

Direction estimation in the single-index model with missing values

YUOXIAO DONG* AND LIPING ZHU

We cast direction estimation in the single-index model into the sufficient dimension reduction framework. Existing sufficient dimension reduction literature with missing values mainly focuses on sliced inverse regression and requires the missing at random (MAR) assumption. In this paper, we propose new methods to handle missing data based on sliced average variance estimation and directional regression. By examining different missingness schemes, we demonstrate that inverse probability weighted estimators for missing predictor are not sensitive to the MAR assumption. The fusion-refined procedures for missing response, on the other hand, may be outperformed by complete case analysis if the response is missing completely at random (MCAR).

AMS 2000 SUBJECT CLASSIFICATIONS: Primary 60K35, 60K35; secondary 60K35.

KEYWORDS AND PHRASES: Directional regression, Missing at random, Missing completely at random, Sliced average variance estimation, Sliced inverse regression.

1. INTRODUCTION

In multivariate analysis with p -dimensional predictor X and univariate response Y , the single-index model assumes that the regression between Y and X can be fully described by a single linear combination $\beta^T X$ of the predictor. Throughout this paper, we assume for some $\beta \in \mathbb{R}^p$,

$$(1) \quad Y = f(\beta^T X, \epsilon),$$

where $\epsilon \perp X$ and “ \perp ” means statistical independence. Model (1) is a compromise between fully parametric and nonparametric modeling. Average derivative estimation (Hardle and Stoker, 1989) considered a special case of (1), and assumed that the conditional mean $E(Y|X)$ has a single index structure. As an incomplete list, model (1) has also been considered in Li and Duan (1989), Li (1991), Yin and Cook (2005).

Estimating the direction β in model (1) can be cast into the framework of sufficient dimension reduction (Cook, 1994, 1996). Sufficient dimension reduction looks for $B \in \mathbb{R}^{p \times d}$ ($d < p$) with the smallest column space, such that $Y \perp X|B^T X$. The column space of B , $\text{Span}(B)$, is called

the central space of Y versus X and denoted by $\mathcal{S}_{Y|X}$. The existence of the central space has been discussed in Yin et al. (2008). The most popular sufficient dimension reduction methods in literature include sliced inverse regression (SIR) (Li, 1991), sliced average variance estimation (SAVE) (Cook and Weisberg, 1991), and directional regression (DR) (Li and Wang, 2007). Model (1) implies that the conditional distribution of $Y|X$ is a function of $\beta^T X$, or equivalently, $Y \perp X|\beta^T X$. Estimating index β in (1) can thus be viewed as estimating the one-dimensional central space $\mathcal{S}_{Y|X} = \text{Span}(\beta)$. In this paper, we aim to recover β in model (1) when there are missing values in the predictor or in the response. The procedures discussed throughout the paper are based on SIR, SAVE and DR.

Missing data is commonplace in multivariate analysis and poses challenges to the existing sufficient dimension reduction methodology. Missing values in the predictors and in the response are treated separately in the literature. Li and Lu (2008) proposed an inverse probability weighted estimator to estimate $\mathcal{S}_{Y|X}$ with the predictors missing at random. In the case when the response is missing, denote Δ as the missingness indicator of Y and $\mathcal{S}_{\Delta|X}$ as the central space of Δ versus X . Ding and Wang (2011) introduced a novel fusion-refined procedure, where estimation of $\mathcal{S}_{Y|X}$ borrows information from $\mathcal{S}_{\Delta|X}$.

The estimators proposed in Li and Lu (2008), Ding and Wang (2011) are both based on SIR. In this paper, we develop the inverse probability weighted procedure and the fusion-refined procedure based on SAVE and DR. SIR is known to be preferable when the link function between Y and $\beta^T X$ is monotone, but SAVE works better when the link function has significant quadratic trend. DR combines the strength of SIR and DR, and enjoys satisfactory performance for a wide range of models. We will see the newly developed estimators based on SAVE and DR inherit such properties when there are missing data. For a complete treatment, we discuss likelihood based estimators in addition to the inverse probability weighted estimators in the presence of missing predictors. For missing response, imputation based estimators as well as fusion-refined procedures are presented.

Missing completely at random (MCAR) and missing at random (MAR) (Rubin, 1976; Little and Rubin, 1987) are two commonly considered missingness schemes. Li and Lu (2008), Ding and Wang (2011) both rely on the MAR

*Corresponding author.

assumption. We compare various methods under different missingness schemes. Because the MAR and MCAR assumptions may be difficult to test in applications, it is important to understand how sensitive the procedures are to the missingness assumptions. Through extensive simulation studies, we find that the inverse probability weighted estimators with missing predictors can work well under either the MAR or the MCAR assumption. On the other hand, although the fusion-refined procedures for missing response can improve over complete analysis under the MAR assumption, this may not be true when the response is MCAR.

The paper is organized as follows. We briefly review the classical SIR, SAVE and DR algorithm without missing values in Section 2. Based on SAVE and DR, we develop new procedures for the missing predictor in Section 3. New fusion-refined and imputation methods for missing response are proposed in Section 4. As we treat missing predictor and missing response separately, simulation studies are provided for each case at the end of Sections 3 and 4 respectively. We conclude the paper with a brief discussion in Section 5.

2. A REVIEW OF SIR, SAVE AND DR

We assume X is multivariate normal with mean 0 and variance Σ throughout this paper. Please note that normality is not the necessary condition for the validity of SIR, SAVE and DR. See, for example, Li and Dong (2009), Dong and Li (2010). Denote $P_\beta = \beta(\beta^T \Sigma \beta)^{-1} \beta^T \Sigma$, and the normality of X implies

$$(2) \quad E(X|\beta^T X) = P_\beta^T X \text{ and } \text{Var}(X|\beta^T X) = \Sigma - P_\beta^T \Sigma P_\beta.$$

Under model (1), SIR is based on the following fact,

$$E(X|Y) = E\{E(X|\beta^T X)|Y\} = P_\beta^T E(X|Y).$$

It follows that $M_{\text{SIR}} = \text{Var}\{E(X|Y)\} \subseteq \Sigma \mathcal{S}_{Y|X}$. For the development of SAVE, we have

$$\text{Var}(X|Y) = E\{\text{Var}(X|\beta^T X)|Y\} + \text{Var}\{E(X|\beta^T X)|Y\}.$$

After plugging in (2) and rearranging the terms, we have

$$\Sigma - \text{Var}(X|Y) = P_\beta^T \{\Sigma - \text{Var}(X|Y)\} P_\beta.$$

It follows that $M_{\text{SAVE}} = E[\{\Sigma - \text{Var}(X|Y)\}^2] \subseteq \Sigma \mathcal{S}_{Y|X}$. DR can be viewed as a combination of SIR and SAVE. For

$$M_{\text{DR}} = E[\{\Sigma - E(XX^T|Y)\}^2] + E^2\{E(X|Y)E(X^T|Y)\} + E\{E(X^T|Y)E(X|Y)\}E\{E(X|Y)E(X^T|Y)\},$$

Li and Wang (2007) demonstrate that $M_{\text{DR}} \subseteq \Sigma \mathcal{S}_{Y|X}$. Denote η as the leading eigenvector of M_{SIR} , M_{SAVE} or M_{DR} . Then $\Sigma^{-1}\eta$ is used to recover the direction β in model (1).

For the sample level estimation, consider an i.i.d. sample (x_i, y_i) . Suppose the sample space of Y can be partitioned into H categories J_1, \dots, J_H , with the h th category having

n_h observations. This partitioning is obvious with discrete Y , and can be achieved by slicing when Y is continuous. Let $I_{i,h} = I(y_i \in J_h)$ denote the indicator function of the i th observation y_i in the h th category. The algorithm for classical SIR, SAVE and DR with no missing values is as follows.

1. Calculate $\hat{\Sigma} = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T / n$, where $\bar{x} = \sum_{i=1}^n x_i / n$.
2. Calculate intraslice mean $\bar{x}_h = \sum_{i=1}^n x_i I_{i,h} / n_h$ and intraslice variance $\hat{\Sigma}_h = \sum_{i=1}^n (x_i - \bar{x}_h)(x_i - \bar{x}_h)^T I_{i,h} / n_h$. Also compute $p_h = n_h / n$, $\nu_h = (\bar{x}_h - \bar{x})(\bar{x}_h - \bar{x})^T$ and $c_h = (\bar{x}_h - \bar{x})^T (\bar{x}_h - \bar{x})$, $h = 1, \dots, H$.
3. For SIR and SAVE, compute $\hat{M}_{\text{SIR}} = \sum_{h=1}^H \nu_h p_h$ and $\hat{M}_{\text{SAVE}} = \sum_{h=1}^H (\hat{\Sigma} - \hat{\Sigma}_h)^2 p_h$ respectively. For DR,

$$\hat{M}_{\text{DR}} = \sum_{h=1}^H (\hat{\Sigma} - \hat{\Sigma}_h - \nu_h)^2 p_h + \left(\sum_{h=1}^H \nu_h p_h \right)^2 + \left(\sum_{h=1}^H c_h p_h \right) \left(\sum_{h=1}^H \nu_h p_h \right).$$

4. Denote the leading eigenvector of \hat{M} (SIR, SAVE, or DR) as $\hat{\eta}$. We estimate β in (1) by $\hat{\beta} = \hat{\Sigma}^{-1} \hat{\eta}$.

When there are missing values in either the response or the predictor, the complete case analysis will discard the cases with missing values, and the above algorithm is implemented based on the cases with complete observations.

3. MISSING PREDICTOR

Li and Lu (2008) proposed inverse probability weighted SIR, and we extend it to SAVE and DR. Besides the inverse probability weighted methods considered in Li and Lu (2008), we present the likelihood based methods as well. For the ease of demonstration, we only discuss the case where one predictor is missing.

3.1 Inverse probability weighted estimators

For $X = (X_1^T, X_2^T)^T$ with $X_1 \in \mathbb{R}^{p-1}$ and $X_2 \in \mathbb{R}$, suppose X_1 is complete, and the missingness of X_2 depends on the observed data but not on the missing data. In symbols, $R \perp\!\!\!\perp X_2 | (Y, X_1)$, where $R = 1$ indicates no missingness and $R = 0$ means at least one entry in X_2 is missing. The challenge here is to recover terms such as Σ , $E(X|Y)$ and $\text{Var}(X|Y)$ in spite of missingness in the data. The key of inverse probability weighted methods is to model the non-missing probability via $\pi = P(R = 1 | Y, X_1)$. Define $\tilde{X}_2 = R X_2 / \pi$ and $\tilde{X}_{22} = R X_2 X_2^T / \pi$. We modify Lemma 1 in Li and Lu (2008) and have the following result.

Proposition 1. *Suppose $\pi = P(R = 1 | Y, X_1)$ is correctly specified, and the MAR assumption $R \perp\!\!\!\perp X_2 | (Y, X_1)$ holds, then $E(\tilde{X}_2) = E(X_2)$, $E(\tilde{X}_{22}) = E(X_2 X_2^T)$, $E(\tilde{X}_2 | Y) = E(X_2 | Y)$, $E(\tilde{X}_{22} | Y) = E(X_2 X_2^T | Y)$, $E(X_1 \tilde{X}_2^T) = E(X_1 X_2^T)$, and $E(X_1 \tilde{X}_2^T | Y) = E(X_1 X_2^T | Y)$.*

The proof of Proposition 1 is straightforward and thus omitted. Please note that Proposition 1 is still valid if we replace the MAR assumption $R \perp\!\!\!\perp X_2 | (Y, X_1)$ with the MCAR assumption $R \perp\!\!\!\perp (Y, X)$, as the latter will guarantee the former.

Proposition 1 suggests we recover Σ and $\text{Var}(X|Y)$ via

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix} \text{ and } \text{Var}(X|Y) = \begin{pmatrix} \Sigma_{11|Y} & \Sigma_{12|Y} \\ \Sigma_{12|Y}^T & \Sigma_{22|Y} \end{pmatrix}$$

where $\Sigma_{11} = \text{Var}(X_1)$, $\Sigma_{12} = E(X_1 \tilde{X}_2^T) - E(X_1)E^T(\tilde{X}_2)$, $\Sigma_{22} = E(\tilde{X}_{22}) - E(\tilde{X}_2)E^T(\tilde{X}_2)$, $\Sigma_{11|Y} = \text{Var}(X_1|Y)$, $\Sigma_{12|Y} = E(X_1 \tilde{X}_2^T|Y) - E(X_1|Y)E^T(\tilde{X}_2|Y)$, and $\Sigma_{22|Y} = E(\tilde{X}_{22}|Y) - E(\tilde{X}_2|Y)E^T(\tilde{X}_2|Y)$.

We describe the sample level inverse probability weighted estimators next. Suppose (x_i, y_i, r_i) are i.i.d. copies of (X, Y, R) . For $i = 1, \dots, n$, $x_i = (x_{1i}^T, x_{2i}^T)^T$, $r_i = 0$ if x_{2i} is missing and $r_i = 1$ otherwise. The inverse probability weights have to be estimated by positing a parametric model $\pi(\theta_\pi) = \pi(Y, X_1; \theta_\pi)$. Based on all (x_{1i}, y_i, r_i) , $\hat{\theta}_\pi$ can be obtained by maximizing the likelihood $\prod_{i=1}^n [\pi(y_i, x_{1i}; \theta_\pi)]^{r_i} [1 - \pi(y_i, x_{1i}; \theta_\pi)]^{1-r_i}$. We follow Li and Lu (2008) and posit a linear regression model $\pi(\theta_\pi)$. Denote the resulting estimators as $\hat{\pi}_i = \pi(y_i, x_{1i}; \hat{\theta}_\pi)$. Then we have

$$\hat{x}_{2i} = r_i x_{2i} / \hat{\pi}_i, \text{ and } \hat{x}_{22i} = r_i x_{2i} x_{2i}^T / \hat{\pi}_i.$$

Although x_{2i} may be missing, \hat{x}_{2i} and \hat{x}_{22i} are available at the sample level. Denote $\bar{x}_1 = \sum_{i=1}^n x_{1i} / n$, $\bar{x}_2 = \sum_{i=1}^n \hat{x}_{2i} / n$, and $\bar{x}_{22} = \sum_{i=1}^n \hat{x}_{22i} / n$. The sample estimators of Σ is thus

$$\hat{\Sigma}^{\text{ipw}} = \begin{pmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{12}^T & \hat{\Sigma}_{22} \end{pmatrix},$$

$$\text{where } \hat{\Sigma}_{11} = \frac{1}{n} \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{1i} - \bar{x}_1)^T,$$

$$\hat{\Sigma}_{12} = \frac{1}{n} \sum_{i=1}^n x_{1i} \hat{x}_{2i}^T - \bar{x}_1 \bar{x}_2^T, \text{ and } \hat{\Sigma}_{22} = \bar{x}_{22} - \bar{x}_2 \bar{x}_2^T.$$

Recall that J_1, \dots, J_H is a partition of the sample space of Y , $I_{i,h} = 1$ if $y_i \in J_h$, and $I_{i,h} = 0$ otherwise. Let $\bar{x}_{1,h} = \sum_{i=1}^n x_{1i} I_{i,h} / n_h$, $\bar{x}_{2,h} = \sum_{i=1}^n \hat{x}_{2i} I_{i,h} / n_h$, and $\bar{x}_{22,h} = \sum_{i=1}^n \hat{x}_{22i} I_{i,h} / n_h$. Then we estimate $\text{Var}(X|Y \in J_h)$ by

$$\hat{\Sigma}_h^{\text{ipw}} = \begin{pmatrix} \hat{\Sigma}_{11,h} & \hat{\Sigma}_{12,h} \\ \hat{\Sigma}_{12,h}^T & \hat{\Sigma}_{22,h} \end{pmatrix},$$

$$\text{where } \hat{\Sigma}_{11,h} = \frac{1}{n_h} \sum_{i=1}^n (x_{1i} - \bar{x}_{1,h})(x_{1i} - \bar{x}_{1,h})^T I_{i,h},$$

$$\hat{\Sigma}_{12,h} = \frac{1}{n_h} \sum_{i=1}^n x_{1i} \hat{x}_{2i}^T I_{i,h} - \bar{x}_{1,h} \bar{x}_{2,h}^T, \text{ and}$$

$$\hat{\Sigma}_{22,h} = \bar{x}_{22,h} - \bar{x}_{2,h} \bar{x}_{2,h}^T.$$

Furthermore, denote $\bar{x}^{\text{ipw}} = (\bar{x}_1^T, \bar{x}_2)^T$, $\bar{x}_h^{\text{ipw}} = (\bar{x}_{1,h}^T, \bar{x}_{2,h})^T$, $\nu_h^{\text{ipw}} = (\bar{x}_h^{\text{ipw}} - \bar{x}^{\text{ipw}})(\bar{x}_h^{\text{ipw}} - \bar{x}^{\text{ipw}})^T$, and $c_h^{\text{ipw}} = (\bar{x}_h^{\text{ipw}} - \bar{x}^{\text{ipw}})^T (\bar{x}_h^{\text{ipw}} - \bar{x}^{\text{ipw}})$. The sample estimators of M_{SIR} and M_{SAVE} become

$$\hat{M}_{\text{SIR}}^{\text{ipw}} = \sum_{h=1}^H \nu_h^{\text{ipw}} p_h, \text{ and } \hat{M}_{\text{SAVE}}^{\text{ipw}} = \sum_{h=1}^H (\hat{\Sigma}_h^{\text{ipw}} - \hat{\Sigma}_h^{\text{ipw}})^2 p_h.$$

Similarly, we have the sample estimator of M_{DR} as

$$\begin{aligned} \hat{M}_{\text{DR}}^{\text{ipw}} &= \sum_{h=1}^H (\hat{\Sigma}_h^{\text{ipw}} - \hat{\Sigma}_h^{\text{ipw}} - \nu_h^{\text{ipw}})^2 p_h \\ &+ \left(\sum_{h=1}^H \nu_h^{\text{ipw}} p_h \right)^2 + \left(\sum_{h=1}^H c_h^{\text{ipw}} p_h \right) \left(\sum_{h=1}^H \nu_h^{\text{ipw}} p_h \right). \end{aligned}$$

Denote the leading eigenvector of \hat{M}^{ipw} (SIR, SAVE, or DR) as $\hat{\eta}^{\text{ipw}}$, and the final inverse probability weighted estimator is $\hat{\beta}^{\text{ipw}} = (\hat{\Sigma}^{\text{ipw}})^{-1} \hat{\eta}^{\text{ipw}}$.

3.2 Likelihood based estimators

For multivariate normal with missing observations, likelihood based estimators for the mean vector and the covariance matrix are well-established in literature. See, for example, Little and Rubin (1987). SIR, SAVE, and DR rely on estimation of intraslice mean vectors and intraslice covariance matrices, and we modify the algorithm in Section 2 as follows:

- 1'. Calculate $\hat{\Sigma}^{\text{mle}}$ and \bar{x}^{mle} using the *mlest* function in the R package 'mvnmle'.
- 2'. Based on the x_i values in the h th slice, use *mlest* and calculate the intraslice mean \bar{x}_h^{mle} and the intraslice variance $\hat{\Sigma}_h^{\text{mle}}$. Also compute $p_h = n_h/n$, $\nu_h^{\text{mle}} = (\bar{x}_h^{\text{mle}} - \bar{x}^{\text{mle}})(\bar{x}_h^{\text{mle}} - \bar{x}^{\text{mle}})^T$ and $c_h^{\text{mle}} = (\bar{x}_h^{\text{mle}} - \bar{x}^{\text{mle}})^T (\bar{x}_h^{\text{mle}} - \bar{x}^{\text{mle}})$, $h = 1, \dots, H$.
- 3'. For SIR and SAVE, we have $\hat{M}_{\text{SIR}}^{\text{mle}} = \sum_{h=1}^H \nu_h^{\text{mle}} p_h$ and $\hat{M}_{\text{SAVE}}^{\text{mle}} = \sum_{h=1}^H (\hat{\Sigma}_h^{\text{mle}} - \hat{\Sigma}_h^{\text{mle}})^2 p_h$. For DR,

$$\begin{aligned} \hat{M}_{\text{DR}}^{\text{mle}} &= \sum_{h=1}^H (\hat{\Sigma}_h^{\text{mle}} - \hat{\Sigma}_h^{\text{mle}} - \nu_h^{\text{mle}})^2 p_h \\ &+ \left(\sum_{h=1}^H \nu_h^{\text{mle}} p_h \right)^2 + \left(\sum_{h=1}^H c_h^{\text{mle}} p_h \right) \left(\sum_{h=1}^H \nu_h^{\text{mle}} p_h \right). \end{aligned}$$

- 4'. Denote the leading eigenvector of \hat{M}^{mle} (SIR, SAVE, or DR) as $\hat{\eta}^{\text{mle}}$. We estimate β in (1) by $\hat{\beta}^{\text{mle}} = (\hat{\Sigma}^{\text{mle}})^{-1} \hat{\eta}^{\text{mle}}$.

3.3 Simulation with missing predictor

We carry out numerical studies in this section to compare different estimators in the presence of missing predictors. Suppose $X = (V_1, \dots, V_5)^T$ is multivariate normal with

Table 1. Missing predictor. The median absolute sample correlation between $\beta^T X$ and $\hat{\beta}^T X$ is reported based on 100 repetitions and sample size $n = 200$

Model	Missing Proportion	Method	$\hat{\beta}^{\text{full}}$	$\hat{\beta}^{\text{cc}}$	V_1 MCAR		V_1 MAR		
					$\hat{\beta}^{\text{mle}}$	$\hat{\beta}^{\text{ipw}}$	$\hat{\beta}^{\text{cc}}$	$\hat{\beta}^{\text{mle}}$	$\hat{\beta}^{\text{ipw}}$
I	50%	SIR	.981	.954	.977	.978	.953	.969	.971
		SAVE	.939	.794	.898	.900	.825	.897	.788
		DR	.973	.914	.957	.949	.928	.957	.932
	70%	SIR	.978	.927	.959	.966	.925	.943	.962
		SAVE	.953	.781	.792	.779	.796	.844	.727
		DR	.969	.891	.939	.915	.896	.920	.832
II	50%	SIR	.377	.352	.418	.503	.386	.489	.502
		SAVE	.986	.968	.975	.975	.968	.975	.968
		DR	.986	.968	.977	.976	.968	.976	.966
	70%	SIR	.335	.308	.447	.480	.348	.512	.492
		SAVE	.987	.949	.954	.966	.949	.951	.937
		DR	.987	.949	.955	.965	.949	.953	.935

mean zero and identity covariance matrix I_5 . The error ϵ is standard normal and independent of X . The two models we consider are

Model I: $Y = \exp(\beta^T X) + \epsilon$, where $\beta = (.5, -.5, .5, 0, 0)^T$;

Model II: $Y = (\beta^T X)^2 + .5\epsilon$, where $\beta = (.5, .5, 0, 0, 0)^T$.

Suppose only V_1 has missing observations with missingness indicator R . We have $R = 1$ when V_1 is observed and $R = 0$ otherwise. Consider two missingness schemes:

$$\begin{aligned}
 (3) \quad & V_1 \text{ MCAR: } P(R = 1) = \rho_x^R (1 - \rho_x)^{1-R}; \\
 & V_1 \text{ MAR: } \pi = P(R = 1 | Y, V_2, \dots, V_5) \\
 & \quad = \frac{\exp(c_x + 0.5V_4 + V_5)}{1 + \exp(c_x + 0.5V_4 + V_5)}.
 \end{aligned}$$

The missing proportion of the predictor V_1 can be controlled by adjusting ρ_x and c_x above. We get approximately 50% missing proportion with $\rho_x = .5$ and $c_x = 0$; and roughly 70% of V_1 is missing with $\rho_x = .3$ and $c_x = -1$. Various estimators are compared across these two different missingness schemes. The estimator based on all observations as if there is no missingness is denoted as $\hat{\beta}^{\text{full}}$; $\hat{\beta}^{\text{cc}}$ denotes the complete case estimator; $\hat{\beta}^{\text{ipw}}$ is the inverse probability weighted estimator described in Section 3.1; and $\hat{\beta}^{\text{mle}}$ denotes the likelihood based estimator in Section 3.2. Three types of estimators (SIR, SAVE and DR) are compared across two missingness schemes (MCAR and MAR). Sample size is fixed at $n = 200$. Based on 100 repetitions, we report in Table 1 the median of the absolute sample correlations between $\beta^T X$ and $\hat{\beta}^T X$.

We first compare different columns in Table 1. $\hat{\beta}^{\text{full}}$ pretends there is no missing data and sets the benchmark. Both $\hat{\beta}^{\text{ipw}}$ and $\hat{\beta}^{\text{mle}}$ have better overall performances than the complete case estimator $\hat{\beta}^{\text{cc}}$. As we discussed in Section 3.1, $\hat{\beta}^{\text{ipw}}$ is not sensitive to the MAR assumption, and has decent performance with V_1 MCAR.

Next we compare SIR, SAVE, and DR across the rows of Table 1. SIR enjoys the best performance for Model I as the link function is monotonic, while it will fail for Model II with quadratic link function. SAVE performs worse than SIR for Model I, but is much better for Model II. As a combination method of SIR and SAVE, DR shares the strength of both methods and has the best overall performance over Models I and II.

4. MISSING RESPONSE

Now we focus on the case when the predictor X is fully observed and the response Y is missing. Ding and Wang (2011) considered fusion-refined and imputation-based estimators for SIR. Their methods can easily extend to SAVE and DR.

4.1 Fusion-refined estimators

Let Δ be the missingness indicator of Y , with $\Delta = 1$ meaning Y is observed and $\Delta = 0$ otherwise. Assume X is fully observed, and Y is MAR with $Y \perp\!\!\!\perp \Delta | X$. Ding and Wang (2011) carefully studied the relationship between the joint central space $\mathcal{S}_{(Y,\Delta)|X}$, and two marginal central spaces $\mathcal{S}_{Y^*|X}$ and $\mathcal{S}_{Y|X}$, where $Y^* = Y\Delta$ is observable. We state the following result that connects these spaces.

Proposition 2. *Suppose there exists $\beta \in \mathbb{R}^p$ such that $Y \perp\!\!\!\perp X | \beta^T X$ and $\Delta \perp\!\!\!\perp X | \beta^T X$. Then under the MAR assumption $\Delta \perp\!\!\!\perp Y | X$, we have*

$$\mathcal{S}_{(Y,\Delta)|X} = \mathcal{S}_{Y^*|X} = \mathcal{S}_{Y|X} = \mathcal{S}_{\Delta|X}.$$

Proposition 2 can be viewed as an extension of Theorem 1 in Ding and Wang (2011). Its proof is obvious and thus omitted. Please note that Proposition 2 is still valid if we replace the MAR assumption $\Delta \perp\!\!\!\perp Y | X$ with the MCAR assumption $\Delta \perp\!\!\!\perp (Y, X)$, as the latter will guarantee the former. While we assume $\mathcal{S}_{\Delta|X}$ aligns with $\mathcal{S}_{Y|X}$ through the same index β ,

Table 2. Missing response. The median absolute sample correlation between $\beta^T X$ and $\hat{\beta}^T X$ is reported based on 100 repetitions and sample size $n = 200$

Model	Missing Proportion	Method	β^{full}			Y MCAR			Y MAR	
			$\hat{\beta}^{\text{full}}$	$\hat{\beta}^{\text{cc}}$	$\hat{\beta}^{\text{mri}}$	$\hat{\beta}^{\text{fr}}$	$\hat{\beta}^{\text{cc}}$	$\hat{\beta}^{\text{mri}}$	$\hat{\beta}^{\text{fr}}$	
I	50%	SIR	.978	.953	.942	.945	.973	.976	.982	
		SAVE	.959	.841	.644	.624	.815	.913	.919	
		DR	.974	.928	.771	.764	.943	.965	.968	
	70%	SIR	.976	.920	.911	.916	.959	.966	.970	
		SAVE	.951	.771	.516	.543	.738	.803	.793	
		DR	.971	.898	.560	.550	.901	.943	.940	
II	50%	SIR	.347	.335	.312	.385	.796	.659	.946	
		SAVE	.986	.970	.959	.967	.970	.946	.953	
		DR	.986	.971	.965	.969	.970	.962	.967	
	70%	SIR	.392	.340	.315	.297	.775	.586	.934	
		SAVE	.987	.938	.906	.901	.893	.868	.849	
		DR	.987	.941	.915	.913	.904	.933	.942	

Ding and Wang (2011) provided a more general treatment where $\mathcal{S}_{\Delta|X}$ may be different from $\mathcal{S}_{Y|X}$. Without ambiguity, we refer to estimators based on $\mathcal{S}_{Y^*|X}$ the fusion-refined estimators.

At the sample level, denote (x_i, y_i, δ_i) as an i.i.d. sample. For $i = 1, \dots, n$, $\delta_i = 0$ means y_i is missing and $\delta_i = 1$ otherwise. Based on fully observable x_i and $y_i^* = y_i \delta_i$, we can carry out the algorithm in Section 2 and estimate $\mathcal{S}_{Y^*|X}$. The final estimator via SIR, SAVE, or DR is denoted as $\hat{\beta}^{\text{fr}}$.

For Y MCAR, it is obvious that $\mathcal{S}_{Y|X}^{\Delta=1} = \mathcal{S}_{Y|X}$, where $\mathcal{S}_{Y|X}^{\Delta=1}$ denotes the central space of Y versus X based on the fully observed data. Under the assumption that Y is MAR, Theorem 5 of Ding and Wang (2011) states that $\mathcal{S}_{Y|X}^{\Delta=1}$ coincides $\mathcal{S}_{Y|X}$. Thus complete case analysis should yield satisfactory results regardless of the missingness scheme.

4.2 Mean regression imputation estimators

Mean regression imputation has been suggested as an alternative to the fusion-refined procedure in Ding and Wang (2011). Let $\hat{z}_i = x_i^T \hat{\beta}^{\text{fr}}$. For subject i with missing response, the key idea is to impute y_i based on $(\hat{z}_j, y_j, \delta_j)$, $j = 1, \dots, n$. Suppose $K(\cdot)$ is a kernel function and denote $K_b(u) = b^{-1}K(u/b)$. We obtain

$$(4) \quad \hat{p}_{ij} = \frac{\delta_j K_b(\hat{z}_j - \hat{z}_i)}{\sum_{j=1}^n \delta_j K_b(\hat{z}_j - \hat{z}_i)},$$

which can be viewed as an estimate of $p_{ij} = P(Y = y_j | X^T \hat{\beta}^{\text{fr}} = x_i^T \hat{\beta}^{\text{fr}})$, the probability of imputing y_j for the i th observation. The mean regression imputation response for the i th observation is thus $y_i^{\text{mri}} = \sum_{j=1}^n y_j \hat{p}_{ij}$. Replace $I_{i,h}$ with $I_{i,h}^{\text{mri}}$ for the algorithm described in Section 2, where $I_{i,h}^{\text{mri}}$ denotes the indicator function of $y_i^{\text{mri}} \in J_h$. The corresponding estimator of β is the mean regression imputation estimator, and we denote it by $\hat{\beta}^{\text{mri}}$.

4.3 Simulation with missing response

We revisit Models I and II in Section 3.3 and consider the case when Y is missing. Denote Δ as the missingness indicator of Y, with $\Delta = 1$ meaning Y observed and $\Delta = 0$ otherwise. Consider two missingness schemes:

$$(5) \quad \begin{aligned} \text{Y MCAR: } P(\Delta = 1) &= \rho_y^\Delta (1 - \rho_y)^{1-\Delta}; \\ \text{Y MAR: } P(\Delta = 1|X) &= \frac{\exp(c_y + \beta^T X)}{1 + \exp(c_y + \beta^T X)}. \end{aligned}$$

The missing proportion of the response Y can be controlled by adjusting ρ_y and c_y above. Setting $\rho_y = .5$ and $c_y = 0$ corresponds to approximately 50% missing proportion; and we get around 70% missing response with $\rho_y = .3$ and $c_y = -1$. Please note that according to Proposition 2, we set the missingness index β to be the same as in the model statement. Various estimators are compared across these two different missingness schemes. As before, we denote $\hat{\beta}^{\text{full}}$ as the estimator based on full data, and $\hat{\beta}^{\text{cc}}$ denotes the complete case estimator. The fusion-refined estimator in Section 4.1 is denoted by $\hat{\beta}^{\text{fr}}$, and the mean regression imputation estimator in Section 4.2 is denoted by $\hat{\beta}^{\text{mri}}$. For $K_b(\cdot)$ involved in (4), we use Gaussian kernel and set the window width $b = n^{-1/5}$ following Ding and Wang (2011).

In Table 2, we report the median absolute sample correlations between $\beta^T X$ and $\hat{\beta}^T X$ based on 100 repetitions. $\hat{\beta}^{\text{full}}$ is based on the full data and sets the benchmark. We clearly see that SIR and DR are preferable for Model I, while SAVE and DR are better than SIR for Model II. The fusion-refined estimator enjoys decent performance under the MAR assumption. With 50% missing proportion for SIR in Model I, $\hat{\beta}^{\text{fr}}$ even yields better result than $\hat{\beta}^{\text{full}}$ (.982 versus .978). This scenario has been described in Ding and Wang (2011), which is possible because $\mathcal{S}_{\Delta|X}$ aligns with $\mathcal{S}_{Y|X}$ and the missingness carries useful information for estimating β . The overall performance of the mean regression imputation estimator $\hat{\beta}^{\text{mri}}$ is similar to $\hat{\beta}^{\text{fr}}$.

ACKNOWLEDGEMENT

We sincerely thank three anonymous referees for their helpful comments that led to a much-improved paper. The research of the first author is supported by National Science Foundation Grant DMS-1106577. The research of the second author is supported by National Natural Science Foundation of China 11071077, Innovation Program of Shanghai Municipal Education Commission 13ZZ055, Pujiang Project of Science and Technology Commission of Shanghai Municipality 12PJ1403200, and Program for New Century Excellent Talents, Ministry of Education of China NCET-12-0901.

Received 2 November 2012

REFERENCES

- COOK, R. D. (1994) On the interpretation of regression plots. *Journal of the American Statistical Association* **89**, 177–189. [MR1266295](#)
- COOK, R. D. (1996) Graphics for regressions with a binary response. *Journal of the American Statistical Association* **91**, 983–992. [MR1424601](#)
- COOK, R. D. & WEISBERG, S. (1991) Discussion of “Sliced inverse regression for dimension reduction”. *Journal of the American Statistical Association* **86**, 316–342. [MR1137117](#)
- DING, X. & WANG, Q. (2011) Fusion-refinement procedure for dimension reduction with missing response at random. *Journal of the American Statistical Association* **106**, 1193–1207. [MR2894774](#)
- DONG, Y. & LI, B. (2010) Dimension reduction for non-elliptically distributed predictors: Second order methods. *Biometrika* **97**(2), 279–294. [MR2650738](#)
- DONG, Y. & ZHU, L. (2012) A note on sliced inverse regression with missing predictors. *Statistical Analysis and Data Mining* **5**(2), 128–138. [MR2910023](#)
- HARDLE, W. & STOKER, T. M. (1989) Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association* **84**, 986–995. [MR1134488](#)
- LI, B. & DONG, Y. (2009) Dimension reduction for non-elliptically distributed predictors. *Annals of Statistics* **37**, 1272–1298. [MR2509074](#)
- LI, B. & WANG, S. (2007) On directional regression for dimension reduction. *Journal of the American Statistical Association* **102**, 997–1008. [MR2354409](#)
- LI, K. C. (1991) Sliced inverse regression for dimension reduction (with discussion). *Journal of the American Statistical Association* **86**, 316–342. [MR1137117](#)
- LI, K. C. & DUAN, N. (1989) Regression analysis under link violation. *Annals of Statistics* **17**, 1009–1052. [MR1015136](#)
- LI, L. & LU, W. (2008) Sufficient dimension reduction with missing predictors. *Journal of the American Statistical Association* **103**, 822–831. [MR2524333](#)
- LITTLE, R. & RUBIN, D. B. (1987) *Statistical analysis with missing data*. New York: Wiley. [MR0890519](#)
- RUBIN, D. B. (1976) Inference and missing data. *Biometrika* **63**, 581–592. [MR0455196](#)
- YIN, X. & COOK, R. D. (2005) Direction estimation in single-index regressions. *Biometrika* **92**(2), 371–384. [MR2201365](#)
- YIN, X., LI, B. & COOK, R. D. (2008) Successive direction extraction for estimating the central subspace in a multiple-index regression. *Journal of Multivariate Analysis* **99**, 1733–1757. [MR2444817](#)

Yuexiao Dong
Temple University
Philadelphia, PA, 19122
USA
E-mail address: ydong@temple.edu

The complete case estimator $\hat{\beta}^{cc}$ under either MCAR and MAR seems to have consistent results with the full data when the missing proportion is 50%, and have reasonable performances when the missing proportion is 70%. This is to be expected as the complete case analysis is unbiased with missing response MCAR or MAR. The fusion-refined $\hat{\beta}^{fr}$ can improve over $\hat{\beta}^{cc}$ under the MAR assumption, as $\hat{\beta}^{fr}$ captures useful information to estimate β from the missingness scheme. On the other hand, $\hat{\beta}^{fr}$ is dominated by $\hat{\beta}^{cc}$ under the MCAR assumption. This is not really surprising, as the only difference between $\hat{\beta}^{cc}$ and $\hat{\beta}^{fr}$ is that the latter includes one additional slice containing all the observations with missing response. When the response is MCAR, the missingness carries no useful information about β and will dilute the estimation of $\mathcal{S}_{Y|X}$.

An interesting finding in Table 2 is the difference between SIR estimator $\hat{\beta}^{cc}$ for Model II. The performance of $\hat{\beta}^{cc}$ based on SIR improves a lot from MCAR to MAR (.335 to .796 with 50% missing; .340 to .775 with 70% missing). SIR based on full data does not work well for Model II due to the symmetric link function, and the estimator with response MCAR will inherit this limitation. When the response is MAR, the observed response is no longer symmetric due to the asymmetry in missingness probability, which explains the improvement of SIR.

5. CONCLUSION

We provide a general treatment of direction estimation in single-index model with missing values. Existing estimators based on SIR, such as inverse probability weighted and fusion-refined estimators, are extended to estimators based on SAVE and DR. We recommend DR-based procedures, which enjoy the best overall performance in a wide range of models, and provide a balance between SIR and SAVE. By considering MAR and MCAR as different missingness schemes, we demonstrate that while inverse probability weighted estimators are not sensitive to the predictor MAR assumption, the fusion-refined procedures could be dominated by complete case analysis when the response is MCAR.

The methods proposed in this paper can be further improved along the following lines. The inverse probability weighted procedures rely on estimation of the missingness probability π , which is estimated in a parametric fashion in this paper. To guard against misspecification of π , augmented inverse probability weighted estimators can be developed as in Li and Lu (2008), Dong and Zhu (2012). The missing predictor and missing response are treated separately, and we only consider the case with a single missing predictor. Further investigation is warranted for multiple missing predictors together with missing response under the sufficient dimension reduction framework.

Liping Zhu
School of Statistics and Management
Shanghai University of Finance and Economics
and the Key Laboratory of Mathematical Economics
Ministry of Education
Shanghai, 200433
China
E-mail address: zhu.liping@mail.shufe.edu.cn