

# Fused sliced inverse regression in survival analysis

Jae Keun Yoo<sup>1,a</sup>

<sup>a</sup>Department of Statistics, Ewha Womans University, Korea

---

## Abstract

Sufficient dimension reduction (SDR) replaces original  $p$ -dimensional predictors to a lower-dimensional linearly transformed predictor. The sliced inverse regression (SIR) has the longest and most popular history of SDR methodologies. The critical weakness of SIR is its known sensitive to the numbers of slices. Recently, a fused sliced inverse regression is developed to overcome this deficit, which combines SIR kernel matrices constructed from various choices of the number of slices. In this paper, the fused sliced inverse regression and SIR are compared to show that the former has a practical advantage in survival regression over the latter. Numerical studies confirm this and real data example is presented.

Keywords: bivariate slicing, fused sliced inverse regression, sufficient dimension reduction, survival analysis

---

## 1. Introduction

Sufficient dimension reduction (SDR) in regression of  $\mathbf{Y} \in \mathbb{R}^1 | \mathbf{X} \in \mathbb{R}^p = (X_1, \dots, X_p)^T$  tries to replace the original  $p$ -dimensional predictors  $\mathbf{X}$  with a lower-dimensional linear projection predictor without loss of information about the conditional distribution of  $Y | \mathbf{X}$ . That is, SDR seeks to find  $\mathbf{M} \in \mathbb{R}^{p \times q}$ :

$$\mathbf{Y} \perp\!\!\!\perp \mathbf{X} | \mathbf{M}^T \mathbf{X}, \quad (1.1)$$

where  $\perp\!\!\!\perp$  stands for independence, and  $q \leq p$ .

Statement (1.1) indicates that the conditional distributions of  $Y | \mathbf{X}$  and  $Y | \mathbf{M}^T \mathbf{X}$  are the same; therefore, the dimension of  $\mathbf{X}$  is replaced by  $\mathbf{M}^T \mathbf{X}$  without loss of information about  $Y | \mathbf{X}$ . A subspace spanned by the columns of  $\mathbf{M}$  satisfying (1.1) is called a dimension reduction subspace. If the intersection of all dimension reduction subspaces is a dimension reduction subspace, it is called the central subspace  $\mathcal{S}_{Y | \mathbf{X}}$ . The central subspace is minimal and unique with a restoration that forms the main stream of SDR literature. Hereafter, the true dimension and orthonormal basis matrix of  $\mathcal{S}_{Y | \mathbf{X}}$  will be denoted as  $d$  and  $\boldsymbol{\eta} \in \mathbb{R}^{p \times d}$ , respectively. The lower-dimensional linear projection predictor  $\boldsymbol{\eta}^T \mathbf{X}$  is called *sufficient predictors*.

One of the most popular SDR methods should be sliced inverse regression (SIR) (Li, 1991). Implementation of SIR requires a categorization of a response variable  $Y$ , called slicing. The selection of the appropriate number of slices are known to be often critical in the application results; in addition, there are no optimal or recommended choices that can be used as a thumb rule. Cook and Zhang (2014) propose a simple solution to overcome this problem by combining sample kernel matrices of SIR constructed from various numbers of slices. Cook and Zhang (2014) call the combining approach

---

<sup>1</sup> Department of Statistics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea.  
E-mail: peter.yoo@ewha.ac.kr

fused sliced inverse regression (FSIR). According to Cook and Zhang (2014), the FSIR results in robust basis estimates of  $\mathcal{S}_{Y|X}$  to the different numbers of slices, varying 3 to 15.

This paper studies FSIR in survival regression (which is out of date) and compares it with SIR under consideration of various slicing schemes. Slicing with a two dimensional response is more complex than that with a dimensional response since the survival regression takes the observed survival time and the censoring indicator as response variables. Therefore, FSIR is expected to have potential advantages in robustness to the basis estimation of  $\mathcal{S}_{Y|X}$  over SIR.

The organization of the paper is as follows. In Section 2, we discuss a fused sliced inverse regression along with applicability to survival regression. Some issues that can arise in slicing under survival regression are also discussed in the same section. Section 3 is devoted to numerical studies and presentation of a real data application. We summarize our work 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  $B \in \mathbb{R}^{p \times q}$ . We also define that  $\Sigma = \text{cov}(\mathbf{X})$ .

## 2. Fused sliced inverse regression in survival regression

### 2.1. Fused sliced inverse regression

Before explaining SIR (Li, 1991), the predictor  $\mathbf{X}$  is normalized to  $\mathbf{Z} = \Sigma^{-1/2}(\mathbf{X} - E(\mathbf{X}))$ . Letting  $\mathcal{S}_{Y|Z}$  be the central subspace for a regression of  $Y|Z$ , then the relationship that  $\mathcal{S}_{Y|X} = \Sigma^{-1/2}\mathcal{S}_{Y|Z}$  holds. Define  $\eta_z$  be  $p \times d$  orthonormal basis matrix for  $\mathcal{S}_{Y|Z}$ . Consider the linearity condition: (C1)  $E(\mathbf{Z}|\eta_z^T \mathbf{Z})$  is linear in  $\eta_z^T \mathbf{Z}$ . According to Li (1991), a proper subspace of  $\mathcal{S}_{Y|Z}$  can be constructed under linearity condition:

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

For the exhaustive estimation of  $\mathcal{S}_{Y|X}$ , it is typically assumed that  $\mathcal{S}(\Sigma^{-1}E(\mathbf{X}|Y)) = \mathcal{S}_{Y|X}$ . An estimation of  $\mathcal{S}_{Y|X}$  through  $E(\mathbf{X}|Y)$  is called sliced inverse regression.

In population, the construction of  $E(\mathbf{Z}|Y)$  should be done without assuming any specific distributional assumption of  $Y|Z$ . If  $Y$  is categorical with  $h$  levels,  $E(\mathbf{Z}|Y = s)$  is the mean of  $\mathbf{Z}$  within each category of  $Y$ . Following this idea, if  $Y$  is continuous or many-valued,  $Y$  is transformed to a categorized response  $\tilde{Y}$  with  $h$  levels. Then compute  $E(\mathbf{Z}|\tilde{Y} = s)$  for  $s = 1, \dots, h$ . This categorization of  $Y$  is called slicing, which is done so that each category has an equal number of observations. The SIR constructs the kernel matrix of either  $\mathbf{M}_{\text{SIR}} = \text{cov}(E(\mathbf{Z}|Y))$  or  $\hat{\mathbf{M}}_{\text{SIR}} = \text{cov}(E(\mathbf{Z}|\tilde{Y}))$ . The sample algorithm for SIR is summarized 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. Denote  $n_s$  as the number of observations for  $\tilde{Y} = s$ .
2. Compute  $\hat{E}(\mathbf{Z}|\tilde{Y} = s) = (1/n_s) \sum_{\tilde{Y}=s} \hat{\mathbf{Z}}_i$ , where  $\hat{\mathbf{Z}}_i = \hat{\Sigma}^{-1}(\mathbf{X}_i - \bar{\mathbf{X}})$  and  $n_s$  stands for the sample size in the  $s$  the slices.
3. Construct  $\hat{\mathbf{M}}_{\text{SIR}}$  as:

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

4. Perform the spectral decomposition on  $\hat{\mathbf{M}}_{\text{SIR}}$ :  $\hat{\mathbf{M}}_{\text{SIR}} = \sum_{i=1}^p \hat{\lambda}_i \hat{\gamma}_i \hat{\gamma}_i^T$ , where  $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_p \geq 0$  are the ordered-eigenvalues of  $\hat{\mathbf{M}}_{\text{SIR}}$  and  $\hat{\gamma}_i$ ,  $i = 1, \dots, p$ , is the corresponding eigenvector to  $\hat{\lambda}_i$ .

5. Determine the structural dimension  $d$ . Let  $\hat{d}$  be an estimate of  $d$ .
6. A set of the eigenvectors corresponding to the first  $\hat{d}$  largest eigenvalues is the estimate of  $\eta_z$ :

$$\hat{\eta}_z = (\hat{\gamma}_1, \dots, \hat{\gamma}_{\hat{d}}).$$

7. Back-transform to obtain the sample basis estimate for  $\eta$  so that  $\hat{\eta} = \hat{\Sigma}^{-1/2} \hat{\eta}_z$ .

In the implementation of SIR in practice, the construction of  $\hat{\mathbf{M}}_{\text{SIR}}$  is the main part, which depends heavily on the selection of  $h$ . It is therefore expected that the estimation results by  $\hat{\mathbf{M}}_{\text{SIR}}$  may be critically different from the numbers of slices.

## 2.2. Fused sliced inverse regression

To overcome the deficit that SIR is sensitive to the number of slices, Cook and Zhang (2014) propose a fused approach to combine the  $\mathbf{M}_{\text{SIR}}$  obtained from the various number of slices. First, define that

$$\mathbf{M}_{\text{FSIR}}^{(h)} = (\mathbf{M}_{\text{SIR}}^{(2)}, \dots, \mathbf{M}_{\text{SIR}}^{(h)}),$$

where  $\mathbf{M}_{\text{SIR}}^{(h)}$  stands for the kernel matrix of SIR with  $h$  slices.

Since  $\mathbf{M}_{\text{FSIR}}^{(h)}$  is a non-decreasing sequence of  $\mathbf{M}_{\text{SIR}}^{(k)}$ , it is straightforward to hold that

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

The assumption of  $\mathcal{S}(\mathbf{M}_{\text{SIR}}^{(k)}) = \mathcal{S}_{Y|Z}$  then induces that

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

Therefore,  $\mathbf{M}_{\text{FSIR}}^{(h)}$  is a kernel matrix, whose column spans  $\mathcal{S}_{Y|Z}$ . The estimation of  $\mathcal{S}_{Y|X}$  through  $\mathbf{M}_{\text{FSIR}}^{(h)}$  will be called FSIR.

The sample version  $\hat{\mathbf{M}}_{\text{FSIR}}^{(h)}$  is constructed by replacing  $\mathbf{M}_{\text{SIR}}^{(k)}$  with  $\hat{\mathbf{M}}_{\text{SIR}}^{(k)}$ :

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

The inference of  $\mathcal{S}_{Y|Z}$  through  $\hat{\mathbf{M}}_{\text{FSIR}}^{(h)}$  is the same as that through  $\hat{\mathbf{M}}_{\text{SIR}}$ . First, spectral-decompose  $\hat{\mathbf{M}}_{\text{FSIR}}^{(h)}$  such that

$$\hat{\mathbf{M}}_{\text{FSIR}}^{(h)} = \sum_{i=1}^p \hat{\alpha}_i \hat{\phi}_i \hat{\phi}_i^T,$$

where  $\hat{\alpha}_1 \geq \dots \geq \hat{\alpha}_p \geq 0$  are the ordered-eigenvalues of  $\hat{\mathbf{M}}_{\text{FSIR}}^{(h)}$  and  $\hat{\phi}_i$ ,  $i = 1, \dots, p$ , is its eigenvector. Then, the set of the eigenvectors corresponding to the first  $\hat{d}$  largest eigenvalues is the estimate of  $\eta_z$ . Then,  $\eta$  is estimated by  $\hat{\Sigma}^{-1/2} \hat{\eta}_z$ .

It should be noted that any choice of  $h$  in  $\hat{\mathbf{M}}_{\text{FSIR}}^{(h)}$  would give the same asymptotic results, but non asymptotic behavior can be affected by choices. According to Cook and Zhang (2014), the FSIR is quite robust to  $h$ , compared to SIR.

### 2.3. Application to survival regression

Survival regression is a study of the conditional distribution of survival time  $T$ , given a set of predictors  $\mathbf{X}$ . The direct regression of  $T|\mathbf{X}$  is not normally plausible since the true survival time  $T$  cannot be fully observed due to censoring. Instead, the data  $(Y_i, \delta_i, \mathbf{X}_i)$ ,  $i = 1, \dots, n$ , are assumed as  $n$  i.i.d. realizations of  $(T, C, \mathbf{X})$ , where  $Y = T\delta + C(1 - \delta)$ ,  $\delta = 0, 1$  is an indicator variable with  $\delta = I(C > T)$ ,  $C$  is a censoring time that is considered right-censoring. Accordingly, SDR in the survival regression seeks to estimate the central subspace  $\mathcal{S}_{T|\mathbf{X}}$ :

$$T \perp\!\!\!\perp \mathbf{X} | \boldsymbol{\eta}^T \mathbf{X}. \quad (2.1)$$

Since  $T$  is not fully unobservable, the direct applications of SIR and FISIR on  $Y|\mathbf{X}$  are not possible to recover  $\mathcal{S}_{T|\mathbf{X}}$ . It is therefore necessary to establish a plausible regression up to recover  $\mathcal{S}_{T|\mathbf{X}}$ . For this, consider a regression of  $(T, C)|\mathbf{X}$ . By construction of  $(T, C)|\mathbf{X}$ , we have  $\mathcal{S}_{T|\mathbf{X}} \subseteq \mathcal{S}_{(T,C)|\mathbf{X}}$ . According to Cook (2003), the central subspace  $\mathcal{S}_{(Y,\delta)|\mathbf{X}}$  of the bivariate regression of  $(Y, \delta)|\mathbf{X}$  is informative to  $\mathcal{S}_{(T,C)|\mathbf{X}}$ , because  $\mathcal{S}_{(Y,\delta)|\mathbf{X}} \subseteq \mathcal{S}_{(T,C)|\mathbf{X}}$ . It should be noted that the estimation of  $\mathcal{S}_{(Y,\delta)|\mathbf{X}}$  is plausible, because  $(Y, \delta, \mathbf{X})$  are collected for survival analysis. Cook (2003) connects the two regressions of  $T|\mathbf{X}$  and  $(Y, \delta)|\mathbf{X}$  under the following condition: (C2)  $C \perp\!\!\!\perp \mathbf{X} | (\boldsymbol{\eta}^T \mathbf{X}, T)$ . Then, condition C2 forces statement (2.1) equivalent to  $(T, C) \perp\!\!\!\perp \mathbf{X} | \boldsymbol{\eta}^T \mathbf{X}$ , that is,  $\mathcal{S}_{(T,C)|\mathbf{X}} = \mathcal{S}_{T|\mathbf{X}}$ . This then also directly implies that  $\mathcal{S}_{(Y,\delta)|\mathbf{X}} \subseteq \mathcal{S}_{(T,C)|\mathbf{X}} = \mathcal{S}_{T|\mathbf{X}}$ . According to Cook (2003), the equality is expected to hold, because proper containment requires carefully balanced conditions. SIR and FSIR are then directly applicable with bivariate slicing of  $Y$  and  $\delta$  to recover  $\mathcal{S}_{T|\mathbf{X}}$ .

### 2.4. Some discussion on bivariate slicing

The bivariate slicing for  $Y$  and  $\delta$  is more systematic than that for  $Y$  alone, which is done as follows. First, slice  $Y$  into  $h$  slices. Each slice is then divided into two groups based on the values of  $\delta$ . So, total numbers of slices are  $h \times 2$ . This slice scheme is quite troublesome in some cases because the slice cannot be divided into two groups if  $\delta$  has only one value of either 0 or 1 in some slices. Chen *et al.* (1999) showed that this may lead biases in the process of the dimension reduction. It then becomes possible that the order of the slices can be changed. That is, slice  $Y$  within each group of  $\delta$ . This slicing scheme subsequently then results in more unbalanced numbers for observations within each of the final slices, which is against the usual policy of the slice scheme that the SIR pursues. It is also not clear which of the two has more advantageous in practice. It is therefore expected that the SIR application to survival regression is more sensitive to the numbers of slices than to usual regressions to slice the response alone. This discussion implies that the FSIR has potential advantages over the SIR in survival regression.

## 3. Numerical studies and real data application

### 3.1. Numerical studies

We consider two types of survival regressions, which are accelerated failure time (AFT) and Cox-proportional Hazards (CPH) models. For both models, the predictors of  $\mathbf{X} = (X_1, \dots, X_{10})^T$  were independently sampled from  $N(0, 1)$ . For  $\boldsymbol{\eta}_0 = (1/\sqrt{55})(1, 2, 3, 4, 5, 0, 0, 0, 0, 0)$ , one-dimensional linear combination  $\boldsymbol{\eta}^T \mathbf{X}$  of  $\mathbf{X}$  was constructed, so that  $\boldsymbol{\eta}$  has the unit length.

With this predictor setting, the following AFT model, shown in Section 5 of Datta *et al.* (2007),

was generated:

$$T|\mathbf{X} \stackrel{\text{iid}}{\sim} \exp(\boldsymbol{\eta}^T \mathbf{X} + \varepsilon),$$

where a random error  $\varepsilon$ , which is independent of  $\mathbf{X}$ , was randomly sampled from  $N(0, 1)$ . The survival regression of  $T|\mathbf{X}$  then follows a log-normal distribution. A censoring variable  $C$  was also randomly generated from a log-normal distribution  $\exp\{N(c_0 \sqrt{2}, 2)\}$  so that  $C \perp\!\!\!\perp \mathbf{X}$ .

For CPH models, first, we followed an example in Section 4.2 of Yoo and Lee (2011). Its hazard and baseline hazard rates are  $\lambda = \exp(\boldsymbol{\eta}^T \mathbf{X})$  and  $\lambda_0(t) = 1$ , respectively, where  $t$  stands for the value of survival time. The second CPH model is constructed by  $\lambda = \exp(\boldsymbol{\eta}^T \mathbf{X})$  and  $\lambda_0(t) = \omega t^{\omega-1}$  for a hazard function and baseline hazard rates, respectively. In the second model, the hazard rate increases with  $\omega > 1$ , and decreases with  $0 < \omega < 1$ . If  $\omega = 1$ , the second CPH model is equivalent to the first one. For the CPH models, a censoring time  $C$  was randomly generated from  $\text{Uniform}(0, c_0)$ , which is independent of  $\mathbf{X}$ . For the second CPH model, two values of 0.5 and 2 were considered for  $\omega$ .

In the AFT and CPH models, the right-censoring scheme was considered, so the observed survival time  $Y$  is equal to  $\min(T, C)$ . The censoring status  $\delta$  is also equal to 1, if  $C > T$ , and 0, otherwise. The sufficient predictor  $\boldsymbol{\eta}^T \mathbf{X}$  is supposed to be estimated better with smaller censoring percentages since censoring variables are generated independently of  $\mathbf{X}$ . Therefore, as discussed in Section 2.3, the censoring percentages affect the asymptotic behaviors of SIR and FISIR in the basis estimation. For this,  $c_0$  was selected to control the average censoring percentages. We considered 30%, 50%, 70% and 80% of censoring. For the second CPH model with  $\omega = 0.5$ , the case of 80% censoring was ruled out, because numerical errors occurred too often.

To measure how well  $\boldsymbol{\eta}$  is estimated, the trace correlation  $r$  (Hooper, 1959) between  $\boldsymbol{\eta}$  and  $\hat{\boldsymbol{\eta}}$  was considered:

$$r = \sqrt{\text{trace}\left(\hat{\boldsymbol{\eta}}\left(\hat{\boldsymbol{\eta}}^T \hat{\boldsymbol{\eta}}\right)^{-1} \hat{\boldsymbol{\eta}}^T \boldsymbol{\eta}\left(\boldsymbol{\eta}^T \boldsymbol{\eta}\right)^{-1} \boldsymbol{\eta}^T\right)}.$$

To convert the correlation (larger, better) to the distance (smaller to better), we define the trace distance of  $r_D = 1 - r$ . With  $n = 100$  and 1,000 iterations of each model, the averages of  $r_D$  are computed and reported in Figures 1–3. In the figure, “F*k.i*” and “S*k.i*” stand for FSIR and SIR, respectively. The script  $i$  in “F*k.i*” and “S*k.i*” indicates the numbers of slices for bivariate slicing of  $Y$  and  $\delta$ . For example, if  $i = 6$ , the number of slices of  $Y$  is equal to 3, because  $\delta$  is categorical with two levels. We considered  $i = 6, 8, 10$  and  $i = 4, 6, 8, 10$  for SIR and FSIR, respectively. For “F*k.i*” and “S*k.i*”,  $k$  varies from 1 to 3 and 1 to 2, respectively. If  $k = 1$ ,  $Y$  is sliced within each category of the censoring status  $\delta$ . For  $k = 2$ ,  $Y$  is sliced first, and then each slice is re-categorized based on the values of  $\delta$ . If  $k = 3$ , the two cases of  $k = 1$  and 2 are fused.

According to Figures 1–3, for the AFT and CPH models, the accuracy in the estimation of  $\boldsymbol{\eta}$  deteriorates as the average censoring percentages increase. Comparing the sensitiveness to the number of slices between FSIR and SIR, SIR becomes more sensitive to the number of slices than FISIR with larger average censoring percentages. FSIR remains quite robust regardless of the censoring percentages. For both SIR and FSIR, the slicing order of  $Y$  and  $\delta$  seem not to matter in the estimation of  $\boldsymbol{\eta}$ . The numerical studies meet our expectations and confirm the potential advantages of FSIR over SIR in survival regression.

### 3.2. Primary biliary cirrhosis data

For illustration purposes, we use primary biliary cirrhosis (PBC) data in Tibshirani (1997) and Yoo and Lee (2011). The data is collected at the Mayo Clinic between 1974 and 1986 and consists of the

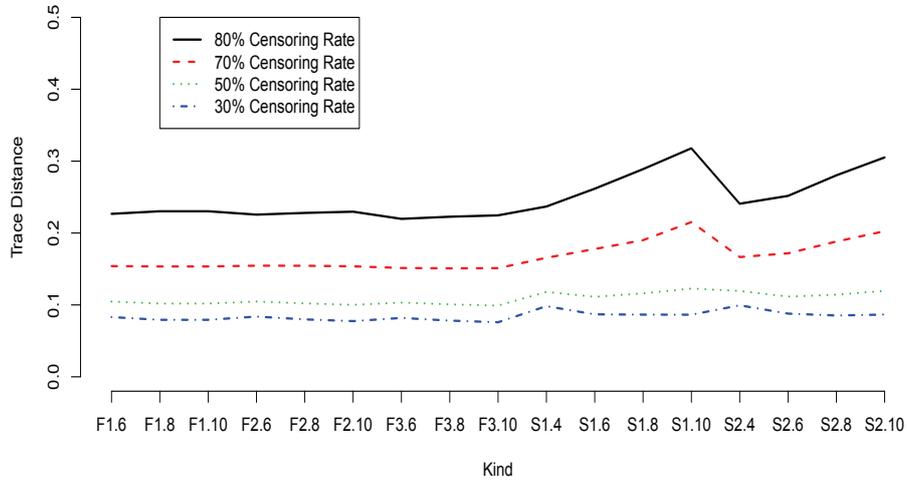


Figure 1: Averages of the trace distances for the accelerated failure time model in Section 3.1.

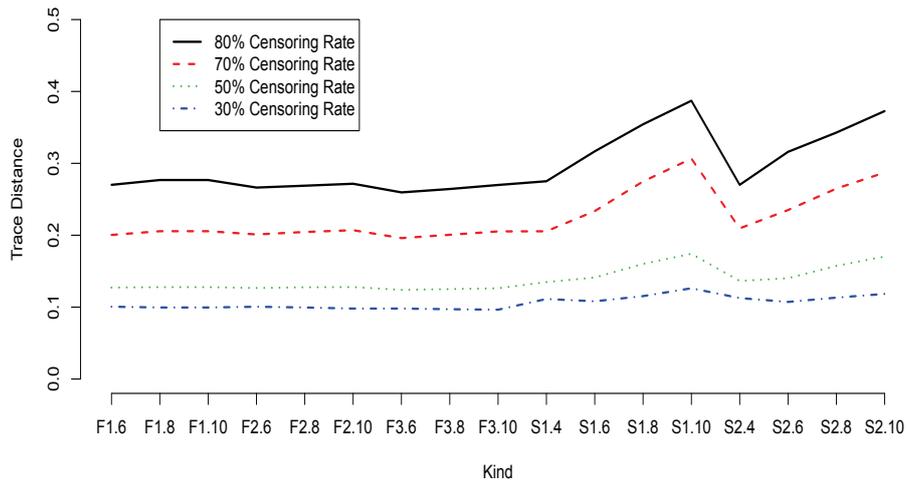
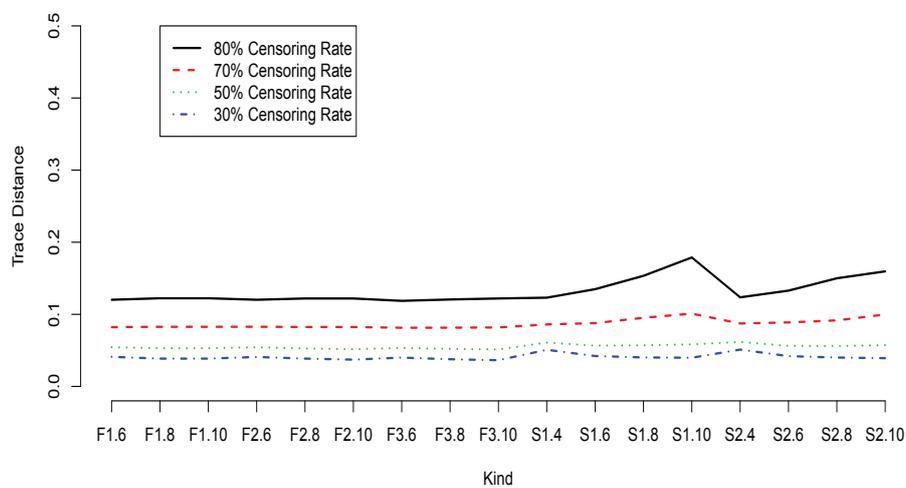


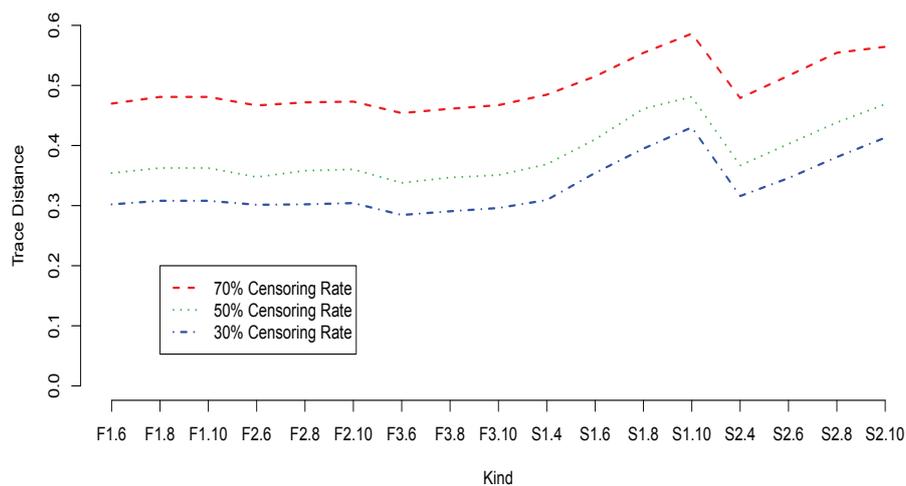
Figure 2: Averages of the trace distances for the Cox-proportional Hazards model with baseline hazards equal to 1 in Section 3.1.

following 19 variables with 276 observations:

- $Y$  = the number of days between registration and the earlier of death or censoring
- $\delta = 1$ , 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



(a)  $\omega = 2$



(b)  $\omega = 0.5$

Figure 3: Averages of the trace distances for the Cox-proportional Hazards model with baseline hazards equal to  $\omega t^{\omega-1}$  in Section 3.1.

- $X_4$  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

Table 1: The average and maximum trace distances for each of FSIR and SIR applications to PBC data in Section 3.2

	Method	$d = 1$	$d = 2$	$d = 3$
Averages	FSIR	0.003	0.013	0.025
	SIR	0.018	0.080	0.093
Maximum	FSIR	0.029 (F1.10 & F2.6)	0.079 (F1.6 & F1.10)	0.101 (F1.10 & F2.6)
	SIR	0.086 (S1.10 & S2.4)	0.263 (S1.4 & S1.10)	0.176 (S1.8 & S2.4)

FSIR = fused sliced inverse regression; SIR = sliced inverse regression.

- $X_8$  Serum bilirubin, in mg/dl
- $X_9$  Serum cholesterol, in mg/dl
- $X_{10}$  Albumin, in g/dl
- $X_{11}$  Urine copper, in  $\mu\text{g/day}$
- $X_{12}$  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

Tibshirani (1997) fitted the data to the CPH model with the 17 predictors, and Yoo and Lee (2011) applied SIR to the PBC data for model-free variable selection.

In the data, the censoring percentage is about 60%, so it is expected that SIR applications with 4, 6, 8, 10 slices yield similar basis estimates. The same slicing schemes as the numerical studies are used for FSIR and SIR. After conducting the various application of FSIR and SIR, the trace distances between  $\hat{\eta}_{Fk,i}^{(m)}$  and  $\hat{\eta}_{Sk,i}^{(m)}$  are computed given  $d = m$  for  $m = 1, 2, 3$ . Since  $d$  is not be determined in this example, the basis are estimated via various choices of  $d$ . The maximum value for  $d$  is three, because  $Sk.4$  can yield only a three-dimensional basis. Table 1 reports the average and maximum distances for each of FSIR and SIR instead of comparing all trace distances. In Table 1,  $(\cdot, \cdot)$  next to the distances stands for the cases to give the maximum distances.

Table 1 also indicates that FSIR is clearly more robust to slicing schemes for any choices of  $d$  than the SIR. Considering the averages, the FSIR and the SIR seem to have no notable differences for the slicing schemes. However, investigating the maximum distance, FSIR produce more robust estimates than SIR. Especially, with  $d = 2$ , the difference between FSIR and SIR in the maximum is 0.16, which seems non-trivial.

#### 4. Conclusion

The goal of the paper is to study FSIR (Cook and Zhang, 2014) in survival regression. Bivariate slicing schemes can affect the estimation of the central subspace by a SIR (Li, 1991) since survival regression considers the observed survival time and censoring status as responses. This issue can be relieved by fusing SIR kernel matrices from the various bivariate slicing schemes. To reach the goal,

the FSIR applies to survival regression by considering various slicing schemes, and is compared with SIR.

Numerical studies confirm that the FSIR has potential advantages in the basis estimation in survival regression over SIR. The real data application also shows that the FSIR results in more robust estimates than SIR.

FSIR cannot have a direct application to correlated or clustered survival data since the SIR is based on iid observations. A study along with this side is in progress.

### Acknowledgments

The authors are grateful to two reviewers and Associate Editor for insightful comments to improve the paper. For the author Jae Keun Yoo, this work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Education (NRF-2014R1A2A1A11049389 and 2009-0093827).

### References

- Chen CH, Li KC, and Wang JL (1999). Dimension reduction for censored regression data, *The Annals of Statistics*, **27**, 1–23.
- Cook RD (2003). Dimension reduction and graphical exploration in regression including survival analysis, *Statistics in Medicine*, **22**, 1399–1413.
- 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.
- Datta S, Le-Rademacher J, and Datta S (2007). Predicting patient survival from microarray data by accelerated failure time modeling using partial least squares and LASSO, *Biometrics*, **63**, 259–271.
- Hooper JW (1959). Simultaneous equations and canonical correlation theory, *Econometrica*, **27**, 245–257.
- Li KC (1991). Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, **86**, 316–327.
- Tibshirani R (1997). The lasso method for variable selection in the Cox model, *Statistics in Medicine*, **16**, 385–395.
- Yoo JK and Lee K (2011). Model-free predictor tests in survival regression through sufficient dimension reduction, *Lifetime Data Analysis*, **17**, 433–444.

Received July 21, 2017; Revised September 4, 2017; Accepted September 14, 2017