Statistica Sinica **22** (2012), 1611-1637
doi:http://dx.doi.org/10.5705/ss.2009.191

SUFFICIENT DIMENSION REDUCTION IN REGRESSIONS WITH MISSING PREDICTORS

Liping Zhu[1], Tao Wang[2] and Lixing Zhu[2]

[1]*Shanghai University of Finance and Economics*
and [2]*Hong Kong Baptist University*

Abstract: Existing sufficient dimension reduction methods often resort to the complete-case analysis when the predictors are subject to missingness. The complete-case analysis is inefficient even with the *missing completely at random* mechanism because all incomplete cases are discarded. In this paper, we introduce a nonparametric imputation procedure for semiparametric regressions with missing predictors. We establish the consistency of the nonparametric imputation under the *missing at random* mechanism that allows the missingness to depend exclusively upon the completely observed response. When the missingness depends on both the completely observed predictors and the response, we propose a parametric method to impute the missing predictors. We demonstrate the estimation consistency of the parametric imputation method through several synthetic examples. Our proposals are illustrated through comprehensive simulations and a data application.

Key words and phrases: Central subspace, missing at random, missing predictors, nonparametric imputation, sliced inverse regression, sufficient dimension reduction.

1. Introduction

The rapid advance of technology has allowed scientists to collect data of unprecedented dimensionality and complexity. The large dimensionality of these data poses many challenges for statisticians in the study of the relationship among various types of variables. In this paper we study the regression of a univariate response variable Y onto the predictor vector $\mathbf{x} = (X_1, \ldots, X_p)^{\mathrm{T}}$. When the dimension p is large, the conventional methods of parametric modeling usually break down due to the *curse of dimensionality.* As a result, sufficient dimension reduction (SDR, Cook (1998)) has attracted considerable attention in the last two decades. It reduces the dimension effectively without loss of regression information, and it imposes no parametric structures on the regression functions. In general, SDR seeks a subspace $\mathcal{S}$ of minimal dimension such that $Y \perp\!\!\!\perp \mathbf{x} \mid P_{\mathcal{S}}\mathbf{x}$, where $\perp\!\!\!\perp$ stands for statistical independence and $P_{\mathcal{S}}$ denotes the orthogonal projection onto $\mathcal{S}$ with respect to the usual inner product. If such a subspace exists, we call it the *central subspace* (CS), denoted by $\mathcal{S}_{Y|\mathbf{x}}$, and call its dimension

$K = \dim(\mathcal{S}_{Y|\mathbf{x}})$ the *structural dimension.* The CS, which can be viewed as a parsimonious population parameter that captures all the regression information, is thus the main object of interest in the SDR inquiry. Since the seminal work of sliced inverse regression (SIR, Li (1991)), numerous methods in the SDR context have been proposed to recover $\mathcal{S}_{Y|\mathbf{x}}$. See, for example, sliced average variance estimation (SAVE, Cook and Weisberg (1991)), principal Hessian directions (PHD, Li (1992)), minimum average variance estimation and its variants (Xia et al. (2002), and Xia (2007)), contour regression (Li, Zha, and Chiaromonte (2005)), and directional regression (Li and Wang (2007)). Among these, SIR is perhaps the most commonly used in the literature, and there have been many elaborations on the original methodology of SIR.

When the dimension p is large, it is common that some predictors are subject to missingness. Missingness complicates the interpretation of high-dimensional data, or even results in spurious or weakened results obtained from usual statistical analysis. To address this issue, many efforts have been devoted within the framework of parametric regressions. For example, with the *missing at random* (MAR) assumption under the forward regression setup, Robins, Rotnitzky, and Zhao (1994) introduced an augmented inverse probability weighted estimation; Yates (1933) considered an imputation procedure for linear models; Rubin (1987) proposed a multiple imputation procedure for parametric regressions. The efficacy of imputation with parametric model assumptions usually relies on correctness of the specification of the underlying model; if the underlying model is misspecified, the resulting estimation could be biased. To get around this issue, Cheng (1994) and Wang and Rao (2002) suggested imputing the missing values with nonparametric estimates of the conditional mean values through multivariate kernel regressions. Owing to the *curse of dimensionality*, their procedures are not applicable to semiparametric regressions with high-dimensional predictors.

Existing SDR methods often resort to complete-case analysis when the predictors are subject to missingness. This is undesirable since the resulting estimators are inconsistent unless the *missing completely at random* (MCAR) assumption holds true (Wang and Chen (2009)). Besides, inference built on the complete measurements is generally inefficient. To retrieve information contained in the missing predictors, Li and Lu (2008) proposed an augmented inverse probability weighted SIR estimator. However, they imposed several parametric model assumptions on the missingness indicator $\boldsymbol{\delta}$ given Y and the completely observed predictors $\mathbf{x}_{obs}$, thus the validity of the inverse probability weighted estimator relies on the correct specification of the parametric model of $\boldsymbol{\delta} \mid (Y, \mathbf{x}_{obs})$; if the underlying model is misspecified, the estimator may be biased (Li and Lu (2008, p.824)). It is desirable to develop some alternatives for estimation and statistical inference in the SDR context.

In this paper we consider missingness in semiparametric regressions under the MAR mechanism that allows the missingness to depend on the completely observed variables. To avoid the *curse of dimensionality* and to simultaneously retain the regression information, we focus on the widely used SIR method proposed by Li (1991), but emphasize that the general idea of this paper can be readily applied to other SDR methods, such as SAVE, PHD and DR.

Estimating the SIR matrix amounts to estimating the conditional expectations $E(X_k \mid Y)$ and $E(X_k X_l \mid Y)$, for $1 \leq k, l \leq p$. The unconditional expectations can be estimated through $E(X_k) = E\{E(X_k \mid Y)\}$ and $E(X_k X_l) = E\{E(X_k X_l \mid Y)\}$. When the missingness depends exclusively on the response, we propose a nonparametric imputation procedure for the missing predictors (Little and Rubin (2002)). To be precise, let $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^{\mathrm{T}}$ denote a vector of missing indicators. The k-th component δ_k takes value 1 if there is no missingness in X_k, and is 0 otherwise. To estimate $E(X_k \mid Y)$ when X_k has missing values, we impute the missing values with $E(\delta_k X_k \mid Y)/E(\delta_k \mid Y)$. Similarly, to estimate $E(X_k X_l \mid Y)$ when either X_k or X_l is missing, we impute the missing values with $E(\delta_k \delta_l X_k X_l \mid Y)/E(\delta_k \delta_l \mid Y)$. The quantities $E(\delta_k \mid Y)$, $E(\delta_k \delta_l \mid Y)$, $E(\delta_k X_k \mid Y)$ and $E(\delta_k \delta_l X_k X_l \mid Y)$ can be estimated using standard nonparametric regressions because δ_k, $\delta_k \delta_l$, $\delta_k X_k$, $\delta_k \delta_l X_k X_l$ and Y are observed. This nonparametric imputation method is justified under the *missing at random* assumption $\boldsymbol{\delta} \perp\!\!\!\perp \mathbf{x} \mid Y$. It has at least three merits.

1. The nonparametric imputation method is more efficient than the complete-case analysis. Our empirical studies also find that it is more efficient than the augmented inverse probability weighted SIR method proposed by Li and Lu (2008), which echoes the theoretical investigation of Rubin (1987).
2. The nonparametric imputation procedure imputes the missing predictors in a model-free fashion. Unlike the augmented inverse probability weighted SIR estimation proposed by Li and Lu (2008), our nonparametric imputation method retains the flavor of SIR (Li (1991)) in the sense that it avoids the parametric model assumptions imposed by Li and Lu (2008).
3. Owing to the merit of SIR, the nonparametric imputation method works for a wide range of semi-parametric regressions satisfying $Y \perp\!\!\!\perp \mathbf{x} \mid P_{\mathcal{S}_{Y|\mathbf{x}}} \mathbf{x}$. In this sense, we extend the application of nonparametric imputation (Little (1992)) to a very general family of semiparametric regressions.

When the missingness depends on both the completely observed predictors $\mathbf{x}_{obs}$ and the response Y, we propose a parametric imputation procedure to estimate the SIR matrix. To be specific, we take $\mathbf{x} = (\mathbf{x}_{mis}^{\mathrm{T}}, \mathbf{x}_{obs}^{\mathrm{T}})^{\mathrm{T}}$, where $\mathbf{x}_{mis} \in \mathbb{R}^{p_1}$ has missingness in a subset of subjects, $\mathbf{x}_{obs} \in \mathbb{R}^{p_2}$ has complete observations for all subjects, and $p = p_1 + p_2$. With slight abuse of notation, we define a vector of missingness indicators $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{p_1})^{\mathrm{T}} \in \mathbb{R}^{p_1}$, whose k-th coordinate δ_k takes value 1 if the k-th coordinate of $\mathbf{x}_{mis}$ is observed and 0 otherwise. In this situation, estimating the SIR matrix amounts to estimating $E(X_{mis,k} \mid \mathbf{x}_{obs}, Y)$ and $E(X_{mis,k}X_{mis,l} \mid \mathbf{x}_{obs}, Y)$ because $E(\mathbf{x}_{obs} \mid Y)$ and $E(\mathbf{x}_{obs}\mathbf{x}_{obs}^{\mathrm{T}} \mid Y)$ can be directly estimated. In order to estimate $E(X_{mis,k} \mid \mathbf{x}_{obs}, Y)$, we impute the missing values with $E(X_{mis,k}\delta_k \mid \mathbf{x}_{obs}, Y)/E(\delta_k \mid \mathbf{x}_{obs}, Y)$. Similarly, in order to estimate $E(X_{mis,k}X_{mis,l} \mid \mathbf{x}_{obs}, Y)$, we impute the missing values with $E(X_{mis,k}X_{mis,l}\delta_k\delta_l \mid \mathbf{x}_{obs}, Y)/E(\delta_k\delta_l \mid \mathbf{x}_{obs}, Y)$. Next we estimate $E(\delta_k \mid \mathbf{x}_{obs}, Y)$, $E(\delta_k\delta_l \mid \mathbf{x}_{obs}, Y)$, $E(X_{mis,k}\delta_k \mid \mathbf{x}_{obs}, Y)$, and $E(X_{mis,k}X_{mis,l}\delta_k\delta_l \mid \mathbf{x}_{obs}, Y)$ after imposing several parametric modeling assumptions, justified under the *missing at random* assumption that $\boldsymbol{\delta} \perp\!\!\!\perp \mathbf{x}_{mis} \mid (\mathbf{x}_{obs}, Y)$. The consistency of the parametric imputation procedure can be established if those parametric models are correctly specified. Our limited experience, gained from simulations, suggests that this parametric imputation procedure performs well in a wide range of semiparametric models.

The rest of this paper is organized as follows. We illustrate in detail the rationale of nonparametric imputation in Section 2 when the missingness is only relevant to the response. When the missingness is related to both the completely observed predictors and the response variable, we suggest imputing the missing values through a parametric imputation scheme in Section 3. Comprehensive simulation results are reported in Section 4 to augment the theoretical results and to compare with some existing methods. A horse colic dataset is analyzed in Section 5. There are concluding remarks in Section 6. The technical details are in the Appendix.

2. SIR with Nonparametric Imputation

In this section we first review SIR with full observation, then describe a nonparametric imputation procedure in the presence of missing predictors. The asymptotic properties of nonparametric imputation are established.

2.1. A brief review

SIR (Li (1991)) is a promising way to estimate $\mathcal{S}_{Y|\mathbf{x}}$. By assuming that $E(\mathbf{x} \mid \Gamma^{\mathrm{T}}\mathbf{x})$ is linear in $\Gamma^{\mathrm{T}}\mathbf{x}$, with Γ denoting a basis of $\mathcal{S}_{Y|\mathbf{x}}$, SIR connects $\mathcal{S}_{Y|\mathbf{x}}$ with the inverse mean $E(\mathbf{x} \mid Y)$ via the relationship span $\left(\boldsymbol{\Sigma}^{-1}\mathbf{M}\right) \subseteq \mathcal{S}_{Y|\mathbf{x}}$, where $\boldsymbol{\Sigma} = \mathrm{cov}\,(\mathbf{x}) = E\,(\mathbf{x}\mathbf{x}^{\mathrm{T}}) - E\,(\mathbf{x})\,E\,(\mathbf{x}^{\mathrm{T}})$, and $\mathbf{M} = \mathrm{cov}\,\{E(\mathbf{x} \mid Y)\} = E\,\{E(\mathbf{x} \mid Y)E(\mathbf{x}^{\mathrm{T}} \mid Y)\} - E\,(\mathbf{x})\,E\,(\mathbf{x}^{\mathrm{T}})$. The linearity condition holds to a reasonable approximation when the dimension p is fairly large (Hall and Li (1993)). To facilitate matters, we write

$$\Phi_0 = E(\mathbf{x}), \quad \Phi_1 = E(\mathbf{x}\mathbf{x}^{\mathrm{T}}), \text{ and } \Phi_2 = E\,\{E(\mathbf{x} \mid Y)E(\mathbf{x}^{\mathrm{T}} \mid Y)\}\,. \qquad (2.1)$$

Clearly $\mathbf{\Sigma} = \Phi_1 - \Phi_0\Phi_0^{\mathrm{T}}$, and $\mathbf{M} = \Phi_2 - \Phi_0\Phi_0^{\mathrm{T}}$.

Implementing SIR with n i.i.d. observations $\{(\mathbf{x}_i^{\mathrm{T}}, Y_i)^{\mathrm{T}}, i = 1, \ldots, n\}$ amounts to offering consistent estimators of the Φ_is. Because Φ_0 and Φ_1 can be estimated by their sample averages, estimating Φ_2 is the key to estimating the SIR matrix. Li (1991) proposed slicing estimation to estimate Φ_2; it is easy to implement. Zhu and Ng (1995) showed that the performance of SIR is robust to the number of slices, which echoes the empirical studies of Li (1991). In this paper, we follow Zhu and Fang (1996) and use kernel regression to estimate Φ_2. To be precise, let $R(Y) = \{R_1(Y), \ldots, R_p(Y)\}^{\mathrm{T}} = E(\mathbf{x} \mid Y) = \{E(X_1 \mid Y), \ldots, E(X_p \mid Y)\}^{\mathrm{T}}$. The kernel estimator of $R(Y)$ is $\widehat{R}_f(Y_i) = \sum_{j\neq i} K_h(Y_i - Y_j)\mathbf{x}_j / \sum_{j\neq i} K_h(Y_i - Y_j)$, where $K_h(\cdot) = K(\cdot/h)/h$, $K(\cdot)$ is a symmetric kernel function and h is a user-specified bandwidth. One can estimate Φ_2 with

$$\widehat{\Phi}_{2,f} = n^{-1}\sum_{i=1}^{n} \widehat{R}_f(Y_i)\widehat{R}_f^{\mathrm{T}}(Y_i).$$

Zhu and Fang (1996) showed that $\widehat{\Phi}_2$ is a root-n consistent estimator of Φ_2 when the predictors have no missing values.

2.2. Nonparametric imputation

When some predictors are subject to missingness, the procedure for estimating the SIR matrix $\mathbf{\Sigma}^{-1}\mathbf{M}$ cannot be applied directly. In this section we consider the *missing at random* (MAR) assumption that

$$\boldsymbol{\delta} \perp\!\!\!\perp \mathbf{x} \mid Y, \tag{2.2}$$

where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^{\mathrm{T}}$ is a vector of missingness indicators. The k-th component δ_k takes value 1 if there is no missingness for the k-th predictor coordinate X_k, and 0 otherwise. This MAR assumption provides in general a better approximation to reality and is less restrictive than the MCAR assumption. A more general MAR assumption will be discussed in the next section.

To implement SIR with missing predictors, we estimate Φ_is as follows. Let $\{(\mathbf{x}_j^{\mathrm{T}}, Y_j)^{\mathrm{T}}, j = 1, \ldots, n\}$ be a set of i.i.d. random vectors, where the response Y_j is always observable, and the p-dimensional predictor vector $\mathbf{x}_j = (X_{1j}, \ldots, X_{pj})^{\mathrm{T}}$ is subject to missingness. In practice, the missing components may vary among the incomplete predictors, thus we define the binary indicator $\delta_{kj} = 1$ if X_{kj} is observed, and $\delta_{kj} = 0$ if X_{kj} is missing, for $k = 1, \ldots, p$, and $j = 1, \ldots, n$.

Estimation of Φ_0 and Φ_2: We first discuss how to estimate Φ_0 and Φ_2 at the population level. Recall that $\Phi_0 = E\{E(\mathbf{x} \mid Y)\}$ and $\Phi_2 = E\{E(\mathbf{x} \mid Y)E(\mathbf{x}^{\mathrm{T}} \mid Y)\}$; this motivates us to estimate both Φ_0 and Φ_2 via the inverse mean $E(\mathbf{x} \mid Y)$.

To be precise, if the predictor value X_k is missing, we follow the idea of Cheng (1994) and impute the value of inverse mean $R_k(Y) = E(X_k \mid Y)$. We resort to the MAR assumption (2.2) to estimate $R_k(Y)$; it implies that $\mathrm{cov}(X_k, \delta_k \mid Y) = E(X_k\delta_k \mid Y) - E(X_k \mid Y)E(\delta_k \mid Y) = 0$. That is,

$$R_k\left(Y\right) = \frac{E\left(X_k\delta_k \mid Y\right)}{E(\delta_k \mid Y)}, \tag{2.3}$$

providing that δ_k does not degenerate. The RHS of (2.3) can be easily estimated even when X_k has missing values: if X_k is missing, we can impute $E\left(X_k\delta_k \mid Y\right)/E(\delta_k \mid Y)$ because it is equal to $E(X_k \mid Y)$. An important observation is that $X_k\delta_k$, δ_k, and Y are observable, although a subset of X_k has missing values, which allows us to estimate $E\left(X_k\delta_k \mid Y\right)$ and $E(\delta_k \mid Y)$ consistently using all observations to improve the efficiency.

The above imputation procedure is easy to implement at the sample level. To be precise, when X_{kj} is missing for some $1 \le j \le n$, we impute $\widehat{X}_{k,j} = \widehat{G}_k(Y_j)/\widehat{g}_k(Y_j)$, where $\widehat{G}_k(Y_j) = \sum_{i \ne j} K_h(Y_i - Y_j)X_{ki}\delta_{ki}/(n-1)$, and $\widehat{g}_k(Y_j) = \sum_{i \ne j} K_h(Y_i - Y_j)\delta_{ki}/(n-1)$. Following arguments of Zhu and Fang (1996) and Zhu and Zhu (2007)), we can see without much difficulty that $\widehat{X}_{k,j}$ is a consistent estimator of $E\left(X_k\delta_k \mid Y\right)/E(\delta_k \mid Y) = E(X_{kj} \mid Y_j)$ in view of (2.2). Therefore, we can estimate $E(X_k)$, the k-th element of Φ_0, by

$$\widehat{\Phi}_{0,k} = n^{-1}\sum_{j=1}^{n}\left\{\delta_{kj}X_{kj} + (1-\delta_{kj})\widehat{X}_{k,j}\right\}. \tag{2.4}$$

Next we deal with Φ_2. Consider the kernel estimator

$$\widehat{R}_k(Y_i) = (n-1)^{-1}\sum_{j \ne i}\frac{K_h(Y_j - Y_i)\left\{\delta_{kj}X_{kj} + (1-\delta_{kj})\widehat{X}_{k,j}\right\}}{\widehat{f}(Y_i)}, \tag{2.5}$$

where $\widehat{f}(Y_i) = \sum_{j \ne i} K_h(Y_j - Y_i)/(n-1)$ is the kernel estimator of the density function of Y. The (k,l)-th element of Φ_2 can be estimated, accordingly, by

$$\widehat{\Phi}_{2,kl} = n^{-1}\sum_{i=1}^{n}\left\{\widehat{R}_k(Y_i)\widehat{R}_l(Y_i)\right\}, \tag{2.6}$$

Remark 1. When Y is categorical or discrete, the nonparametric imputation method is still readily applicable. In such situations, we can still impute the missing value X_{kj} with the inverse mean $R_k(Y_j)$, while estimator of $R_k(Y_j)$ is simplified to $\widehat{X}^d_{k,j} = \sum_{i \ne j} \mathbf{1}(Y_i = Y_j)X_{ki}\delta_{ki}/\sum_{i \ne j} \mathbf{1}(Y_i = Y_j)\delta_{ki}$. Then $\widehat{\Phi}_{0,k}$

takes a similar form to (2.4) in which $\widehat{X}_{k,j}$ is replaced with $\widehat{X}^d_{k,j}$. To estimate Φ_2, we take, in parallel to the continuous case,

$$\widehat{R}^d_k(Y_j) = \sum_{i\neq j} \mathbf{1}\{Y_i = Y_j\} \left\{\delta_{ki}X_{ki} + (1-\delta_{ki})\widehat{X}^d_{k,i}\right\} \Big/ \sum_{i\neq j} \mathbf{1}\{Y_i = Y_j\},$$

The (k,l)-th element of Φ_2 can be estimated in a similar fashion.

Estimation of Φ_1: Note that $E(\mathbf{x}\mathbf{x}^{\mathrm{T}}) = E\{E(\mathbf{x}\mathbf{x}^{\mathrm{T}} \mid Y)\}$. Accordingly, estimating Φ_1 is in the spirit of estimating Φ_0. Note that when either X_k or X_l is missing, X_kX_l is then missing, which implies, $\delta_k\delta_l = 0$. Similarly, $\delta_k\delta_l = 1$ if and only if both X_k and X_l are observed. Thus we propose to impute a missing value X_kX_l with an estimator of the inverse mean $R_{kl}(Y) = E(X_kX_l \mid Y)$. The MAR assumption (2.2) implies that $X_kX_l \perp\!\!\!\perp \delta_k\delta_l \mid Y$. Consequently, $R_{kl}(Y) = E(X_kX_l\delta_k\delta_l \mid Y)/E(\delta_k\delta_l \mid Y)$ for $1 \leq k,l \leq p$.

If $X_{kj}X_{lj}$ is missing for some $1 \leq j \leq n$, a reasonable imputation value can be $\widehat{X}_{kl,j} = \widehat{G}_{kl}(Y_j)/\widehat{g}_{kl}(Y_j)$, where $\widehat{G}_{kl}(Y_j) = (n-1)^{-1}\sum_{i\neq j} K_h(Y_i - Y_j)X_{ki}X_{li}\delta_{ki}\delta_{li}$ and $\widehat{g}_{kl}(Y_j) = (n-1)^{-1}\sum_{i\neq j} K_h(Y_i - Y_j)\delta_{ki}\delta_{li}$. Then we can estimate $E(X_kX_l)$, the (k,l)-th element of Φ_1, by

$$\widehat{\Phi}_{1,kl} = n^{-1}\sum_{j=1}^{n}\left\{\delta_{kj}\delta_{lj}X_{kj}X_{lj} + (1-\delta_{kj}\delta_{lj})\widehat{X}_{kl,j}\right\}. \tag{2.7}$$

Remark 2. When the response is categorical or discrete, we take $\widehat{X}^d_{kl,j} = \sum_{i\neq j}\mathbf{1}(Y_i = Y_j)X_{ki}X_{li}\delta_{ki}\delta_{li}/\sum_{i\neq j}\mathbf{1}(Y_i = Y_j)\delta_{ki}\delta_{ki}$. Then $\widehat{\Phi}_{1,kl}$ takes a similar form to (2.7) through replacing $\widehat{X}_{kl,j}$ with $\widehat{X}^d_{kl,j}$.

Estimation of $\mathbf{\Sigma}^{-1}\mathbf{M}$: Through using the proposed nonparametric imputation procedures in (2.4), (2.6), and (2.7), we can estimate $\mathbf{\Sigma}$ with $\mathbf{\Sigma}_n = \widehat{\Phi}_1 - \widehat{\Phi}_0\widehat{\Phi}_0^{\mathrm{T}}$, and $\mathbf{M}$ with $\mathbf{M}_n = \widehat{\Phi}_2 - \widehat{\Phi}_0\widehat{\Phi}_0^{\mathrm{T}}$. A natural estimate of $\mathbf{\Sigma}^{-1}\mathbf{M}$ is thus $\mathbf{\Sigma}_n^{-1}\mathbf{M}_n$.

2.3. Asymptotic properties

In this section we study the asymptotic behavior of the nonparametric imputation method. For ease of exposition, let $f(y)$ be the density function of Y, assumed to be bounded away from zero and above. Let $\pi_k(Y) = E(\delta_k \mid Y)$, $r_k(Y) = E(X_k\delta_k \mid Y)$, $R_k(Y) = E(X_k \mid Y)$, $g_k(Y) = \pi_k(Y)f(Y)$ and $G_k(Y) = r_k(Y)f(Y)$, for $k = 1,\ldots,p$. We set regularity conditions to ensure the desired asymptotic properties. These technical conditions are not the weakest possible, but they are imposed to facilitate the proofs.

(1) The d-th order kernel function $K(\cdot)$ is symmetric for some $d \geq 2$. It has support on the interval (-1,1). In addition, $\int_{-1}^{1} K^2(u)du < \infty$.

(2) The $(d-1)$-th derivatives of functions $f(y)$, $f(y)R_k(y)$, $R_k(y)$, $g_k(y)$, and $G_k(y)$ are locally Lipschitz.

(3) The bandwidth $h = O(n^{-c})$, where $1/(2d) < c < 1/2$.

(4) $E\left(\mathbf{x}^{\mathrm{T}}\mathbf{x}\right)^2 < \infty$.

Theorem 1. *Under the MAR assumption* (2.2) *and Conditions* (1)−(4), *as* $n \to \infty$,

$$\sqrt{n}\{\mathrm{vec}(\boldsymbol{\Sigma}_n^{-1}\mathbf{M}_n) - \mathrm{vec}(\boldsymbol{\Sigma}^{-1}\mathbf{M})\} \quad \text{is asymptotically normal,} \tag{2.8}$$

where vec *denotes the operator that stacks all columns of a matrix to a vector.*

Theorem 1 states the convergence rate of $\boldsymbol{\Sigma}_n^{-1}\mathbf{M}_n$ when Y is continuous. We establish the asymptotic normality for a general class of kernel functions; in implementations we choose the second-order Epanechnikov kernel function. The asymptotic normality can be derived similarly when Y is discrete or categorical; details of the proof are omitted from the present context.

3. SIR with Parametric Imputation

In this section we consider a more general missing-data mechanism than (2.2), allowing the missingness to depend on both the completely observed predictors and the response variable Y. Take $\mathbf{x} = (\mathbf{x}_{mis}^{\mathrm{T}}, \mathbf{x}_{obs}^{\mathrm{T}})^{\mathrm{T}}$, where $\mathbf{x}_{mis} \in \mathbb{R}^{p_1}$ has missingness in a subset of subjects, $\mathbf{x}_{obs} \in \mathbb{R}^{p_2}$ has complete observations for all subjects, $p = p_1 + p_2$, and the vector of missingness indicators $\boldsymbol{\delta}$ as before. Here we make the MAR assumption that

$$\boldsymbol{\delta} \perp\!\!\!\perp \mathbf{x}_{mis} \mid (\mathbf{x}_{obs}, Y)\,. \tag{3.1}$$

Note that (2.2) implies (3.1). The MAR assumption (3.1) in the inverse regression setup was introduced in Li and Lu (2008), and has its roots in Rubin (1976). In this section, we introduce a parametric imputation procedure for SIR under (3.1).

Similar to (2.1), we consider the partitions:

$$\begin{aligned}
\Phi_0 &= \left(E(\mathbf{x}_{mis}^{\mathrm{T}}), E(\mathbf{x}_{obs}^{\mathrm{T}})\right)^{\mathrm{T}},\\
\Phi_1 &= \begin{pmatrix} E(\mathbf{x}_{mis}\mathbf{x}_{mis}^{\mathrm{T}}) & E(\mathbf{x}_{mis}\mathbf{x}_{obs}^{\mathrm{T}}) \\ E(\mathbf{x}_{obs}\mathbf{x}_{mis}^{\mathrm{T}}) & E(\mathbf{x}_{obs}\mathbf{x}_{obs}^{\mathrm{T}}) \end{pmatrix}, \text{ and}\\
\Phi_2 &= \begin{pmatrix} E\left\{E(\mathbf{x}_{mis}|Y)E(\mathbf{x}_{mis}^{\mathrm{T}}|Y)\right\} & E\left\{E(\mathbf{x}_{mis}|Y)E(\mathbf{x}_{obs}^{\mathrm{T}}|Y)\right\} \\ E\left\{E(\mathbf{x}_{obs}|Y)E(\mathbf{x}_{mis}^{\mathrm{T}}|Y)\right\} & E\left\{E(\mathbf{x}_{obs}|Y)E(\mathbf{x}_{obs}^{\mathrm{T}}|Y)\right\} \end{pmatrix}.
\end{aligned}$$

To implement SIR, one must estimate eight quantities: $E(\mathbf{x}_{obs})$, $E(\mathbf{x}_{obs}\mathbf{x}_{obs}^{\mathrm{T}})$, $E\left\{E(\mathbf{x}_{obs} \mid Y)E(\mathbf{x}_{obs}^{\mathrm{T}} \mid Y)\right\}$, $E(\mathbf{x}_{mis})$, $E(\mathbf{x}_{mis}\mathbf{x}_{mis}^{\mathrm{T}})$, $E(\mathbf{x}_{mis}\mathbf{x}_{obs}^{\mathrm{T}})$, $E\{E(\mathbf{x}_{mis} \mid Y)$ $E(\mathbf{x}_{mis}^{\mathrm{T}} \mid Y)\}$, and $E\left\{E(\mathbf{x}_{mis} \mid Y)E(\mathbf{x}_{obs}^{\mathrm{T}} \mid Y)\right\}$. Note that the first three can be estimated as usual, because they involve no missing observations. Refer to Section 2.1 or to Zhu and Fang (1996) for more details about SIR estimation with no missing values. New consistent estimators are desired for the last five quantities; we discuss how to estimate them in the sequel.

To facilitate matters, we denote by $X_{mis,k}$ and δ_k the k-th coordinates of $\mathbf{x}_{mis}$ and $\boldsymbol{\delta}$, respectively, for $k = 1, \ldots, p_1$.

Estimation of $E(\mathbf{x}_{mis})$, $E(\mathbf{x}_{mis}\mathbf{x}_{obs}^{\mathrm{T}})$, $E\{E(\mathbf{x}_{mis}|Y)E(\mathbf{x}_{mis}^{\mathrm{T}}|Y)\}$, and $E\{E(\mathbf{x}_{mis}|Y)E(\mathbf{x}_{obs}^{\mathrm{T}}|Y)\}$:

We discuss how to estimate $E(X_{mis,k})$ in detail; the rationale for estimating the other quantities is similar. We first note that

$$E(X_{mis,k}) = E\{E(X_{mis,k}|\mathbf{x}_{obs}, Y)\} = E\left\{\frac{E(X_{mis,k}\delta_k|\mathbf{x}_{obs}, Y)}{E(\delta_k|\mathbf{x}_{obs}, Y)}\right\}. \tag{3.2}$$

The first equality follows from the law of iterative expectations. In view of (3.1), one can easily have $\mathrm{cov}(X_{mis,k}, \delta_k \mid \mathbf{x}_{obs}, Y) = 0$. Thus, $E(X_{mis,k} \mid \mathbf{x}_{obs}, Y) = E(X_{mis,k}\delta_k \mid \mathbf{x}_{obs}, Y)/E(\delta_k \mid \mathbf{x}_{obs}, Y)$, which entails the second equality of (3.2). As $X_{mis,k}\delta_k$, δ_k, $\mathbf{x}_{obs}$ and Y are observable, we can estimate the RHS of (3.2) via standard procedures, to be described below.

When the dimension p_2 of $\mathbf{x}_{obs}$ is small compared with the sample size n, we can follow the ideas in Section 2.2 to estimate $E(X_{mis,k}\delta_k \mid \mathbf{x}_{obs}, Y)$ and $E(\delta_k \mid \mathbf{x}_{obs}, Y)$ directly through local smoothing techniques, such as the kernel regression. An anonymous referee cautioned that the kernel-based imputation method may run into the *curse of dimensionality* when p_2 of $\mathbf{x}_{obs}$ is fairly large. To address this issue, we posit several parameter models for estimating $E(X_{mis,k}\delta_k|\mathbf{x}_{obs}, Y)$ and $E(\delta_k|\mathbf{x}_{obs}, Y)$. Let

$$E(\delta_k \mid \mathbf{x}_{obs}, Y) = \pi_k(\mathbf{x}_{obs}, Y; \boldsymbol{\alpha}_{1,k}), \text{ and} \tag{3.3}$$

$$E(X_{mis,k}\delta_k \mid \mathbf{x}_{obs}, Y) = \psi_k(\mathbf{x}_{obs}, Y; \boldsymbol{\gamma}_{1,k}), \text{ for } k = 1, \ldots, p_1. \tag{3.4}$$

These functions are indexed by the parameters $\boldsymbol{\alpha}_{1,k}$ and $\boldsymbol{\gamma}_{1,k}$; to estimate them, it suffices to estimate the involved parameters. Because δ_k is binary, a natural choice for $\pi_k(\mathbf{x}_{obs}, Y; \boldsymbol{\alpha}_{1,k})$ is logistic regression, but other parametric models can be easily accommodated as well. With the logistic regression, we can use maximum likelihood to estimate the parameters $\boldsymbol{\alpha}_{1,k}$'s for $k = 1, \ldots, p_1$. Similarly for $\psi_k(\mathbf{x}_{obs}, Y; \boldsymbol{\gamma}_{1,k})$, a convenient option is the linear regression model, but other parameter models can be accommodated as well. Given the data $\{(\mathbf{x}_1, Y_1, \boldsymbol{\delta}_1), \ldots, (\mathbf{x}_n, Y_n, \boldsymbol{\delta}_n)\}$, where $\mathbf{x}_i = (\mathbf{x}_{mis,i}^{\mathrm{T}}, \mathbf{x}_{obs,i}^{\mathrm{T}})^{\mathrm{T}}$, $\boldsymbol{\gamma}_{1,k}$ can be estimated by least squares. Denote by $\widehat{\boldsymbol{\alpha}}_{1,k}$ and $\widehat{\boldsymbol{\gamma}}_{1,k}$ the respective estimators of $\boldsymbol{\alpha}_{1,k}$ and $\boldsymbol{\gamma}_{1,k}$. These estimates are root-n consistent provided the parametric models $\pi_k(\mathbf{x}_{obs}, Y; \boldsymbol{\alpha}_{1,k})$ and $\psi_k(\mathbf{x}_{obs}, Y; \boldsymbol{\gamma}_{1,k})$ are correctly specified. Thus, one can estimate $E(X_{mis,k} \mid \mathbf{x}_{obs}, Y)$ by

$$\widehat{E}(X_{mis,k} \mid \mathbf{x}_{obs}, Y) = \frac{\widehat{E}(X_{mis,k}\delta_k \mid \mathbf{x}_{obs}, Y)}{\widehat{E}(\delta_k \mid \mathbf{x}_{obs}, Y)} = \frac{\psi_k(\mathbf{x}_{obs}, Y; \widehat{\boldsymbol{\gamma}}_{1,k})}{\pi_k(\mathbf{x}_{obs}, Y; \widehat{\boldsymbol{\alpha}}_{1,k})}. \tag{3.5}$$

With (3.5), a natural option for estimating $E(X_{mis,k})$ is

$$\widehat{E}(X_{mis,k}) = n^{-1}\sum_{i=1}^{n}\left\{\delta_{ki}X_{ki} + (1-\delta_{ki})\,\widehat{E}(X_{mis,k} \mid \mathbf{x}_{obs,i}, Y_i)\right\}.$$

Note that $E(\mathbf{x}_{mis}\mathbf{x}_{obs}^{\mathrm{T}}) = E\left\{E(\mathbf{x}_{mis} \mid \mathbf{x}_{obs}, Y)\mathbf{x}_{obs}^{\mathrm{T}}\right\}$. Similarly, we can estimate the k-th row of $E(\mathbf{x}_{mis}\mathbf{x}_{obs}^{\mathrm{T}})$, $E(X_{mis,k}\mathbf{x}_{obs}^{\mathrm{T}})$, by

$$\widehat{E}(X_{mis,k}\mathbf{x}_{obs}^{\mathrm{T}}) = n^{-1}\sum_{i=1}^{n}\left\{\delta_{ki}X_{mis,k,i} + (1-\delta_{ki})\,\widehat{E}(X_{mis,k} \mid \mathbf{x}_{obs,i}, Y_i)\right\}\mathbf{x}_{obs,i}^{\mathrm{T}}.$$

Note that $E(\mathbf{x}_{mis} \mid Y) = E\left\{E(\mathbf{x}_{mis} \mid \mathbf{x}_{obs}, Y) \mid Y\right\}$. To estimate the k-th component $E(X_{mis,k} \mid Y)$, we can use

$$\widehat{E}(X_{mis,k} \mid Y) = n^{-1}\sum_{i=1}^{n}\frac{K_h(Y_i - Y)\left\{\delta_{ki}X_{ki} + (1-\delta_{ki})\widehat{E}(X_{mis,k} \mid \mathbf{x}_{obs,i}, Y_i)\right\}}{\widehat{f}(Y)}.$$

Recall that $E(\mathbf{x}_{obs} \mid Y)$ can be estimated with the usual kernel smoother $\widehat{R}_f(Y)$ defined in Section 2.1. With consistent estimates $\widehat{E}(X_{mis,k} \mid Y)$ and $\widehat{R}_f(Y)$, one can follow similar procedures as proposed in this section to construct consistent estimators for $E\left\{E(\mathbf{x}_{mis} \mid Y)E(\mathbf{x}_{mis}^{\mathrm{T}} \mid Y)\right\}$ and $E\left\{E(\mathbf{x}_{mis} \mid Y)E(\mathbf{x}_{obs}^{\mathrm{T}} \mid Y)\right\}$. Details are omitted for the sake of brevity.

Estimation of $E(\mathbf{x}_{mis}\mathbf{x}_{mis}^{\mathrm{T}})$: Here, estimation is similar to that of $E(\mathbf{x}_{mis})$. We only sketch the outline below. Note that

$$\begin{aligned} E(X_{mis,k}X_{mis,l}) &= E\left\{E(X_{mis,k}X_{mis,l} \mid \mathbf{x}_{obs}, Y)\right\} \\ &= E\left\{\frac{E(X_{mis,k}X_{mis,l}\delta_k\delta_l \mid \mathbf{x}_{obs}, Y)}{E(\delta_k\delta_l \mid \mathbf{x}_{obs}, Y)}\right\}. \end{aligned}$$

This follows from (3.1), since $\mathrm{cov}(X_{mis,k}X_{mis,l}, \delta_k\delta_l \mid \mathbf{x}_{obs}, Y) = 0$.

When the dimension p_1 of $\mathbf{x}_{obs}$ is large, we posit several parameter models for estimating $E(X_{mis,k}X_{mis,l}\delta_k\delta_l \mid \mathbf{x}_{obs}, Y)$ and $E(\delta_k\delta_l \mid \mathbf{x}_{obs}, Y)$. Let

$$E(\delta_k\delta_l \mid \mathbf{x}_{obs}, Y) = \pi_{k,l}(\mathbf{x}_{obs}, Y; \boldsymbol{\alpha}_{2,kl}), \text{ and} \tag{3.6}$$

$$E(X_{mis,k}X_{mis,l}\delta_k\delta_l \mid \mathbf{x}_{obs}, Y) = \psi_{k,l}(\mathbf{x}_{obs}, Y; \boldsymbol{\gamma}_{2,kl}), \text{ for } k, l = 1, \ldots, p_1. \tag{3.7}$$

Because $\delta_k\delta_l$ is also binary, we take logistic regression model for $\pi_{k,l}(\mathbf{x}_{obs}, Y; \boldsymbol{\alpha}_{2,kl})$, and use the maximum likelihood to estimate the parameters $\boldsymbol{\alpha}_{2,kl}$. Similarly, we take the linear regression model for $\psi_{k,l}(\mathbf{x}_{obs}, Y; \boldsymbol{\gamma}_{2,kl})$ and estimate $\boldsymbol{\gamma}_{2,kl}$ with least squares. Denote by $\widehat{\boldsymbol{\alpha}}_{2,kl}$ and $\widehat{\boldsymbol{\gamma}}_{2,kl}$ the estimators of $\boldsymbol{\alpha}_{2,kl}$ and $\boldsymbol{\gamma}_{2,kl}$. These estimators are root-n consistent if the parametric models $\pi_{k,l}(\mathbf{x}_{obs}, Y; \boldsymbol{\alpha}_{2,kl})$ and $\psi_{k,l}(\mathbf{x}_{obs}, Y; \boldsymbol{\gamma}_{2,kl})$ are correctly specified. Thus,

$$\begin{aligned}\widehat{E}(X_{mis,k}X_{mis,l} \mid \mathbf{x}_{obs}, Y) &= \frac{\widehat{E}(X_{mis,k}X_{mis,l}\delta_k\delta_l \mid \mathbf{x}_{obs}, Y)}{\widehat{E}(\delta_k\delta_l \mid \mathbf{x}_{obs}, Y)}\\ &= \frac{\psi_{k,l}(\mathbf{x}_{obs}, Y; \widehat{\boldsymbol{\gamma}}_{2,kl})}{\pi_{k,l}(\mathbf{x}_{obs}, Y; \widehat{\boldsymbol{\alpha}}_{2,kl})}. \end{aligned} \tag{3.8}$$

With (3.8), one can easily estimate $E(X_{mis,k}X_{mis,l})$ by

$$\begin{aligned}&\widehat{E}(X_{mis,k}X_{mis,l})\\ &\quad = n^{-1}\sum_{i=1}^{n}\left\{\delta_{ki}\delta_{li}X_{mis,k,i}X_{mis,l,i} + (1-\delta_{ki}\delta_{li})\,\widehat{E}(X_{mis,k}X_{mis,l} \mid \mathbf{x}_{obs,i}, Y_i)\right\}.\end{aligned}$$

Estimation of $\boldsymbol{\Sigma}^{-1}\mathbf{M}$: Through replacing the unknowns in Φ_0, Φ_1, and Φ_2 with their corresponding estimates, we estimate $\boldsymbol{\Sigma}$ with $\boldsymbol{\Sigma}'_n = \widehat{\Phi}_1 - \widehat{\Phi}_0\widehat{\Phi}_0^{\mathrm{T}}$, and $\mathbf{M}$ by $\mathbf{M}'_n = \widehat{\Phi}_2 - \widehat{\Phi}_0\widehat{\Phi}_0^{\mathrm{T}}$. A natural estimate of $\boldsymbol{\Sigma}^{-1}\mathbf{M}$ is $\boldsymbol{\Sigma}'^{-1}_n\mathbf{M}'_n$.

The consistency of the parametric imputation of SIR is stated in a theorem.

Theorem 2. *Under Conditions* (1)−(4)*, the MAR assumption* (3.1)*, and that the parametric models* (3.3)−(3.4)*,* (3.6)−(3.7) *are correctly specified,* $\widehat{\boldsymbol{\alpha}}_{1,k}$*,* $\widehat{\boldsymbol{\gamma}}_{1,k}$*,* $\widehat{\boldsymbol{\alpha}}_{2,kl}$*, and* $\widehat{\boldsymbol{\gamma}}_{2,kl}$ *are root-n consistent, for* $k, l = 1, \ldots, p_1$*, as* $n \to \infty$*,*

$$\sqrt{n}\{\mathrm{vec}(\boldsymbol{\Sigma}'^{-1}_n\mathbf{M}'_n) - \mathrm{vec}(\boldsymbol{\Sigma}^{-1}\mathbf{M})\} \quad \text{is asymptotically normal.}$$

4. Simulations

In this section we examine the finite-sample performance of the proposed nonparametric and parametric imputation procedures through synthetic studies. We compared the performance of seven proposals in our simulations.

(1) NP-KIR: Nonparametric imputation with kernel estimation of the SIR matrix. Refer to Section 2 for details.

(2) NP-SIR: Nonparametric imputation with the slicing estimation of the SIR matrix. This differs from NP-KIR in that NP-SIR uses the slicing estimation to impute missing observations and to estimate the SIR matrix.

(3) P-KIR: Parametric imputation with kernel estimation of the SIR matrix. Refer to Section 3 for details.

(4) P-SIR: Parametric imputation with the slicing estimation of the SIR matrix. This differs from P-KIR in that P-SIR uses the slicing estimation to impute missing observations and to estimate the SIR matrix.

(5) CC-KIR: The naive complete-case analysis that uses only the complete observations in estimating $\mathcal{S}_{Y|\mathbf{x}}$. We use the kernel estimation to estimate the SIR matrix.

(6) CC-SIR: The naive complete-case analysis that uses only the complete observations in estimating $\mathcal{S}_{Y|\mathbf{x}}$. We use the slicing estimation to estimate the SIR matrix.

(7) FC-KIR: The full-case analysis that uses all n observations that have no missing predictors. We use the kernel estimation to estimate the SIR matrix.

(8) FC-SIR: The full-case analysis that uses all n observations that have no missing predictors. This differs from FC-KIR in that we use the slicing estimation to estimate the SIR matrix.

(9) AIPW: The augmented inverse probability weighted SIR introduced by Li and Lu (2008).

(10) AIPWM: The marginal augmented inverse probability weighted SIR introduced by Li and Lu (2008).

Because the slicing estimation is insensitive to the number of slices H (Li (1991), Zhu and Ng (1995)), we only report the results with $H = 5$ and $H = 10$. For procedures (1), (3), (5) and (7) that use kernel smoothing, we used cross-validation to obtain an optimal bandwidth h_{opt}, then used $h = n^{-2/15}h_{opt}$ as the resulting bandwidth.

We adopted two models for the purposes of comparison:

$$Y = (\boldsymbol{\gamma}_1^{\mathrm{T}}\mathbf{x})(\boldsymbol{\gamma}_1^{\mathrm{T}}\mathbf{x} + \boldsymbol{\gamma}_2^{\mathrm{T}}\mathbf{x} + 3) + 0.5\varepsilon, \tag{4.1}$$

$$Y = \frac{\boldsymbol{\gamma}_1^{\mathrm{T}}\mathbf{x}}{(\boldsymbol{\gamma}_2^{\mathrm{T}}\mathbf{x} + 1.5)^2 + 0.5} + 0.5\varepsilon. \tag{4.2}$$

These models were used in Li (1991) and Li and Lu (2008). We followed their settings and drew $\mathbf{x} = (X_1, \ldots, X_p)^{\mathrm{T}}$ from a multivariate normal distribution with mean zero and covariance $0.3^{|k-l|}$ between X_k and X_l, for $1 \leq k, l \leq p$. The error term ε was independent standard normal. We set $\boldsymbol{\gamma}_1 = (1, 0, 0, 0, 0)^{\mathrm{T}}$ and $\boldsymbol{\gamma}_2 = (0, 1, 0, 0, 0)^{\mathrm{T}}$. When $p > 5$, the rest of the components of $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ were set to zero. Simulations were repeated 500 times, each of sample size $n = 200$. We chose $p = 5$ and $p = 10$ to compare different methods.

Example 1. In this example, we took a subset of predictors to have missing observations, with the probability π given the response Y as

$$\pi_k(Y) = \text{prob}(\delta_k = 1 \mid Y) = \frac{\exp(c_0 - 0.25Y)}{1 + \exp(c_0 - 0.25Y)}, \tag{4.3}$$

c_0 is a scalar constant that controls the missingness proportion. We chose $c_0 = -1, 0$ and 1 to evaluate the effect of the missing proportion on the efficacy of the various methods. The empirical missingness proportion, indicated by "mp" in Tables 1−2 and determined by c_0, is also reported in Tables 1−2. We can see that the larger c_0 values indicate fewer missing values.

We considered two cases: (i) only X_1 has missing values, and (ii) both X_1 and X_2 have missing values. To measure the estimation accuracy, we adopted the trace correlation coefficient proposed by Ferré (1998). To be precise, let $\widehat{\mathcal{S}}_{Y|\mathbf{x}}$ be an estimator of $\mathcal{S}_{Y|\mathbf{x}}$, and let P and $\widehat{P}$ be the respective projection operators in the standard inner product of $\mathcal{S}_{Y|\mathbf{x}}$ and its estimator $\widehat{\mathcal{S}}_{Y|\mathbf{x}}$. The trace correlation is $R^2(K) = \text{tr}(P\widehat{P})/K$; $K = \dim(\mathcal{S}_{Y|\mathbf{x}}) = 2$ was assumed to be known in advance. This measure describes the "closeness" between the estimated and the true subspaces; it ranges from 0 and 1, with larger values indicating better estimation. We report the median and the median absolute deviation of $R^2(K)$ values over 500 repetitions in Table 1, for $p = 5$ and in Table 2 for $p = 10$.

For the full-case analyses such as FC-KIR and FC-SIR, the data contain no missing values, and we only report one result for each model. When only X_1 has missing values, AIPWM is equivalent to AIPW, hence we only report results for AIPW in Tables 1−2.

It can be seen from Tables 1−2 that, in most scenarios, the proposals using the kernel smoothing and their counterparts using the slicing estimation have comparable performance. Yet, kernel smoothing slightly outperforms the slicing estimation when both X_1 and X_2 have a large proportion of missing values. We use Table 1 as an example because Table 2 conveys a similar message. In model (4.1) with $c = -1$, the median trace correlation is 0.915 for P-KIR, and 0.594 for P-SIR with $H = 5$, 0.564 for P-SIR with $H = 10$. In model (4.2) with $c = -1$, the median trace correlation is 0.923 for NP-KIR, 0.772 for NP-SIR with $H = 5$ and 0.813 for NP-SIR with $H = 10$.

Full-case analysis serves naturally as a benchmark. The results of both FC-KIR and FC-SIR are reported in Tables 1−12. They perform the best across all scenarios because all data can be used precisely when there is no missingness.

Complete-case analysis performs the worst in most scenarios. In Table 1 as an example when both X_1 and X_2 have missing values with $c = -1$, the median trace correlation for CC-KIR is as small as 0.520 in model (4.1) and 0.537 in

Table 1. The median and the median absolute deviation of the trace correlation coefficients for models (4.1) and (4.2) with $p = 5$. The missingness follows (4.3).

$n = 200, p = 5$	only X_1 has missing values			both X_1 and X_2 have missing values		
	model (4.1)					
	$c_0=-1$	$c_0=0$	$c_0=1$	$c_0=-1$	$c_0=0$	$c_0=1$
NP-KIR	0.934±0.049	0.934±0.055	0.945±0.048	0.917±0.065	0.921±0.069	0.929±0.064
NP-SIR ($H=5$)	0.930±0.044	0.940±0.039	0.942±0.040	0.858±0.102	0.903±0.071	0.924±0.054
NP-SIR ($H=10$)	0.931±0.041	0.942±0.037	0.946±0.034	0.880±0.089	0.918±0.061	0.932±0.047
P-KIR	0.928±0.072	0.922±0.074	0.938±0.065	0.915±0.064	0.899±0.092	0.923±0.070
P-SIR ($H=5$)	0.856±0.107	0.860±0.109	0.888±0.090	0.594±0.134	0.734±0.171	0.863±0.095
P-SIR ($H=10$)	0.832±0.137	0.838±0.124	0.887±0.100	0.564±0.099	0.663±0.186	0.848±0.121
CC-KIR	0.793±0.212	0.909±0.085	0.952±0.045	0.520±0.125	0.751±0.256	0.900±0.102
CC-SIR ($H=5$)	0.787±0.175	0.883±0.091	0.928±0.055	0.518±0.099	0.744±0.190	0.883±0.092
CC-SIR ($H=10$)	0.759±0.190	0.879±0.105	0.926±0.059	0.499±0.114	0.739±0.194	0.877±0.102
FC-KIR	0.973±0.023	-	-	-	-	-
FC-SIR ($H=5$)	0.949±0.042	-	-	-	-	-
FC-SIR ($H=10$)	0.954±0.036	-	-	-	-	-
AIPW ($H=5$)	0.934±0.052	0.937±0.048	0.937±0.049	0.538±0.070	0.604±0.148	0.834±0.156
AIPW ($H=10$)	0.937±0.050	0.939±0.045	0.941±0.042	0.599±0.151	0.716±0.262	0.851±0.157
AIPWM ($H=5$)	-	-	-	0.670±0.215	0.847±0.134	0.912±0.069
AIPWM ($H=10$)	-	-	-	0.701±0.244	0.848±0.137	0.915±0.072
mp	0.755±0.029	0.560±0.037	0.345±0.037	0.930±0.014	0.775±0.029	0.540±0.029
	model (4.2)					
	$c_0=-1$	$c_0=0$	$c_0=1$	$c_0=-1$	$c_0=0$	$c_0=1$
NP-KIR	0.947±0.038	0.954±0.039	0.952±0.039	0.923±0.062	0.944±0.045	0.950±0.038
NP-SIR ($H=5$)	0.840±0.111	0.873±0.088	0.885±0.077	0.772±0.154	0.844±0.111	0.886±0.082
NP-SIR ($H=10$)	0.859±0.098	0.905±0.072	0.910±0.064	0.813±0.131	0.878±0.086	0.910±0.062
P-KIR	0.957±0.030	0.961±0.031	0.957±0.034	0.611±0.167	0.763±0.253	0.904±0.097
P-SIR ($H=5$)	0.906±0.063	0.899±0.068	0.893±0.072	0.543±0.071	0.662±0.162	0.813±0.126
P-SIR ($H=10$)	0.926±0.047	0.926±0.051	0.915±0.057	0.519±0.041	0.613±0.138	0.822±0.127
CC-KIR	0.804±0.203	0.905±0.085	0.936±0.051	0.537±0.161	0.753±0.244	0.907±0.085
CC-SIR ($H=5$)	0.700±0.207	0.835±0.122	0.884±0.092	0.501±0.117	0.702±0.195	0.842±0.119
CC-SIR ($H=10$)	0.690±0.197	0.831±0.130	0.895±0.081	0.468±0.128	0.645±0.209	0.844±0.130
FC-KIR	0.957±0.032	-	-	-	-	-
FC-SIR ($H=5$)	0.905±0.065	-	-	-	-	-
FC-SIR ($H=10$)	0.927±0.054	-	-	-	-	-
AIPW ($H=5$)	0.871±0.095	0.883±0.079	0.891±0.074	0.512±0.074	0.700±0.238	0.844±0.123
AIPW ($H=10$)	0.901±0.072	0.917±0.061	0.912±0.062	0.546±0.124	0.748±0.220	0.878±0.092
AIPWM ($H=5$)	-	-	-	0.722±0.241	0.835±0.136	0.881±0.086
AIPWM ($H=10$)	-	-	-	0.762±0.201	0.871±0.102	0.906±0.066
mp	0.720±0.037	0.495±0.037	0.265±0.029	0.920±0.014	0.735±0.029	0.455±0.037

model (4.2). These naive procedures are typically regarded as not very efficient in practice.

When only X_1 has missing values, NP-KIR, NP-SIR, P-KIR, P-SIR, and AIPW (equivalently, AIPWM) have comparable performance in both models (4.1) and (4.2). By contrast, when both X_1 and X_2 have missing values, NP-KIR and P-KIR perform much better than AIPW and AIPWM. For example, in Table 1 when both X_1 and X_2 have missing values with $c = 0$, the median trace correlation is 0.921 for NP-KIR and 0.8999 for P-KIR in model (4.1), compared with 0.716 for AIPW with $H = 10$ and 0.848 for AIPWM with $H = 10$.

Table 2. The median and the median absolute deviation of the trace correlation coefficients for models (4.1) and (4.2) with $p = 10$. The missingness follows (4.3).

$n = 200, p = 10$	only X_1 has missing values			both X_1 and X_2 have missing values		
	model (4.1)					
	$c_0 = -1$	$c_0 = 0$	$c_0 = 1$	$c_0 = -1$	$c_0 = 0$	$c_0 = 1$
NP-KIR	0.841±0.079	0.842±0.077	0.850±0.085	0.824±0.082	0.826±0.088	0.837±0.089
NP-SIR ($H = 5$)	0.841±0.055	0.855±0.053	0.864±0.053	0.721±0.125	0.791±0.088	0.833±0.065
NP-SIR ($H = 10$)	0.844±0.059	0.866±0.055	0.875±0.050	0.751±0.115	0.822±0.073	0.842±0.064
P-KIR	0.834±0.118	0.844±0.117	0.859±0.097	0.877±0.105	0.849±0.129	0.867±0.123
P-SIR ($H = 5$)	0.730±0.101	0.734±0.102	0.771±0.106	0.513±0.056	0.597±0.131	0.712±0.127
P-SIR ($H = 10$)	0.732±0.114	0.738±0.119	0.772±0.117	0.543±0.086	0.575±0.117	0.698±0.155
CC-KIR	0.570±0.162	0.746±0.179	0.850±0.112	0.414±0.126	0.542±0.147	0.755±0.176
CC-SIR ($H = 5$)	0.595±0.157	0.727±0.109	0.798±0.092	0.385±0.169	0.559±0.129	0.737±0.132
CC-SIR ($H = 10$)	0.549±0.122	0.708±0.129	0.796±0.099	0.286±0.212	0.521±0.112	0.706±0.160
FC-KIR	0.928±0.050	-	-	-	-	-
FC-SIR ($H = 5$)	0.867±0.059	-	-	-	-	-
FC-SIR ($H = 10$)	0.874±0.057	-	-	-	-	-
AIPW ($H = 5$)	0.846±0.067	0.854±0.064	0.859±0.059	0.507±0.081	0.559±0.118	0.745±0.141
AIPW ($H = 10$)	0.851±0.068	0.862±0.069	0.868±0.063	0.522±0.099	0.666±0.199	0.774±0.154
AIPWM ($H = 5$)	-	-	-	0.598±0.156	0.728±0.148	0.823±0.085
AIPWM ($H = 10$)	-	-	-	0.620±0.181	0.736±0.164	0.823±0.095
mp	0.760±0.029	0.560±0.037	0.345±0.037	0.930±0.014	0.780±0.029	0.540±0.037
	model (4.2)					
	$c_0 = -1$	$c_0 = 0$	$c_0 = 1$	$c_0 = -1$	$c_0 = 0$	$c_0 = 1$
NP-KIR	0.865±0.076	0.878±0.074	0.881±0.071	0.834±0.100	0.849±0.099	0.870±0.080
NP-SIR ($H = 5$)	0.681±0.144	0.743±0.123	0.768±0.111	0.592±0.156	0.702±0.148	0.755±0.127
NP-SIR ($H = 10$)	0.706±0.142	0.774±0.111	0.794±0.107	0.638±0.162	0.735±0.131	0.792±0.108
P-KIR	0.890±0.058	0.899±0.057	0.894±0.060	0.524±0.089	0.597±0.164	0.765±0.178
P-SIR ($H = 5$)	0.804±0.090	0.795±0.099	0.781±0.102	0.494±0.046	0.542±0.086	0.653±0.147
P-SIR ($H = 10$)	0.831±0.080	0.822±0.089	0.805±0.095	0.489±0.040	0.522±0.065	0.647±0.152
CC-KIR	0.527±0.164	0.738±0.201	0.825±0.118	0.365±0.144	0.506±0.138	0.755±0.153
CC-SIR ($H = 5$)	0.484±0.128	0.623±0.156	0.726±0.128	0.278±0.155	0.470±0.127	0.645±0.149
CC-SIR ($H = 10$)	0.459±0.115	0.613±0.155	0.727±0.151	0.204±0.149	0.441±0.113	0.631±0.164
FC-KIR	0.889±0.063	-	-	-	-	-
FC-SIR ($H = 5$)	0.779±0.105	-	-	-	-	-
FC-SIR ($H = 10$)	0.810±0.096	-	-	-	-	-
AIPW ($H = 5$)	0.749±0.117	0.771±0.110	0.774±0.105	0.472±0.101	0.597±0.163	0.720±0.133
AIPW ($H = 10$)	0.775±0.116	0.797±0.104	0.802±0.104	0.497±0.131	0.647±0.162	0.745±0.135
AIPWM ($H = 5$)	-	-	-	0.598±0.160	0.710±0.156	0.750±0.138
AIPWM ($H = 10$)	-	-	-	0.621±0.171	0.738±0.141	0.784±0.115
mp	0.725±0.029	0.490±0.037	0.265±0.029	0.920±0.022	0.740±0.029	0.460±0.029

When both X_1 and X_2 have missing values, we can also see that the median absolute deviation of trace correlations of imputation procedures such as NP-KIR, NP-SIR, P-KIR, and P-SIR are significantly smaller than those of the inverse probability weighted methods such as AIPW and AIPWM, in both models, and that NP-KIR performs the best in most scenarios in terms of median absolute deviation values.

Example 2. In this example, we examined the performance of the different proposals under the MAR assumption (3.1). To be precise, we took the probability

Table 3. The median and the median absolute deviation of the trace correlation coefficients for models (4.1) and (4.2) with $p = 5$. The missingness follows (4.4).

$n = 200, p = 5$	only X_1 has missing values			both X_1 and X_2 have missing values		
	model (4.1)					
	$c_0 = -1$	$c_0 = 0$	$c_0 = 1$	$c_0 = -1$	$c_0 = 0$	$c_0 = 1$
NP-KIR	0.963±0.032	0.967±0.030	0.971±0.026	0.906±0.100	0.937±0.069	0.960±0.037
NP-SIR ($H=5$)	0.925±0.051	0.932±0.045	0.937±0.040	0.811±0.143	0.887±0.080	0.917±0.061
NP-SIR ($H=10$)	0.930±0.048	0.943±0.042	0.944±0.042	0.843±0.134	0.898±0.078	0.930±0.054
P-KIR	0.954±0.035	0.961±0.036	0.971±0.025	0.637±0.199	0.738±0.245	0.901±0.113
P-SIR ($H=5$)	0.928±0.044	0.933±0.042	0.940±0.041	0.553±0.085	0.662±0.154	0.829±0.117
P-SIR ($H=10$)	0.926±0.047	0.943±0.040	0.948±0.040	0.545±0.078	0.630±0.167	0.811±0.139
CC-KIR	0.826±0.191	0.907±0.095	0.951±0.043	0.589±0.171	0.775±0.229	0.905±0.091
CC-SIR ($H=5$)	0.737±0.188	0.846±0.136	0.906±0.075	0.598±0.137	0.709±0.194	0.833±0.139
CC-SIR ($H=10$)	0.745±0.193	0.836±0.149	0.907±0.077	0.597±0.140	0.707±0.194	0.848±0.128
FC-KIR	0.971±0.026	-	-	-	-	-
FC-SIR ($H=5$)	0.947±0.039	-	-	-	-	-
FC-SIR ($H=10$)	0.955±0.037	-	-	-	-	-
AIPW ($H=5$)	0.931±0.047	0.938±0.039	0.938±0.040	0.574±0.112	0.738±0.247	0.881±0.101
AIPW ($H=10$)	0.937±0.046	0.947±0.038	0.946±0.040	0.598±0.149	0.766±0.224	0.911±0.078
AIPWM ($H=5$)	-	-	-	0.786±0.195	0.888±0.093	0.918±0.061
AIPWM ($H=10$)	-	-	-	0.824±0.162	0.898±0.089	0.930±0.053
mp	0.632±0.020	0.454±0.016	0.283±0.014	0.631±0.016	0.456±0.017	0.282±0.014
	model (4.2)					
	$c_0=-1$	$c_0=0$	$c_0=1$	$c_0=-1$	$c_0=0$	$c_0=1$
NP-KIR	0.940±0.045	0.946±0.043	0.948±0.044	0.923±0.063	0.938±0.051	0.940±0.049
NP-SIR ($H=5$)	0.870±0.084	0.888±0.071	0.896±0.072	0.773±0.153	0.831±0.119	0.877±0.086
NP-SIR ($H=10$)	0.889±0.068	0.905±0.059	0.914±0.059	0.822±0.115	0.864±0.089	0.896±0.075
P-KIR	0.930±0.061	0.946±0.043	0.953±0.040	0.673±0.261	0.722±0.301	0.887±0.115
P-SIR ($H=5$)	0.809±0.134	0.856±0.116	0.881±0.092	0.651±0.201	0.683±0.207	0.811±0.137
P-SIR ($H=10$)	0.832±0.143	0.879±0.097	0.900±0.076	0.682±0.247	0.680±0.207	0.806±0.154
CC-KIR	0.696±0.226	0.840±0.140	0.900±0.082	0.509±0.161	0.667±0.217	0.807±0.167
CC-SIR ($H=5$)	0.641±0.192	0.768±0.166	0.844±0.104	0.495±0.155	0.611±0.178	0.775±0.161
CC-SIR ($H=10$)	0.636±0.191	0.799±0.147	0.862±0.100	0.475±0.137	0.604±0.184	0.778±0.161
FC-KIR	0.959±0.029	-	-	-	-	-
FC-SIR ($H=5$)	0.904±0.065	-	-	-	-	-
FC-SIR ($H=10$)	0.934±0.045	-	-	-	-	-
AIPW ($H=5$)	0.863±0.106	0.880±0.086	0.893±0.078	0.519±0.084	0.601±0.161	0.783±0.179
AIPW ($H=10$)	0.891±0.081	0.902±0.072	0.913±0.062	0.564±0.125	0.649±0.217	0.806±0.157
AIPWM ($H=5$)	-	-	-	0.598±0.150	0.749±0.204	0.852±0.112
AIPWM ($H=10$)	-	-	-	0.689±0.227	0.802±0.162	0.881±0.099
mp	0.693±0.014	0.505±0.015	0.316±0.014	0.694±0.014	0.505±0.016	0.315±0.015

π of the missingness given both Y and $(X_2, \ldots, X_p)$ to be

$$\begin{aligned}\pi_1(Y, X_2, \ldots, X_p) &= \text{prob}(\delta_1 = 1 \mid Y, X_2, \cdots, X_p)\\ &= \frac{\exp(c_0 + 0.25Y + 0.5X_2 - X_p)}{1 + \exp(c_0 + 0.25Y + 0.5X_2 - X_p)}. \end{aligned} \tag{4.4}$$

Other settings remain the same as before. The results are summarized in Table 3 for $p = 5$, and Table 4 for $p = 10$.

Similar conclusions can be drawn as in Example 1 for both the full-case and complete-case analyses. The AIPW, NP-KIR, NP-SIR, PKIR, and P-SIR have

Table 4. The median and the median absolute deviation of the trace correlation coefficients for models (4.1) and (4.2) with $p = 10$. The missingness follows (4.4).

$n = 200, p = 10$	only X_1 has missing values			both X_1 and X_2 have missing values		
	model (4.1)					
	$c_0 = -1$	$c_0 = 0$	$c_0 = 1$	$c_0 = -1$	$c_0 = 0$	$c_0 = 1$
NP-KIR	0.907±0.064	0.916±0.061	0.918±0.062	0.796±0.199	0.872±0.118	0.894±0.085
NP-SIR ($H=5$)	0.830±0.069	0.849±0.062	0.856±0.058	0.664±0.137	0.778±0.096	0.823±0.081
NP-SIR ($H=10$)	0.842±0.070	0.856±0.064	0.864±0.059	0.695±0.150	0.779±0.107	0.829±0.079
P-KIR	0.903±0.063	0.905±0.063	0.920±0.060	0.518±0.086	0.565±0.123	0.748±0.223
P-SIR ($H=5$)	0.857±0.058	0.858±0.060	0.863±0.054	0.522±0.079	0.573±0.117	0.703±0.138
P-SIR ($H=10$)	0.862±0.058	0.863±0.061	0.872±0.056	0.543±0.111	0.552±0.101	0.663±0.152
CC-KIR	0.601±0.176	0.780±0.182	0.872±0.103	0.481±0.067	0.570±0.148	0.768±0.180
CC-SIR ($H=5$)	0.553±0.094	0.682±0.139	0.772±0.114	0.498±0.054	0.550±0.096	0.666±0.154
CC-SIR ($H=10$)	0.562±0.104	0.674±0.157	0.773±0.109	0.503±0.063	0.547±0.089	0.669±0.152
FC-KIR	0.923±0.058	-	-	-	-	-
FC-SIR ($H=5$)	0.866±0.059	-	-	-	-	-
FC-SIR ($H=10$)	0.875±0.057	-	-	-	-	-
AIPW ($H=5$)	0.848±0.060	0.857±0.057	0.861±0.054	0.524±0.079	0.641±0.173	0.779±0.129
AIPW ($H=10$)	0.851±0.067	0.864±0.058	0.871±0.056	0.551±0.107	0.670±0.182	0.805±0.107
AIPWM ($H=5$)	-	-	-	0.690±0.167	0.784±0.113	0.830±0.078
AIPWM ($H=10$)	-	-	-	0.726±0.163	0.800±0.112	0.841±0.080
mp	0.630±0.033	0.455±0.037	0.285±0.033	0.795±0.029	0.635±0.037	0.440±0.037
	model (4.2)					
	$c_0=-1$	$c_0=0$	$c_0=1$	$c_0=-1$	$c_0=0$	$c_0=1$
NP-KIR	0.855±0.080	0.866±0.073	0.879±0.070	0.838±0.099	0.841±0.099	0.859±0.078
NP-SIR ($H=5$)	0.703±0.124	0.748±0.102	0.763±0.098	0.595±0.145	0.674±0.151	0.731±0.123
NP-SIR ($H=10$)	0.741±0.108	0.784±0.091	0.799±0.092	0.630±0.155	0.706±0.139	0.767±0.111
P-KIR	0.841±0.111	0.865±0.085	0.892±0.067	0.651±0.255	0.616±0.199	0.756±0.210
P-SIR ($H=5$)	0.646±0.147	0.697±0.154	0.752±0.122	0.595±0.163	0.551±0.117	0.649±0.153
P-SIR ($H=10$)	0.653±0.167	0.732±0.155	0.779±0.125	0.663±0.214	0.575±0.143	0.645±0.173
CC-KIR	0.489±0.154	0.669±0.181	0.795±0.122	0.344±0.146	0.460±0.127	0.666±0.173
CC-SIR ($H=5$)	0.428±0.140	0.559±0.152	0.677±0.143	0.273±0.141	0.414±0.129	0.561±0.147
CC-SIR ($H=10$)	0.408±0.125	0.575±0.166	0.700±0.145	0.232±0.140	0.383±0.120	0.565±0.171
FC-KIR	0.892±0.055	-	-	-	-	-
FC-SIR ($H=5$)	0.771±0.106	-	-	-	-	-
FC-SIR ($H=10$)	0.807±0.097	-	-	-	-	-
AIPW ($H=5$)	0.727±0.126	0.750±0.112	0.765±0.108	0.475±0.104	0.529±0.117	0.652±0.165
AIPW ($H=10$)	0.762±0.125	0.789±0.105	0.802±0.095	0.495±0.111	0.563±0.141	0.682±0.161
AIPWM ($H=5$)	-	-	-	0.540±0.133	0.645±0.161	0.726±0.128
AIPWM ($H=10$)	-	-	-	0.565±0.157	0.675±0.172	0.735±0.143
mp	0.695±0.029	0.505±0.037	0.320±0.029	0.865±0.022	0.705±0.029	0.490±0.037

comparable performance even with the full-case procedures FC-KIR and FC-SIR in model (4.1) when only X_1 has missing values. When both X_1 and X_2 have missing values, the performance of AIPW, P-KIR, and P-SIR deteriorate. The complete-case procedure performs the worst, particularly when both X_1 and X_2 have a large number of missing values. NP-KIR is the winner in most scenarios. We remark here that the nonparametric imputation procedures such as NP-KIR and NP-SIR perform quite well even when the missing probability is misspecified, suggesting that these procedures are robust to the misspecification of missingness

Table 5. The estimated coefficients of the predictors obtained with six proposals.

	X_1	X_2	X_3	X_4	X_5	X_6
NP-KIR	-0.4565	-0.5894	-0.4976	-0.4013	0.1812	-0.0494
NP-SIR	0.0353	-0.6348	-0.5072	-0.4707	0.2132	-0.2671
P-KIR	-0.4565	-0.5894	-0.4976	-0.4013	0.1812	-0.0494
P-SIR	-0.4565	-0.5894	-0.4976	-0.4013	0.1812	-0.0494
AIPW	0.3667	-0.7289	-0.4594	0.3218	0.0680	-0.1218
AIPWM	-0.3785	-0.4652	-0.7781	-0.1222	0.1335	0.0439

mechanism. Conclusions for model (4.2) are similar.

5. An Application

In this section we illustrate our proposals through the horse colic data set available from the Machine Learning Repository at the University of California-Irvine. The objective is to understand whether the lesion of a horse is surgical. Several attributes were collected, but we focus on six factors with continuous measurements as predictors: rectal temperature in degrees Celsius (X_1), the heart rate in beats per minute (X_2), respiratory rate (X_3), the number of red cells by volume in the blood (X_4), total protein (X_5) and the abdomcentesis total protein (X_6) in gms/dL. The response Y is binary, taking value 1 if the lesion of the horse is surgical, and 2 otherwise. Because the response is retrospective, and all cases are either operated upon or autopsied, the response is always known. There are 368 sample points in total, and all six predictors are subject to missingness. Among them, 69 instances in X_1, 26 in X_2, 71 in X_3, 37 in X_4, 43 in X_5, and 235 in X_6 had missing values. The complete data contains only 105 instances. Because the response is binary, the SIR method can identify at most one direction (Cook and Lee (1999)), we take $\dim(\mathcal{S}_{Y|\mathbf{x}}) = 1$.

Six proposals, NP-KIR and NP-SIR, P-KIR and P-SIR, AIPW and AIPWM, were applied to estimate $\mathcal{S}_{Y|\mathbf{x}}$. Because the performance of the slicing estimation is robust to the number of slices, we chose $H = 10$. Our simulations indicate that complete-case analysis, such as CC-KIR or CC-KIR, do not perform well, thus we did not include then in our comparison. The estimated directions of these six proposals are summarized in Table 5. The results obtained with NP-KIR, P-KIR and P-SIR are similar, and slightly different from NP-SIR. However, they are quite different from those obtained with AIPW and AIPWM.

To evaluate the estimation accuracy of these six proposals, we adopted the bootstrap procedure proposed by Ye and Weiss (2003), with $\widehat{\mathcal{S}}_{Y|\mathbf{x}}$ the estimated $\mathcal{S}_{Y|\mathbf{x}}$ using the original data. Then bootstrapped the original data set 1,000 times. For each bootstrap, we applied the six proposals to estimate $\mathcal{S}_{Y|\mathbf{x}}$. With $\widehat{\mathcal{S}}^b_{Y|\mathbf{x}}$ the

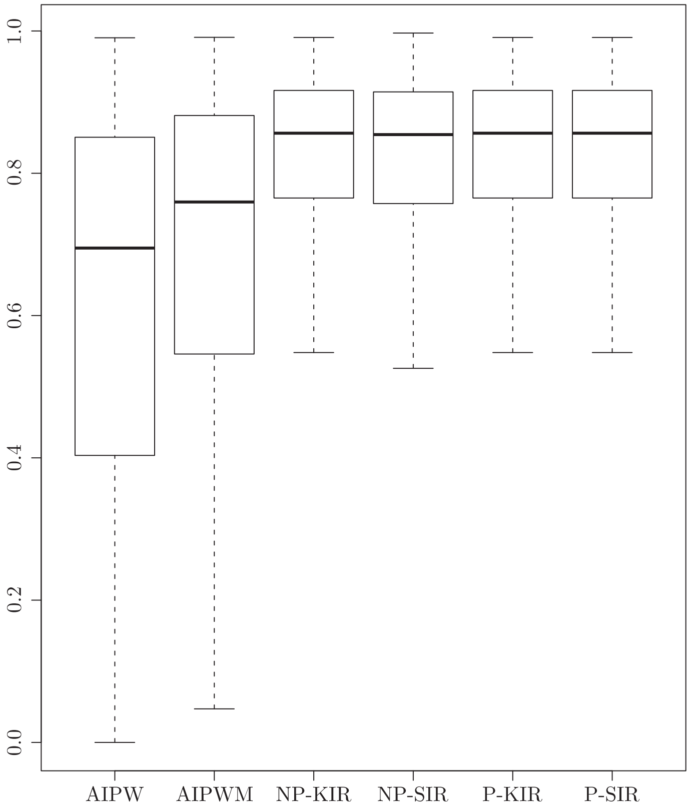

Figure 1. The boxplots of the bootstrap trace correlations for six proposals used in the horse colic data. The vertical axis denotes the trace correlation coefficients.

estimated $\mathcal{S}_{Y|\mathbf{x}}$ using bootstrap data, for $b = 1, \ldots, 1,000$, we calculated the trace correlation coefficient $\{R^2(K)\}^b$ between $\widehat{\mathcal{S}}_{Y|\mathbf{x}}$ and $\widehat{\mathcal{S}}^b_{Y|\mathbf{x}}$ (Ferré (1998)). We took $\{R^2(K)\}^b$ as a measure of variability of the estimator $\widehat{\mathcal{S}}_{Y|\mathbf{x}}$, and preferred proposals with smaller variability, assuming no or minimal bias. The boxplot of the bootstrap trace correlations between $\widehat{\mathcal{S}}_{Y|\mathbf{x}}$ and the bootstrap estimator $\widehat{\mathcal{S}}^b_{Y|\mathbf{x}}$ from these six proposals are presented in Figure 1. It can be seen that NP-KIR, P-KIR, and P-SIR behave comparably, and slightly better than NP-SIR. All our proposals outperformed AIPW and AIPWM significantly.

In terms of predictor contribution, the ratios of these coefficients can be judged as the ratios of standard coefficients from a linear model. In particular, rectal temperature in degrees Celsius (X_1), the heart rate in beats per minute (X_2), respiratory rate (X_3), and the number of red cells by volumes in blood (X_4) play dominant roles in determining whether the lesion of the horse is surgical.

Next we evaluated the predictive power of all six proposals with leave-one-

Table 6. The prediction power obtained with leave-one-out cross-validation.

AIPW	AIPWM	NP-KIR	NP-SIR	P-KIR	P-SIR	Logistic
0.617	0.630	0.685	0.668	0.685	0.685	0.681

out cross-validation. Predictive power is assessed through the proportion of n response variables that are correctly predicted from logistic regression. Because the estimated $\mathcal{S}_{Y|\mathbf{x}}$ is assumed to have one dimension, we fit logistic regression using a single linear combination of the predictors for each proposal. The predictive power using leave-one-out cross-validation is reported in Table 6. It can be seen clearly that NP-KIR, P-KIR and P-SIR perform the best, followed by NP-SIR. AIPW performs the worst, which complies with the bootstrap results reported in Figure 1. In addition, we fit a logistic regression with all six predictors X_is. It yielded the predictive power of 0.681. This indicates that our proposals reduce the dimension effectively, while retaining the regression information of $Y \mid \mathbf{x}$.

6. A Brief Discussion

In this paper we introduce nonparametric and parametric imputation procedures to tackle some general SDR problems when a subset of predictors has missing observations. The nonparametric imputation method inherits the merit of SDR in that it imposes no parametric assumptions on modeling. The associate editor pointed out the MAR assumption (2.2) is stringent. Thus we also consider (3.1). To address the issue of *curse of dimensionality* under (3.1), we introduce a parametric imputation procedure. As such, some of the nonparametric flavor is lost in our SDR estimation with missing predictors. This is the price we pay for a complicated missing mechanism. Our limited simulation studies suggest that the proposed parametric imputation method works well even when simple models, such as linear and logistic regressions, were used.

An anonymous referee pointed out that to check the appropriateness of the two missingness schemes (2.2) and (3.1) is of both practical and theoretical interest. For example, how can one check whether the missingness depends exclusively on the response? This can be formulated as a problem of testing the conditional independence between $\boldsymbol{\delta}$ and $\mathbf{x}$ when Y is given. This is an open and important question even when all observations are complete. Bergsm (2011) proposed to test the conditional independence through partial copulas given complete observations. How to adapt the existing methodologies designed for testing the conditional independence in the complete-data case to the missing-data problems deserves further investigation.

Acknowledgement

Liping Zhu and Tao Wang were supported by National Natural Science Foundation of China (11071077), and Lixing Zhu was supported by the Research Grants Council of Hong Kong (HKU 2030/07P). We are grateful to the Editor, an associate editor, and two anonymous referees for their suggestions which have led to a significant improvement of our early manuscript.

Appendix

We first present some useful lemmas, followed by the proofs of the theorems.

Lemma A.1. Let $H_0(Y) = E\{H(X_l, \delta_l, Y) \mid Y\}$. If *condition (1)* holds and the $(d-1)$-st derivative of $H_0(y)f(y)$ satisfies the local Lipschitz condition,

$$\frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{1}{n-1}\sum_{j\neq i}K_h(Y_j - Y_i)H(X_{lj}, \delta_{lj}, Y_j) - H_0(Y_i)f(Y_i)\Big\}^2 = O_P(h^d). \quad \text{(A.1)}$$

Proof of Lemma A.1 The result is a slightly modified version of Lemma A.1 of Zhu and Zhu (2007) and Zhu and Fang (1996). The proof is omitted.

Lemma A.2. Suppose conditions (1)−(3) are satisfied. Then

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{R}_k(Y_i)R_l(Y_i) - E\{R_k(Y)R_l(Y)\} = \frac{1}{n}\sum_{i=1}^{n}\ell_1(X_{ki}, Y_i, \delta_{ki}) + o_P\left(\frac{1}{\sqrt{n}}\right), \quad \text{(A.2)}$$

where $\ell_1(X_{ki}, Y_i, \delta_{ki}) = \{R_k(Y_i) + \{X_{ki} - R_k(Y_i)\}\,\delta_{ki}\,\{3\pi(Y_i) - 2\}/\pi(Y_i)\}\,R_l(Y_i)$.

Proof of Lemma A.2. We split the proof into two steps.
Step 1 : In this step, we show that, through replacing $\widehat{X}_{kj}$ with $R_k(Y_j)$ in (2.5),

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{R}'_k(Y_i)R_l(Y_i) - E\{R_k(Y)R_l(Y)\} = \frac{1}{n}\sum_{i=1}^{n}\ell_{11}(X_{ki}, Y_i, \delta_{ki}) + o_P\left(\frac{1}{\sqrt{n}}\right), \quad \text{(A.3)}$$

where $\widehat{R}'_k(Y_i) = \widehat{G}'_k(Y_i)/\widehat{f}(Y_i)$, $\widehat{G}'_k(Y_i) = (n-1)^{-1}\sum_{j=1, j\neq i}^{n} K_h(Y_j - Y_i)\{\delta_{kj}X_{kj} + (1-\delta_{kj})R_k(Y_j)\}$, and $\ell_{11}(X_{ki}, Y_i, \delta_{ki}) = \{\delta_{ki}X_{ki} + (1-\delta_{ki})R_k(Y_i)\}\,R_l(Y_i) - E\{R_k(Y)R_l(Y)\}$.

Step 1.1 : Using (2.2), (2), Lemma A.1, and the Cauchy-Schwarz Inequality, we can show without much difficulty that

$$n^{-1}\sum_{i=1}^{n}\Big\{\widehat{G}'_k(Y_i) - R_k(Y_i)f(Y_i)\Big\}\Big\{\widehat{f}(Y_i) - f(Y_i)\Big\} = O_P(h^d),$$

It follows that

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\widehat{R}'_k(Y_i)R_l(Y_i) = & \, n^{-1}\sum_{i=1}^{n}\left\{\widehat{G}'_k(Y_i)-R_k(Y_i)f(Y_i)\right\}\frac{R_l(Y_i)}{f(Y_i)} \\ & -n^{-1}\sum_{i=1}^{n}\left\{\widehat{f}(Y_i)-f(Y_i)\right\}R_k(Y_i)\frac{R_l(Y_i)}{f(Y_i)} \\ & +n^{-1}\sum_{i=1}^{n}R_k(Y_i)R_l(Y_i)+o_P\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Step 1.2 : Following similar arguments used for proving Lemma A.3 of Zhu and Zhu (2007), we can show under (2.2), the Local Lipschitz condition of $R_k(y)f(y)$ and $R_k(y)$ in condition (2), and Lemma A.1, that

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}\left\{\widehat{G}'_k(Y_i)-R_k(Y_i)f(Y_i)\right\}\frac{R_l(Y_i)}{f(Y_i)} \\ & = n^{-1}\sum_{i=1}^{n}\left\{\delta_{ki}X_{ki}+(1-\delta_{ki})R_k(Y_i)\right\}R_l(Y_i)-E\left\{R_k(Y)R_l(Y)\right\}+o_P\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Step 1.3 : Lemma A.2 of Zhu and Zhu (2007) has

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}\left\{\widehat{f}(Y_i)-f(Y_i)\right\}\frac{R_k(Y_i)R_l(Y_i)}{f(Y_i)} \\ & \quad = n^{-1}\sum_{i=1}^{n}R_k(Y_i)R_l(Y_i)-E\left\{R_k(Y)R_l(Y)\right\}+o_P\left(\frac{1}{\sqrt{n}}\right). \end{aligned}$$

Thus, by combining the results in *Steps* 1.1−1.3, (A.3) is proved. To show (A.2), it suffices to show that the difference between the LHS of (A.2) and that of (A.3) admits another asymptotically linear representation.

Step 2 : In this step, we show that

$$n^{-1}\sum_{i=1}^{n}\left\{\widehat{R}'_k(Y_i)-\widehat{R}_k(Y_i)\right\}R_l(Y_i) = n^{-1}\sum_{i=1}^{n}\ell_{12}(X_{ki},\delta_{ki},Y_i)+O_P(h^d), \tag{A.4}$$

where $\ell_{12}(X_{ki},\delta_{ki},Y_i) = 2\delta_{ki}R_l(Y_i)\left\{R_k(Y_i)-X_{ki}\right\}\left\{1-\pi_k(Y_i)\right\}/\pi_k(Y_i)$.

Step 2.1 : Following similar arguments used for proving Lemma A.3 of Zhu and Zhu (2007), we can show that

$$\begin{aligned}
&\frac{1}{n(n-1)}\sum_{i\neq j}\frac{K_h(Y_j-Y_i)(1-\delta_{kj})R_l(Y_i)G_k(Y_j)\widehat{g}_k(Y_j)}{\{f(Y_i)g_k^2(Y_j)\}}\\
&\quad=\frac{2}{n}\sum_{i=1}^{n}\left[\{1-\pi_k(Y_i)\}+(1-\delta_{ki})+\frac{\delta_{ki}\{1-\pi_k(Y_i)\}}{\pi_k(Y_i)}\right]R_l(Y_i)R_k(Y_i)\\
&\qquad-5E\left[\{1-\pi_k(Y_i)\}R_l(Y_i)R_k(Y_i)\right]+O_P(h^d),\ \text{and,}\\
&\frac{1}{n(n-1)}\sum_{i\neq j}\frac{K_h(Y_j-Y_i)(1-\delta_{kj})R_l(Y_i)\widehat{G}_k(Y_j)}{\{f(Y_i)g_k(Y_j)\}}\\
&\quad=\frac{2}{n}\sum_{i=1}^{n}R_l(Y_i)\left\{R_k(Y_i)\{1-\pi_k(Y_i)\}+X_{ki}\delta_{ki}\frac{1-\pi_k(Y_i)}{\pi_k(Y_i)}+R_k(Y_i)(1-\delta_{ki})\right\}\\
&\qquad-5E\left[R_l(Y)R_k(Y)\{1-\pi_k(Y_i)\}\right]+O_P(h^d).
\end{aligned}$$

Therefore, after some straightforward algebraic calculations, we obtain that

$$\begin{aligned}
&\frac{1}{n(n-1)}\sum_{i\neq j}K_h(Y_j-Y_i)(1-\delta_{kj})\left\{R_k(Y_j)-\widehat{X}_{k,j}\right\}\frac{R_l(Y_i)}{f(Y_i)}\\
&\quad=\frac{2}{n}\sum_{i=1}^{n}\frac{\delta_{ki}R_l(Y_i)\{R_k(Y_i)-X_{ki}\}\{1-\pi_k(Y_i)\}}{\pi_k(Y_i)}+O_P(h^d).
\end{aligned}$$

Step 2.2 : The LHS of (A.4) can be expanded as

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{R}_k'(Y_i)-\widehat{R}_k(Y_i)\right\}R_l(Y_i)\\
&\quad=\sum_{i\neq j}\frac{K_h(Y_j-Y_i)(1-\delta_{kj})\left\{R_k(Y_j)-\widehat{X}_{kj}\right\}R_l(Y_i)}{n(n-1)f(Y_i)}\left\{1+\frac{f(Y_i)-\widehat{f}(Y_i)}{\widehat{f}(Y_i)}\right\}.
\end{aligned}$$

The strong consistency of $\widehat{f}(y)$, together with the results in *Step* 2.1, leads to (A.4). Then (A.2) follows by combining the results from these two steps.

Lemma A.3. Suppose (1)−(3) are satisfied. Then

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\delta_{ki}X_{ki}+(1-\delta_{ki})\widehat{X}_{ki}\right\}-E(X_k)=\frac{1}{n}\sum_{i=1}^{n}\ell_{2k}(X_{ki},Y_i,\delta_{ki})+o_P\left(\frac{1}{\sqrt{n}}\right),\tag{A.5}$$

where $\ell_{2k}(X_{ki},Y_i,\delta_{ki})=R_k(Y_i)+\delta_{ki}\{X_{ki}-R_k(Y_i)\}/\pi(Y_i)-E(X_k)$.

Proof of Lemma A.3. Without loss of generality, we assume that $E(X_k) = 0$. Following similar arguments used for proving Lemma A.3 of Zhu and Zhu (2007), we can show that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{(1-\delta_{ki})\left\{\widehat{G}_k(Y_i)-G_k(Y_i)\right\}}{g_k(Y_i)} \\
&\quad=\frac{1}{n}\sum_{i=1}^{n}\frac{X_{ki}\delta_{ki}\left\{1-\pi_k(Y_i)\right\}}{\pi_k(Y_i)}-E\left\{R_k(Y)-r_k(Y)\right\}+O_P\left(\frac{1}{\sqrt{n}}\right), \text{ and} \\
&\frac{1}{n}\sum_{i=1}^{n}\frac{(1-\delta_{ki})G_k(Y_i)\left\{\widehat{g}_k(Y_i)-g_k(Y_i)\right\}}{g_k^2(Y_i)} \\
&\quad=\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_{ki}R_k(Y_i)\left\{1-\pi_k(Y_i)\right\}}{\pi_k(Y_i)}-E\left\{R_k(Y)-r_k(Y)\right\}+O_P\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

Accordingly, the LHS of (A.5) can be expanded as

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left\{\delta_{ki}X_{ki}+(1-\delta_{ki})\widehat{X}_{ki}\right\} \\
&\quad=\frac{1}{n}\sum_{i=1}^{n}\left[R_k(Y_i)+\frac{\delta_{ki}\left\{X_{ki}-R_k(Y_i)\right\}}{\pi(Y_i)}\right]+O_P\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

which completes the proof.

The following lemma is a parallel to Lemma A.3; its proof is skipped.

Lemma A.4. Suppose (1)–(3) are satisfied. Then

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left\{\delta_{ki}\delta_{li}X_{ki}X_{lj}+(1-\delta_{ki}\delta_{li})\widehat{X}_{kl,i}\right\}-E(X_kX_l) \\
&\quad=\frac{1}{n}\sum_{i=1}^{n}\ell_3(X_{ki},X_{li},\delta_{ki},\delta_{li},Y_i)+o_P\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

where $\ell_3(X_{ki},X_{li},\delta_{ki},\delta_{li},Y_i)=R_{kl}(Y_i)+\delta_{ki}\delta_{li}\{X_{ki}X_{li}-R_{kl}(Y_i)\}/\pi_{kl}(Y_i)-E(X_kX_l)$, and $\pi_{kl}(Y)=E\left(\delta_k\delta_l \mid Y\right)$.

Proof of Theorem 1. It suffices to show that $\mathbf{M}_n-\mathbf{M}$ and $\mathbf{\Sigma}_n-\mathbf{\Sigma}$ admit asymptotically linear representations, respectively, by noting that

$$
\begin{aligned}
&\mathbf{\Sigma}_n^{-1}\mathbf{M}_n-\mathbf{\Sigma}^{-1}\mathbf{M} \\
&\quad=\mathbf{\Sigma}^{-1}(\mathbf{\Sigma}-\mathbf{\Sigma}_n)\mathbf{\Sigma}_n^{-1}(\mathbf{M}_n-\mathbf{M})+\mathbf{\Sigma}^{-1}(\mathbf{\Sigma}-\mathbf{\Sigma}_n)\mathbf{\Sigma}_n^{-1}\mathbf{M}+\mathbf{\Sigma}^{-1}(\mathbf{M}_n-\mathbf{M}).
\end{aligned}
$$

By invoking Lemmas A.3 and A.4,

$$\begin{aligned}\mathbf{\Sigma}_n - \mathbf{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\{\ell_3(\mathbf{x}_i, \boldsymbol{\delta}_i, Y_i) - \ell_2(\mathbf{x}_i, \boldsymbol{\delta}_i, Y_i)E(\mathbf{x}^{\mathrm{T}}) - E(\mathbf{x})\ell_2^{\mathrm{T}}(\mathbf{x}_i, \boldsymbol{\delta}_i, Y_i)\} \\ +o_P\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Moreover, by invoking Lemmas A.2 and A.3, we have

$$\begin{aligned}\mathbf{M}_n - \mathbf{M} = \frac{1}{n}\sum_{i=1}^{n}\{\ell_1(\mathbf{x}_i, \boldsymbol{\delta}_i, Y_i) - \ell_2(\mathbf{x}_i, \boldsymbol{\delta}_i, Y_i)E(\mathbf{x})^{\mathrm{T}} - E(\mathbf{x})\ell_2^{\mathrm{T}}(\mathbf{x}_i, \boldsymbol{\delta}_i, Y_i)\} \\ +o_P\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

The subsequent development can proceed, after simple algebraic calculations, to see that $\mathbf{\Sigma}_n^{-1}\mathbf{M}_n - \mathbf{\Sigma}^{-1}\mathbf{M}$ can be expanded as

$$\begin{aligned}n^{-1}\sum_{i=1}^{n}\Big[\mathbf{\Sigma}^{-1}\{\ell_1(\mathbf{x}_i, \boldsymbol{\delta}_i, Y_i) - \ell_2(\mathbf{x}_i, \boldsymbol{\delta}_i, Y_i)E(\mathbf{x}^{\mathrm{T}}) - E(\mathbf{x})\ell_2^{\mathrm{T}}(\mathbf{x}_i, \boldsymbol{\delta}_i, Y_i)\} \\ -\mathbf{\Sigma}^{-1}\{\ell_3(\mathbf{x}_i, \boldsymbol{\delta}_i, Y_i) - \ell_2(\mathbf{x}_i, \boldsymbol{\delta}_i, Y_i)E(\mathbf{x}^{\mathrm{T}}) - E(\mathbf{x})\ell_2^{\mathrm{T}}(\mathbf{x}_i, \boldsymbol{\delta}_i, Y_i)\}\mathbf{\Sigma}^{-1}\mathbf{M}\Big] \\ +o_P\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

The proof is completed by the Lindeberg-Levy Central Limit Theorem.

Proof of Theorem 2. The proof of Theorem 2 is easier than that of Theorem 1 in that we replace the missing measurements with parametric imputation which has a faster convergence rate than the nonparametric imputation in Theorem 1. Thus, following parallel arguments to those for proving Theorem 1, we can prove Theorem 2 without much difficulty. We omit details here.

References

Bergsm, W. P. (2011). Nonparametric testing of conditional independence by means of the partial copula. Available at `http://arxiv.org/PS_cache/arxiv/pdf/1101/1101.4607v1.pdf`

Cheng, P. (1994). Nonparametric estimation of mean functionals with data missing at random. *J. Amer. Statist. Assoc.* **89**, 81-87.

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

Cook, R. D. and Lee, H. (1999). Dimension reduction in binary response regression. *J. Amer. Statist. Assoc.* **94**, 1187-1200.

Cook, R. D. and Weisberg, S. (1991). Discussion of "Sliced inverse regression for dimension reduction". *J. Amer. Statist. Assoc.* **86**, 316-342.

Ferr e, L. (1998). Determingng the dimension in sliced inverse regression and related methods. *J. Amer. Statist. Assoc.* **93**, 132-140.

Hall, P. and Li, K. C. (1993). On almost linearity of low dimensional projection from high dimensional data. *Ann. Statist.* **21**, 867-889.

Li, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *J. Amer. Statist. Assoc.* **86**, 316-342.

Li, K. C. (1992). On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma. *J. Amer. Statist. Assoc.* **87**, 1025-1039.

Li, L. X. and Lu, W. B. (2008). Sufficient dimension reduction with missing predictors. *J. Amer. Statist. Assoc.* **103**, 822-831.

Li, B. and Wang, S. L. (2007). On directional regression for dimension reduction. *J. Amer. Statist. Assoc.* **102**, 997-1008.

Li, B., Zha, H. and Chiaromonte, F. (2005). Contour regression: a general approach to dimension reduction. *Ann. Statist.* **33**, 1580-1616.

Little, R. J. A. (1992), Regression with missing X's: a review. *J. Amer. Statist. Assoc.* **87**, 1227-1237.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data.* 2nd edition. Wiley, New Jersey.

Robins, J. M., Rotnitzky, A. and Zhao, L. (1994). Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.* **89**, 846-866.

Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581-592.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* Wiley, New York.

Wang, D. and Chen, S. X. (2009). Empirical likelihood for estimating equation with missing values. *Ann. Statist.* **37**, 490-517.

Wang, Q. and Rao, J. N. K. (2002). Empirical likelihood-based inference under imputation for missing response data. *Ann. Statist.* **30**, 896-924.

Xia, Y. C. (2007). A constructive approach to the estimation of dimension reduction directions. *Ann. Statist.* **35**, 2654-2690

Xia, Y. C., Tong, H., Li, W. K. and Zhu, L. X. (2002). An adaptive estimation of optimal regression subspace. *J. Roy. Statist. Soc. Ser. B* **64**, 363-410.

Ye, Z. and Weiss, R. E. (2003). Using the bootstrap to select one of a new class of dimension reduction methods. *J. Amer. Statist. Assoc.* **98**, 968-979.

Yates, F. (1933). The analysis of replicated experiments when the field results are incomplete. *Emporium J. Experimental Agriculture* **1**, 129-142.

Zhu, L. X. and Fang, K. T. (1996). Asymptotics for the kernel estimators of sliced inverse regression. *Ann. Statist.* **24**, 1053-1067.

Zhu, L. X. and Ng, K. W. (1995). Asymptotics of sliced inverse regression. *Statist. Sinica* **5**, 727-736.

Zhu, L. P. and Zhu, L. X. (2007). On kernel method for sliced average variance estimation. *J. Multivariate Anal.* **98**, 970-991.

School of Statistics and Management and the Key Laboratory of Mathematical Economics, Ministry of Education, Shanghai University of Finance and Economics, Shanghai 200433, P. R. China.

E-mail: zhu.liping@mail.shufe.edu.cn

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, P. R. China.

E-mail: 10466029@hkbu.edu.hk

Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, P. R. China.

E-mail: lzhu@hkbu.edu.hk

(Received July 2009; accepted November 2011)