On robustness in dimension determination in fused sliced inverse regression

Jae Keun Yoo[1,a], Yoo Na Cho[a]

[a]Department of Statistics, Ewha Womans University, Korea

Abstract

The goal of sufficient dimension reduction (SDR) is to replace original p-dimensional predictors with a lower-dimensional linearly transformed predictor. The sliced inverse regression (SIR) (Li, *Journal of the American Statistical Association*, **86**, 316–342, 1991) is one of the most popular SDR methods because of its applicability and simple implementation in practice. However, SIR may yield different dimension reduction results for different numbers of slices and despite its popularity, is a clear deficit for SIR. To overcome this, a fused sliced inverse regression was recently proposed. The study shows that the dimension-reduced predictors is robust to the numbers of the slices, but it does not investigate how robust its dimension determination is. This paper suggests a permutation dimension determination for the fused sliced inverse regression that is compared with SIR to investigate the robustness to the numbers of slices in the dimension determination. Numerical studies confirm this and a real data example is presented.

Keywords: central subspace, fused sliced inverse regression, permutation test, sufficient dimension reduction

1. Introduction

Consider a regression of $Y \in \mathbb{R}^1 | \mathbf{X} \in \mathbb{R}^p = (X_1, \ldots, X_p)^{\mathrm{T}}$. Then, sufficient dimension reduction (SDR) in regression replaces the original p-dimensional predictors $\mathbf{X}$ with a lower-dimensional linearly transformed predictor $\mathbf{M}^{\mathrm{T}}\mathbf{X}$ without loss of information about $Y|\mathbf{X}$ such that

$$Y \perp\!\!\!\perp \mathbf{X}|\mathbf{M}^{\mathrm{T}}\mathbf{X}, \tag{1.1}$$

where $\mathbf{M} \in \mathbb{R}^{p \times q}$, $\perp\!\!\!\perp$ stands for independence, and $q \leq p$.

Statement (1.1) implies that the two regressions of $Y|\mathbf{X}$ and $Y|\mathbf{M}^{\mathrm{T}}\mathbf{X}$ are equivalent; therefore, $\mathbf{X}$ can be replaced by $\mathbf{M}^{\mathrm{T}}\mathbf{X}$ without loss of information about $Y|\mathbf{X}$. A column subspace of $\mathbf{M}$ satisfying (1.1) is defined as a dimension reduction subspace. If the intersection of all possible dimension reduction subspaces is a dimension reduction subspace, it is minimal and unique among all possible dimension reduction subspaces. The intersection is defined as the central subspace $\mathcal{S}_{Y|\mathbf{X}}$. Therefore, the estimation of $\mathcal{S}_{Y|\mathbf{X}}$ is the main goal of SDR. The rest of the paper denotes the true dimension and orthonormal basis matrix of $\mathcal{S}_{Y|\mathbf{X}}$ as d and $\boldsymbol{\eta} \in \mathbb{R}^{p \times d}$, respectively. The lower-dimensional linearly transformed predictor $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$ is called *sufficient predictors*.

The sliced inverse regression (SIR) (Li, 1991) is one of the most popular SDR methods due to large applicability and simple implementation in practice. The key step of the SIR application is a

---

categorization of a response variable Y. This categorization is called slicing. The performance of SIR does not theoretically depend on the numbers of slices; however, its sample behaviors to estimate $\mathcal{S}_{Y|\mathbf{X}}$ are sensitive to the numbers of slices. Even critically, there is no optimal or recommended one, which can be adopted as a golden standard. To relieve this issue, Cook and Zhang (2014) recently suggest an approach to fuse sample kernel matrices of SIR constructed from various choices of slices. This fusing approach is known as fused sliced inverse regression (FSIR). According to Cook and Zhang (2014), the FSIR results in robust basis estimation of $\mathcal{S}_{Y|\mathbf{X}}$ to the numbers of slices.

The inference of $\mathcal{S}_{Y|\mathbf{X}}$ requires two steps. First, the structural dimension d of $\mathcal{S}_{Y|\mathbf{X}}$ should be determined, and the basis $\boldsymbol{\eta}$ is estimated given that $d = \hat{d}$. However, Cook and Zhang (2014) do not provide the clear dimension determination and its asymptotic behaviors.

This paper investigates robustness in the dimension estimation of $\mathcal{S}_{Y|\mathbf{X}}$ in FSIR by employing a permutation approach. If FSIR is sensitive to determine the structural dimension, then this accordingly affects the robustness in the basis estimation. This study can completely prove the potential advantages of FSIR over SIR in robust and possibly better estimation of $\mathcal{S}_{Y|\mathbf{X}}$.

The organization of the paper is as follows. In Section 2, we briefly discuss a FSIR and suggest a permutation dimension determination. Section 3 is devoted to numerical studies and the presentation of a real data application. Our work is summarized in Section 4. We will define the following notations, which will be used frequently throughout the rest of the paper. A subspace $\mathcal{S}(\mathbf{B})$ stands for a subspace spanned by the columns of $\mathbf{B} \in \mathbb{R}^{p \times q}$; in addition, we define that $\boldsymbol{\Sigma} = \mathrm{cov}(\mathbf{X})$.

2. Permutation dimension determination in fused sliced inverse regression

2.1. Sliced inverse regression

The method of SIR (Li, 1991) will be explained through the normalized predictor $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - E(\mathbf{X}))$ of $\mathbf{X}$. For notational convenience, define $\mathcal{S}_{Y|\mathbf{Z}}$ and $\boldsymbol{\eta}_z$ as the central subspace for a regression of $Y|\mathbf{Z}$ and its $p \times d$ orthonormal basis matrix , respectively. Proposition 6.3 in Cook (1998) guarantees that the structural dimensions of $\mathcal{S}_{Y|\mathbf{X}}$ and $\mathcal{S}_{Y|\mathbf{Z}}$ are equal and that $\mathcal{S}_{Y|\mathbf{X}} = \boldsymbol{\Sigma}^{-1/2}\mathcal{S}_{Y|\mathbf{Z}}$. Assume that the following linearity condition holds: (C1) $E(\mathbf{Z}|\boldsymbol{\eta}_z^{\mathrm{T}}\mathbf{Z})$ is linear in $\boldsymbol{\eta}_z^{\mathrm{T}}\mathbf{Z}$. Li (1991) shows the following relation under the linearity condition:

$$\mathcal{S}(E(\mathbf{Z}|Y)) \subseteq \mathcal{S}_{Y|\mathbf{Z}} \Leftrightarrow \mathcal{S}\left(\boldsymbol{\Sigma}^{-1} E(\mathbf{X}|Y)\right) \subseteq \mathcal{S}_{Y|\mathbf{X}}.$$

To guarantee the exhaustive estimation of $\mathcal{S}_{Y|\mathbf{Z}}$, it is usually assumed that $\mathcal{S}(E(\mathbf{Z}|Y)) = \mathcal{S}_{Y|\mathbf{Z}}$. An estimation of $\mathcal{S}_{Y|\mathbf{X}}$ through $E(\mathbf{X}|Y)$ is called SIR.

Since any specific models of $Y|\mathbf{Z}$ are not assumed in SIR, $E(\mathbf{Z}|Y)$ should be restored nonparametically. If Y is categorical with h levels, the construction of $E(\mathbf{Z}|Y)$ is straightforward, which is a set of the means of $\mathbf{Z}$ within each level of Y. If Y is continuous or many-valued, the computation of $E(\mathbf{Z}|Y)$ is unclear. For its simple but effective estimation in the case, Y is categorized with h levels. The categorized response will be denoted as $\tilde{Y}$, and the computation of $E(\mathbf{Z}|\tilde{Y} = s)$ is straightforward. Assuming that $\mathcal{S}(E(\mathbf{Z}|Y)) = \mathcal{S}(E(\mathbf{Z}|\tilde{Y} = s))$, $E(\mathbf{Z}|Y)$ is easily and nonparametrically recovered, assuming that $\mathcal{S}(E(\mathbf{Z}|Y)) = \mathcal{S}(E(\mathbf{Z}|\tilde{Y} = s))$. This categorization of Y is often called slicing, which is the essential part in the SIR application in practice. Next, SIR constructs the kernel matrix of either $\mathbf{M}_{\mathrm{SIR}} = \mathrm{cov}(E(\mathbf{Z}|Y))$ or $\mathbf{M}_{\mathrm{SIR}} = \mathrm{cov}(E(\mathbf{Z}|\tilde{Y}))$, depending on the type of Y. Then the columns of $\mathbf{M}_{\mathrm{SIR}}$ spans $\mathcal{S}_{Y|\mathbf{Z}}$.

The sample SIR algorithm is as:

1. Construct $\tilde{Y}$ by partitioning the range of Y into h non-overlapping intervals to have equal numbers

of the observations as much as possible, if Y is continuous or many-valued. If Y is categorical, the original Y becomes $\tilde{Y}$. Denote n_s as the number of observations for $\tilde{Y} = s$ for $s = 1, \ldots, h$.

2. Construct $\hat{\mathbf{Z}}_i = \hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{X}_i - \bar{\mathbf{X}})$, and then compute $\hat{E}(\mathbf{Z}|\tilde{Y} = s) = (1/n_s)\sum_{\tilde{Y}=s} \hat{\mathbf{Z}}_i$ for $s = 1, \ldots, h$.

3. Compute the sample version $\hat{\mathbf{M}}_{\text{SIR}}$ of $\mathbf{M}_{\text{SIR}}$ as:

$$\hat{\mathbf{M}}_{\text{SIR}} = \widehat{\text{cov}}\left(E\left(\mathbf{Z}|\tilde{Y}\right)\right) = \sum_{s=1}^{h} \frac{n_s}{n} \hat{E}\left(\mathbf{Z}|\tilde{Y} = s\right) \hat{E}\left(\mathbf{Z}|\tilde{Y} = s\right)^{\text{T}}.$$

4. Spectral-decompose $\hat{\mathbf{M}}_{\text{SIR}}$ such that $\hat{\mathbf{M}}_{\text{SIR}} = \sum_{i=1}^{p} \hat{\alpha}_j \hat{\phi}_j \hat{\phi}_j^{\text{T}}$, where $\hat{\alpha}_1 \geq \cdots \geq \hat{\alpha}_p \geq 0$ are the ordered-eigenvalues of $\hat{\mathbf{M}}_{\text{SIR}}$ and $\hat{\phi}_i$, $i = 1, \ldots, p$, is the corresponding eigenvector to $\hat{\alpha}_i$.

5. The structural dimension d is determined through a sequential test, and set $d = \hat{d}$.

6. The eigenvectors $\hat{\phi}_i$ corresponding to the first $\hat{d}$ largest $\hat{\alpha}_i$s become the estimate of $\boldsymbol{\eta}_z$:

$$\hat{\boldsymbol{\eta}}_z = \left(\hat{\phi}_1, \ldots, \hat{\phi}_{\hat{d}}\right)$$

to have the sample basis estimate for $\boldsymbol{\eta}$ so that $\hat{\boldsymbol{\eta}} = \hat{\boldsymbol{\Sigma}}^{-1/2}\hat{\boldsymbol{\eta}}_z$.

In the sample implementation of SIR, the construction of $\hat{\mathbf{M}}_{\text{SIR}}$ clearly depends on the numbers of h. In practice, different numbers of h yield different asymptotic results in basis and dimension estimation of $\mathcal{S}_{Y|\mathbf{Z}}$, and hence this should be worrisome in the data application. Moreover, there is no golden standard on how many slices should be used.

2.2. Fused sliced inverse regression

To overcome the deficit of the sensitiveness of SIR to the numbers of slices, a fused approach is developed by Cook and Zhang (2014) to combine the $\mathbf{M}_{\text{SIR}}$ constructed from various numbers of slices. Now define

$$\mathbf{E}_{\text{F}}^{(h)} = \left(E\left(\mathbf{Z}|\tilde{Y}_{(h_1)}\right), E\left(\mathbf{Z}|\tilde{Y}_{(h_2)}\right), \ldots, E\left(\mathbf{Z}|\tilde{Y}_{(h_K)}\right)\right) \quad \text{and} \quad \mathbf{M}_{\text{FSIR}}^{(h)} = \mathbf{E}_{\text{F}}^{(h)}\mathbf{E}_{\text{F}}^{(h)\text{T}},$$

where $Y_{(h_K)}$ indicates the categorized response into h_K slices.

Since $\mathbf{M}_{\text{FSIR}}^{(h)}$ is a non-decreasing sequence over h, the following relation holds that

$$\mathcal{S}\left(\mathbf{M}_{\text{SIR}}^{(k)}\right) \subseteq \mathcal{S}\left(\mathbf{M}_{\text{FSIR}}^{(h)}\right), \quad k = 2, 3, \ldots, h.$$

Furthermore, the assumed equivalence of $\mathcal{S}(\mathbf{M}_{\text{SIR}}^{(k)}) = \mathcal{S}_{Y|\mathbf{Z}}$ guarantees that

$$\mathcal{S}\left(\mathbf{M}_{\text{FSIR}}^{(h)}\right) = \mathcal{S}_{Y|\mathbf{Z}} \quad \text{and} \quad \mathcal{S}\left(\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{M}_{\text{FSIR}}^{(h)}\right) = \mathcal{S}_{Y|\mathbf{X}}.$$

This directly implies that the columns of $\mathbf{M}_{\text{FSIR}}^{(h)}$ spans $\mathcal{S}_{Y|\mathbf{Z}}$. The inference on $\mathcal{S}_{Y|\mathbf{X}}$ through $\mathbf{M}_{\text{FSIR}}^{(h)}$ is called FSIR.

The sample version $\hat{\mathbf{M}}_{\text{FSIR}}^{(h)}$ is computed by constructing $\hat{\mathbf{M}}_{\text{SIR}}^{(k)}$ via usual SIR application:

$$\hat{\mathbf{M}}_{\text{FSIR}}^{(h)} = \left(\hat{\mathbf{M}}_{\text{SIR}}^{(2)}, \hat{\mathbf{M}}_{\text{SIR}}^{(3)}, \ldots, \hat{\mathbf{M}}_{\text{SIR}}^{(h)}\right).$$

The inference procedure of $\mathcal{S}_{Y|\mathbf{Z}}$ via $\hat{\mathbf{M}}_{\text{FSIR}}^{(h)}$ is the same as that through SIR. Let $\hat{\mathbf{M}}_{\text{FSIR}}^{(h)} = \sum_{i=1}^{p} \hat{\lambda}_i \hat{\gamma}_i \hat{\gamma}_i^{\text{T}}$, be the spectral decomposition of $\hat{\mathbf{M}}_{\text{FSIR}}^{(h)}$, where $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p \geq 0$ are the ordered-eigenvalues of $\hat{\mathbf{M}}_{\text{FSIR}}^{(h)}$ and $\hat{\gamma}_i$ is its corresponding eigenvector for $i = 1, \ldots, p$.

Cook and Zhang (2014) provide that FSIR is robust in the basis estimation of $\mathcal{S}_{Y|\mathbf{Z}}$ to the number of slices compared with SIR, and they estimate $\boldsymbol{\eta}_z$ with the true structural dimension given. However, its result clearly affects the basis estimation since the dimension should be determined before the basis estimation. That is, if the dimension determination is not robust to the numbers of slices, then it affects the selection of the eigenvectors $\hat{\lambda}_i$. Accordingly, this falsely impacts the robustness in the estimation of $\mathcal{S}_{Y|\mathbf{Z}}$.

For this, in the next section, a permutation approach is suggested to estimate the structural dimension d of $\mathcal{S}_{Y|\mathbf{Z}}$ through FSIR.

2.3. Permutation dimension determination

It is necessary to first define related hypothesis before estimating the true structural dimension. For this, the following sequence of hypothesis (Rao, 1965) is tested. Starting with $m = 0$, test $H_0 : d = m$ versus $H_1 : d = m+1$. If $H_0 : d = m$ is rejected, increment m by 1 and repeat the test, stopping the first time H_0 is not rejected and setting $\hat{d} = m$. To conduct this sequential test, a test statistic is necessary. To construct the test statistics, it should be observed that the dimension of $\mathcal{S}_{Y|\mathbf{Z}}$ is equal to the true rank of $\mathbf{M}_{\text{FSIR}}^{(h)}$, because $\mathcal{S}(\mathbf{M}_{\text{FSIR}}^{(h)}) = \mathcal{S}_{Y|\mathbf{Z}}$. Then the inference on d is equal to the rank estimation of $\mathbf{M}_{\text{FSIR}}^{(h)}$. This is equivalent to test how many non-zero eigenvalues exist for its spectral decomposition. Based on this discussion, the related test statistics for the dimension determination is suggested as

$$\hat{\Lambda}_m = n \sum_{i=m+1}^{p} \hat{\lambda}_i,$$

where n is a sample size.

To avoid the difficulty in deriving the asymptotics of $\hat{\Lambda}_m$ under H_0, a permutation dimension determination approach is adopted here. The permutation approach has had a long history in SDR literature, but it is still popularly used (Cook and Weisberg, 1991; Cook and Yin, 2002; Yin and Cook, 2002; Lee *et al.*, 2013; Dong *et al.*, 2016). One possible drawback is that it needs an additional condition of

$$\left(Y, \boldsymbol{\Gamma}_1^{\text{T}} \mathbf{Z}\right) \perp\!\!\!\perp \boldsymbol{\Gamma}_2^{\text{T}} \mathbf{Z},$$

where $\boldsymbol{\Gamma}_1 = (\gamma_1, \ldots, \gamma_m)$ is a set of eigenvectors corresponding to the m largest eigenvalues of $\mathbf{M}_{\text{FSIR}}^{(h)}$ under $H_0 : d = m$, and $\boldsymbol{\Gamma}_2 \in \mathbb{R}^{p \times (p-m)}$ is the orthogonal complement of $\boldsymbol{\Gamma}_1$. If $\mathbf{Z}$ is normally distributed, the condition is satisfied under H_0. Since the predictors are often transformed to normality to satisfy linearity, the additional condition would not be strong in practice.

The permutation dimension determination algorithm is:

1. Construct $\hat{\mathbf{M}}_{\text{FSIR}}^{(h)}$. Under $H_0 : d = m$, obtain $\hat{\Lambda}_m$ and divide eigenvector matrices into the following two groups:

$$\hat{\boldsymbol{\Gamma}}_1 = (\hat{\gamma}_1, \ldots, \hat{\gamma}_m) \quad \text{and} \quad \hat{\boldsymbol{\Gamma}}_2 = \left(\hat{\gamma}_{m+1}, \ldots, \hat{\gamma}_p\right).$$

2. Compute two sets predictors of $\hat{V}_i \in \mathbb{R}^{m \times 1} = \hat{\boldsymbol{\Gamma}}_1^{\text{T}} \hat{\mathbf{Z}}_i$ and $\hat{U}_i \in \mathbb{R}^{(p-m) \times 1} = \hat{\boldsymbol{\Gamma}}_2^{\text{T}} \hat{\mathbf{Z}}_i$, for $i = 1, \ldots, n$.

3. Randomly permute the index i of the $\hat{U}_i$ alone and construct the permuted set $\hat{U}_i^*$.

4. Compute the test statistic $\hat{\Lambda}_m^*$ via the FSIR application on a regression of $Y_i|(\hat{\mathbf{V}}_i, \hat{\mathbf{U}}_i^*)$.

5. Repeat steps (3)–(4) N times, where N is the total number of permutations. The p-value of the hypothesis testing is the fraction of $\hat{\Lambda}_m^*$ that exceed $\hat{\Lambda}_m$.

In this permutation determination procedure, it had better be noted that the p-values may not have an increasing tendency as m increases. The number of permutation $N = 1000$ will be used in numerical studies and real data analysis.

3. Numerical studies and real data application

3.1. Numerical studies

We considered the following two regression models:

- Model 1: $Y|\mathbf{X} = (X_1, \ldots, X_{10})^{\mathrm{T}} = X_1 + \varepsilon$.

- Model 2: $Y|\mathbf{X} = (X_1, \ldots, X_{10})^{\mathrm{T}} = X_1 + \exp(X_2)\varepsilon$.

For both models, the predictors of $\mathbf{X} = (X_1, \ldots, X_{10})^{\mathrm{T}}$ and the random error ε were independently sampled either from $N(0, 1)$ or from t_{10}.

For Model 1, the structural dimension of $\mathcal{S}_{Y|\mathbf{X}}$ is equal to one, and $\mathcal{S}_{Y|\mathbf{X}}$ is spanned by $\boldsymbol{\eta} = (1, 0, \ldots, 0)^{\mathrm{T}}$. The structure of the regression is quite simple. All information on the regression for predictors is completely given in the first conditional moment only through X_1. However, in Model 2, the two columns of $\boldsymbol{\eta} = ((1, 0, \ldots, 0), (0, 1, 0, \ldots, 0))^{\mathrm{T}}$ spans $\mathcal{S}_{Y|\mathbf{X}}$, and its true dimension is equal to two. The regression depends on its first two conditional moments. One sufficient predictor X_1 explains the conditional mean, while the other sufficient predictor X_2 contributes to the conditional variance. Therefore, Model 2 has a more complex regression than Model 1.

Tho sample sizes of $n = 100$ and 200 were considered, and each model was iterated 500 times with $h = 4, 5, 7, 9$. The 5% nominal level was used for all simulation. For SIR, if predictors are normally-distributed, the dimension was determined through a χ^2 test; if not, a weighted sum of independent χ^2s was used.

As a summary of the numerical studies, the percentages of the correct dimension determination for each model were initially computed. Furthermore, to investigate the robustness in the dimension estimation, for each iteration of the model, the concordant decisions with its true dimension were investigated with two or more slice choices. Let $\hat{d}_h$ be the estimate of the dimension by SIR or FSIR with h slices. Then, for each model, the following percentages were additionally computed.

2 pairs	$\hat{d}_4 = \hat{d}_5 = d$	$\hat{d}_4 = \hat{d}_7 = d$	$\hat{d}_4 = \hat{d}_9 = d$	
	$\hat{d}_5 = \hat{d}_7 = d$	$\hat{d}_5 = \hat{d}_9 = d$	$\hat{d}_7 = \hat{d}_9 = d$	
3 pairs	$\hat{d}_4 = \hat{d}_5 = \hat{d}_7 = d$	$\hat{d}_4 = \hat{d}_5 = \hat{d}_9 = d$	$\hat{d}_4 = \hat{d}_7 = \hat{d}_9 = d$	$\hat{d}_5 = \hat{d}_7 = \hat{d}_9 = d$
All	$\hat{d}_4 = \hat{d}_5 = \hat{d}_7 = \hat{d}_9 = d$			

For Model 1 and 2, the dimension d is one and two, respectively. If SIR and FSIR are robust to the numbers of slices in the dimension estimation, the correct decision percentages for each number of slices and the 2–4 pairs of the equal decision percentages must be close to 95%. These are reported in Figure 1. For each figure, the horizontal axis represents various choices of slices. For example, "479" stands for the percentages of $\hat{d}_4 = \hat{d}_7 = \hat{d}_9 = d$ and "all" does for those of $\hat{d}_4 = \hat{d}_5 = \hat{d}_7 = \hat{d}_9 = d$.

(a) Model 1 with $N(0, 1)$

(b) Model 1 with t_{10}
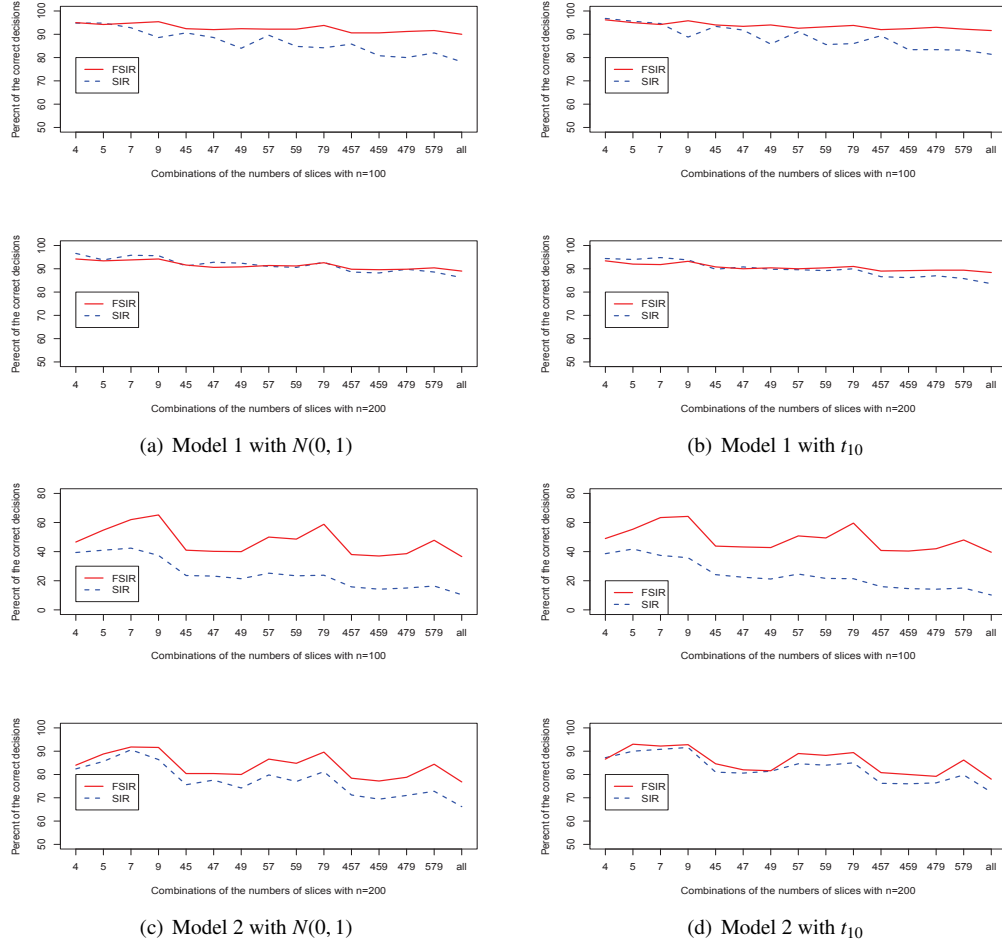
(c) Model 2 with $N(0, 1)$

(d) Model 2 with t_{10}

Figure 1: *Summary for numerical studies; Solid line, FSIR; Dashed line, SIR. SIR = sliced inverse regression; FSIR = fused SIR.*

Figure 1 shows characteristic behaviors in the dimension estimation. In most numerical studies, FSIR shows better dimension determination results than SIR. With smaller sample size, FSIR is more robust to the numbers of slices than SIR. With complex regression models, FSIR is impacted by the numbers of slices, which is less affected than SIR. FSIR shows more consistent behaviors in the dimension estimation to the distribution of the variables than SIR. As discussed in Cook and Zhang (2014), FSIR yields more robust and better estimation result to the number of slices than SIR. This directly impacts the dimension estimation, so FSIR yields more robust results than SIR. This numerical studies confirm that FSIR should have a potential advantage over SIR in both dimension and basis estimation of $\mathcal{S}_{Y|\mathbf{X}}$.

3.2. Real data example: primary biliary cirrhosis data

For illustration purposes, primary biliary cirrhosis (PBC) data in Tibshirani (1997) and Yoo (2017) is analyzed. This data was collected at the Mayo Clinic between 1974 and 1986 and contains the

following 19 variables with 276 observations:

Y = the number of days between registration and the earlier of death or censoring

δ = 1, if Y is time to death; 0 otherwise

X_1 Treatment code: 1 = D-penicillamine, 2 = placebo

X_2 Age in years

X_3 Gender: 0 = male, 1 = female

X_4 Presence of ascites: absent = 0 or present = 1

X_5 Presence of hepatomegaly: absent = 0 or present = 1

X_6 Presence of spiders: 0 = no or 1 = yes

X_7 Presence of edema: absent and no diuretic therapy = 0, present but no diuretic therapy or edema resolved by diuretics = 0.5, or present despite diuretic therapy = 1

X_8 Serum bilirubin, in mg/dl

X_9 Serum cholesterol, in mg/dl

X_{10} Albumin, in g/dl

X_{11} Urine copper, in μg/day

X_{12} Alkaline phosphatase, in U/liter

X_{13} SGOT, in U/ml

X_{14} Triglycerides, in mg/dl

X_{15} Platelet count; coded value is number of platelets per cubic ml of blood divided by 1,000

X_{16} Prothrombin time, in seconds

X_{17} Histologic state of disease, graded 1, 2, 3, or 4

We also considered PBC data because it often used in survival regression. According to Cook (2003), the SIR application in survival regression requires bivariate slicing of the observed survival time and the censoring status. In this analysis, the observed survival time was sliced first into 2, 3, 4, and 5 categories, and then the observations with each category were secondly partitioned by two groups depending on the censoring status. Therefore, the numbers of slices under consideration were 4, 6, 8, and 10 for SIR. The same slicing scheme was applied for FSIR; in addition, we considered three fusing cases of $(4, 6)$, $(4, 6, 8)$, and $(4, 6, 8, 10)$. Table 1 reports the p-values for $H_0 : d = m$ for $m = 0, 1, 2, 3$.

According to Table 1, with level 5%, the SIR application to PBC data determines that $\hat{d} = 2$ with 4 and 10 slices and $\hat{d} = 3$ with 6 and 8 slices. Therefore, with SIR, the dimension determination would be confusing. However, FSIR determines that $\hat{d} = 3$ for all three cases and supports the different estimation results in SIR that originated from the different numbers of slices.

According to Yoo (2017), for the same data, FSIR shows more robust basis estimation than SIR for each case of $d = 1$, $d = 2$, and $d = 3$. Therefore, the robustness in the dimension estimation coincides with the basis estimation.

Table 1: P-values for $H_0 : d = m$ by SIR and FSIR; SIR#, SIR with # slices; FSIR#, FSIR upto # slices

	SIR4	SIR6	SIR8	SIR10	FSIR6	FSIR8	FSIR10
$H_0 : d = 0$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$H_0 : d = 1$	0.000	0.000	0.000	0.000	0.000	0.000	0.000
$H_0 : d = 2$	0.454	0.020	0.009	0.062	0.032	0.008	0.010
$H_0 : d = 3$	N/A	0.725	0.286	0.955	0.610	0.331	0.631

SIR = sliced inverse regression; FSIR = fused SIR.

4. Conclusion

This paper investigates robustness in the dimension determination by FSIR (Cook and Zhang, 2014) over SIR (Li, 1991). A permutation approach is employed to avoid difficulty in deriving related test statistics.

Numerical studies show that FSIR has more robust dimension estimation to the numbers of slices than SIR, and confirms that FSIR has a potential advantage in an inference on the central subspace over SIR.

Usage of the asymptotic distribution of the test statistics in FSIR enables avoiding the additional requirement of the permutation test. It also reduces the computing time for the related p-values and leaves a room for improving robustness in the dimension determination.

Acknowledgments

References

Cook RD (1998). *Regression Graphics: Ideas for Studying Regressions through Graphics*, Wiley, New York.

Cook RD (2003). Dimension reduction and graphical exploration in regression including survival analysis, *Statistics in Medicine*, **22**, 1399–1413.

Cook RD and Weisberg S (1991). Discussion of sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, **86**, 328–332.

Cook RD and Yin X (2001). Dimension reduction and visualization in discriminant analysis, *Australian and New Zealand Journal of Statistics*, **43**, 147–199.

Cook RD and Zhang X (2014). Fused estimators of the central subspace in sufficient dimension reduction, *Journal of the American Statistical Association*, **109**, 815–827.

Dong Y, Yang C, and Yu Z (2016). On permutation tests for predictor contribution in sufficient dimension reduction, *Journal of Multivariate Analysis*, **149**, 81–91.

Lee K, Oh S and Yoo JK (2013). Method-free permutation predictor hypothesis tests in sufficient dimension reduction, *Communications for Statistical Applications and Methods*, **20**, 291–300.

Li KC (1991). Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, **86**, 316–342.

Rao CR (1965). *Linear Statistical Inference and Its Application*, Wiley, New York.

Tibshirani R (1997). The lasso method for variable selection in the Cox model, *Statistics in Medicine*, **16**, 385–395.

Yin X and Cook RD (2002). Dimension reduction for the conditional kth moment in regression, *Journal of the Royal Statistical Society, Series B*, **64**, 159–175.

Yoo JK (2017). Fused sliced inverse regression in survival analysis, *Communications for Statistical Applications and Methods*, **24**, 533–541.