufficient dimension reduction through discretization–expectation estimation, Biometrika 97 (2010) 295–304.