Journal of Multivariate Analysis 174 (2019) 104542

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Dimension reduction estimation for central mean subspace with missing multivariate response

Guo-Liang Fan [a], Hong-Xia Xu [b], Han-Ying Liang [c,*]

[a] School of Economics and Management, Shanghai Maritime University, Shanghai, 201306, PR China
[b] Department of Mathematics, Shanghai Maritime University, Shanghai, 201306, PR China
[c] School of Mathematical Sciences, Tongji University, Shanghai 200092, PR China

A R T I C L E   I N F O

Article history:
Received 23 March 2019
Received in revised form 21 August 2019
Accepted 21 August 2019
Available online 29 August 2019

AMS 2010 subject classifications:
primary 62H12
secondary 62F12

Keywords:
Central mean subspace
High dimensionality
Missing data
Multivariate response
Sufficient dimension reduction

A B S T R A C T

Multivariate response data often arise in practice and they are frequently subject to missingness. Under this circumstance, the standard sufficient dimension reduction (SDR) methods cannot be used directly. To reduce the dimension and estimate the central mean subspace, a profile least squares estimation method is proposed based on an inverse probability weighted technique. The profile least squares method does not need any distributional assumptions on the covariates and hence differs from existing SDR methods. The resulting estimator of the central mean subspace is proved to be asymptotically normal and root n consistent under some mild conditions. The structural dimension is determined by a BIC-type criterion and the consistency of its estimator is established. Comprehensive simulations and a real data analysis show that the proposed method works promisingly.

1. Introduction

With rapid advance of technology, observations for a large number of variables can be easily collected. High-dimensional regression analysis is one of the most popular tools that help us to gain insight into relationships between two sets of high dimensional variables. Suppose that $\mathbf{y} = (Y_1, \ldots, Y_q)^\top \in \mathbb{R}^q$ is multivariate response and $\mathbf{x} = (X_1, \ldots, X_p)^\top \in \mathbb{R}^p$ is a covariate vector. The general object of interest is to focus on the conditional mean function $E(\mathbf{y}|\mathbf{x})$. However, nonparametric estimation of $E(\mathbf{y}|\mathbf{x})$ involves a high-dimensional predictor $\mathbf{x}$, which suffers from the "curse of dimensionality". Sufficient dimension reduction (SDR) has been proposed to reduce the dimension of $\mathbf{x}$ while preserving its information in $E(\mathbf{y}|\mathbf{x})$. The aim is to search a matrix $\boldsymbol{\beta} \in \mathbb{R}^{p\times d}$ such that $E(\mathbf{y}|\mathbf{x}) = E(\mathbf{y}|\boldsymbol{\beta}^\top\mathbf{x})$, where $E(\mathbf{y}|\boldsymbol{\beta}^\top\mathbf{x})$ can be estimated efficiently if d is small. Noting that $\boldsymbol{\beta}$ is not identifiable, Cook and Li [5] defined the central mean subspace span$(\boldsymbol{\beta})$ as the column space of $\boldsymbol{\beta}$ with the smallest column dimension such that $E(\mathbf{y}|\mathbf{x}) = E(\mathbf{y}|\boldsymbol{\beta}^\top\mathbf{x})$. The column dimension of span$(\boldsymbol{\beta})$, denoted by d_0, is called the structural dimension. When the data are fully observed and $\mathbf{y}$ is univariate ($q = 1$), since the work of Li [12] who proposed the sliced inverse regression method, there have been many approaches developed to estimate span$(\boldsymbol{\beta})$. Specifically, Li [13] proposed the principal Hessian directions method to recover span$(\boldsymbol{\beta})$ through the eigenvectors. Discussions on the principal Hessian directions method were given by Cheng and Zhu [3], Cook [4] and Cook and Li [5], etc. Assuming that $E(\mathbf{y}|\boldsymbol{\beta}^\top\mathbf{x})$ is a sufficiently smooth function of $\boldsymbol{\beta}^\top\mathbf{x}$, Xia et al. [22] suggested a minimum average variance estimation method to recover span$(\boldsymbol{\beta})$. Ma and Zhu [15,17] proposed semiparametric methods

* Corresponding author.
E-mail address: hyliang@tongji.edu.cn (H.-Y. Liang).

https://doi.org/10.1016/j.jmva.2019.104542
0047-259X/

to estimate span($\boldsymbol{\beta}$) when either $E(\mathbf{y}|\boldsymbol{\beta}^\top\mathbf{x})$ or $E(\mathbf{x}|\boldsymbol{\beta}^\top\mathbf{x})$ is assumed to be smooth. When the data are fully observed and $\mathbf{y}$ is multivariate, Zhu and Zhong [28] proposed a profile least squares approach to perform estimation and inference on the central mean subspace.

In practice, missing data are a prevailing problem in any data analyses. A variable is considered missing if the value of the variable is not observed in an observation. In most analyses in the medical literature, the most common way of dealing with missing response data is to just omit those participants who have any missing data among its variables. Such an analysis is called a complete case (CC) analysis. The CC analysis is quite popular, because it is the default analysis for most standard statistical softwares. However, in many situations, the CC analysis is not an appropriate way to proceed and it may decrease the power of the analysis by decreasing the effective sample size. Data are said to be missing at random (MAR) if, given the observed data, the failure to observe a value does not depend on the data that are unobserved. For example, in cancer clinical trials, information on the size of a primary tumor is often missing, and the size of the primary tumor may depend on the type of the primary tumor, which is often fully observed. If the probability of primary tumor size being missing only depends on the type of primary tumor, then the missingness is considered to be MAR. Many approaches have been developed on how to handle missing values in multivariate analysis. For example, Liao et al. [14] investigated existing imputation approaches for phenomic data, proposed a novel "imputability" concept with a quantitative imputability measure to characterize whether a missing value is imputable or not, and proposed a self-training selection scheme to select the best imputation approach; Zhao and Long [25] investigated several approaches of using regularized regression and Bayesian lasso regression to impute the missing values for high-dimensional data; Wei et al. [21] developed a computationally efficient alternating expectation conditional maximization algorithm for parameter estimation of the generalized hyperbolic factor analyzers model in the presence of the missing values as well as heavy-tailed and/or asymmetric clusters. In regression analysis, when the response is univariate ($q = 1$), some interesting problems have been investigated with the MAR setting. For example, Hu et al. [11] focused on the estimation of the marginal mean response when the response is MAR and covariates are available; Yang and Wang [24] developed a dimension-reduction based kernel imputation method for the sliced inverse regression; Deng and Wang [6] developed dimension reduction estimating methods for probability density with response MAR when covariables are present. Other related works can be found in Guo et al. [9], who proposed selection probability assisted recovery and complete case assisted recovery methods, and Zhu et al. [27], who proposed a parametric imputation procedure based on the sliced inverse regression.

In this article, we address the dimension reduction estimation for span($\boldsymbol{\beta}$) when the responses $\mathbf{y}$ are MAR and the predictors $\mathbf{x}$ are present. Specifically, the random sample $\{(\mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\delta}_i)\}$ for $i \in \{1, \ldots, n\}$ of the incomplete data comes from $\{\mathbf{x}, \mathbf{y}, \boldsymbol{\delta}\}$, where all the $\mathbf{x}$ are observed, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_q)^\top$, and $\delta_k = 1$ if Y_k is observed and $\delta_k = 0$ otherwise for $k \in \{1, \ldots, q\}$. With incomplete observations, we consider the estimation of the central mean subspace span($\boldsymbol{\beta}$) under MAR, which implies that $\boldsymbol{\delta}$ and $\mathbf{y}$ are conditionally independent given $\mathbf{x}$, or equivalently, for $k \in \{1, \ldots, q\}$,

$$P(\delta_k = 1|\mathbf{x}, Y_k) = P(\delta_k = 1|\mathbf{x}) \stackrel{\text{def}}{:=} \pi_k(\mathbf{x}),$$

where $\pi_k(\cdot)$ is called a selection probability function. In general, to understand how the conditional mean functions of $\mathbf{y}$ vary with $\mathbf{x}$, we let $E(\mathbf{y}|\mathbf{x}) = \mathbf{m}(\boldsymbol{\beta}^\top\mathbf{x})$, or equivalently, $E(Y_l|\mathbf{x}) = m_l(\boldsymbol{\beta}^\top\mathbf{x})$, where $\mathbf{m}(\boldsymbol{\beta}^\top\mathbf{x}) = \{m_1(\boldsymbol{\beta}^\top\mathbf{x}), \ldots, m_q(\boldsymbol{\beta}^\top\mathbf{x})\}^\top$, $\boldsymbol{\beta}$ is a $p \times d_0$ matrix with an unknown d_0. All $m_k(\cdot)$, $\boldsymbol{\beta}$ and d_0 have to be estimated based on the observed data, and all $m_k(\cdot)$ share an identical $\boldsymbol{\beta}$ to ensure that the span($\boldsymbol{\beta}$) is identifiable (see [28]).

The rest of this article is organized as follows. In Section 2, we propose a profile least squares method to estimate the central mean subspace based on the inverse probability weighted (IPW) approach, and establish asymptotic properties for the resultant estimators. We present the performance of the proposed methods via a comprehensive simulation study and a real data analysis in Section 3. The concluding remarks are given in Section 4. Some additional simulations and all technical proofs are deferred to the Appendix.

2. Methodology

In this section, we seek a $\boldsymbol{\beta}$ with the minimal column dimension d_0 such that $E(\mathbf{y}|\mathbf{x}) = E(\mathbf{y}|\boldsymbol{\beta}^\top\mathbf{x})$. That is, in the sense of the SDR, replace $\mathbf{x}$ with $\boldsymbol{\beta}^\top\mathbf{x}$, which is sufficient to describe how $E(\mathbf{y}|\mathbf{x})$ varies with $\mathbf{x}$. Using the parameterization design in Fan et al. [8], we let $\boldsymbol{\beta} = (\mathbf{I}_{d_0\times d_0}, \boldsymbol{\beta}_{-d_0}^\top)^\top$, where $\mathbf{I}_{d_0\times d_0}$ is a $d_0 \times d_0$ identity matrix and $\boldsymbol{\beta}_{-d_0}$ is a $(p - d_0) \times d_0$ matrix composed of the last $(p - d_0)$ rows of $\boldsymbol{\beta}$. The space of all such $\boldsymbol{\beta}$ matrices forms a one-to-one mapping to the dimension reduction space. Thus, the problem of estimating span($\boldsymbol{\beta}$) is equivalent to a problem of estimating $\boldsymbol{\beta}_{-d_0}$. This parameterization implies that the first d_0 covariates of $\mathbf{x}$ contribute to $\mathbf{y}$. Otherwise one can always rotate the order of the entries in $\mathbf{x}$ to guarantee that the first d_0 components of $\mathbf{x}$ are useful.

Since the structural dimension d_0 of span($\boldsymbol{\beta}$) is unknown in advance, we will demonstrate our proposed estimation procedure with a working dimension d. A profile least squares method for estimating $\boldsymbol{\beta}$, or equivalently, $\boldsymbol{\beta}_{-d}$ is proposed in the sequel. Let $\mathbf{x}_d = (X_1, \ldots, X_d)^\top$ and $\mathbf{x}_{-d} = (X_{d+1}, \ldots, X_p)^\top$. Hence, $\boldsymbol{\beta}^\top\mathbf{x} = \mathbf{x}_d + \boldsymbol{\beta}_{-d}^\top\mathbf{x}_{-d}$ and $\mathbf{m}(\boldsymbol{\beta}^\top\mathbf{x}) = \mathbf{m}(\mathbf{x}_d + \boldsymbol{\beta}_{-d}^\top\mathbf{x}_{-d})$.

In practice, the selection probability functions $\pi_k(\cdot)$ are usually unknown and need to be estimated. The Nadaraya–Watson estimation approach is often used to estimate $\pi_k(\cdot)$. However, a fully nonparametric estimation is usually unattractive because the estimation precision decreases rapidly as the dimension of $\mathbf{x}$ increases, i.e., fully nonparametric estimation may suffer from the curse of dimensionality (see [10]). In this situation, a parametric approach might be more applicable to estimate $\pi_k(\cdot)$, that is, we assume $\pi_k(\mathbf{x}) = \pi_k(\mathbf{x}, \boldsymbol{\gamma}_k)$. The logistic regression, based on $\{(\mathbf{x}_i, \mathbf{y}_i, \boldsymbol{\delta}_i)\}$ for $i \in \{1, \dots, n\}$, can yield consistent estimators of the regression coefficients in the model, provided that $\pi_k(\mathbf{x}_i, \boldsymbol{\gamma}_k)$ are correctly specified. Specifically, suppose that for $k \in \{1, \dots, q\}$,

$$\pi_k(\mathbf{x}_i, \boldsymbol{\gamma}_k) = \frac{1}{1 + \exp(-\boldsymbol{\gamma}_k^\top \mathbf{x}_i)}, \tag{2.1}$$

where $\boldsymbol{\gamma}_k = (\gamma_{k1}, \dots, \gamma_{kp})^\top$ are unknown parameter vectors. Hence, the estimators of the selection probability functions are given by $\pi_k(\mathbf{x}, \widehat{\boldsymbol{\gamma}}_k)$, where $\widehat{\boldsymbol{\gamma}}_k$ are the maximum likelihood estimators (MLEs) of $\boldsymbol{\gamma}_k$. We adopt the following least squares criterion to estimate $\boldsymbol{\beta}_{-d}$ based on the IPW method,

$$\widehat{\boldsymbol{\beta}}_{-d} := \underset{\mathbf{b} \in \mathbb{R}^{(p-d)\times d}}{\operatorname{argmin}} \sum_{i=1}^{n} \Big\{\mathbf{y}_i - \widehat{\mathbf{m}}(\mathbf{x}_{d,i}^\top + \mathbf{b}^\top \mathbf{x}_{-d,i})\Big\}^\top \widehat{\mathbf{W}}_i \Big\{\mathbf{y}_i - \widehat{\mathbf{m}}(\mathbf{x}_{d,i}^\top + \mathbf{b}^\top \mathbf{x}_{-d,i})\Big\}, \tag{2.2}$$

where $\widehat{\mathbf{W}}_i = \operatorname{diag}\Big\{\delta_{i1}/\pi_1(\mathbf{x}_i, \widehat{\gamma}_1), \dots, \delta_{iq}/\pi_q(\mathbf{x}_i, \widehat{\gamma}_q)\Big\}$,

$$\begin{aligned}
\widehat{\mathbf{m}}(\mathbf{x}_{d,i}^\top + \mathbf{b}^\top \mathbf{x}_{-d,i}) &= \big\{\widehat{m}_1(\mathbf{x}_{d,i}^\top + \mathbf{b}^\top \mathbf{x}_{-d,i}), \dots, \widehat{m}_q(\mathbf{x}_{d,i}^\top + \mathbf{b}^\top \mathbf{x}_{-d,i})\big\}^\top, \\
\widehat{m}_k(\mathbf{x}_{d,i}^\top + \mathbf{b}^\top \mathbf{x}_{-d,i}) &= \frac{\sum_{j=1, j\neq i}^{n} \delta_{jk} K_h(\mathbf{x}_{d,j}^\top + \mathbf{b}^\top \mathbf{x}_{-d,j} - \mathbf{x}_{d,i}^\top - \mathbf{b}^\top \mathbf{x}_{-d,i}) Y_{jk}}{\sum_{j=1, j\neq i}^{n} \delta_{jk} K_h(\mathbf{x}_{d,j}^\top + \mathbf{b}^\top \mathbf{x}_{-d,j} - \mathbf{x}_{d,i}^\top - \mathbf{b}^\top \mathbf{x}_{-d,i})},
\end{aligned} \tag{2.3}$$

for $k \in \{1, \dots, q\}$, $K_h(\cdot) = K(\cdot/h)/h^d$ is a product of d univariate kernel functions and $0 < h := h_n \to 0$ is the bandwidth.

In order to formulate the main results in this paper, the following regularity conditions are listed.

(C1) The density function $f(\boldsymbol{\beta}^\top \mathbf{x})$ of $\boldsymbol{\beta}^\top \mathbf{x}$ is locally Lipschitz continuous and bounded away from zero and infinity. In addition, $\mathbf{m}(\boldsymbol{\beta}^\top \mathbf{x})$ and $E(\mathbf{x}|\boldsymbol{\beta}^\top \mathbf{x})$ are locally Lipschitz continuous.

(C2) The univariate kernel function $K(\cdot)$ is symmetric, has a compact support and derivatives up to order s. In addition,

$$\int K(u)du = 1; \quad \int u^k K(u)du = 0, \ k \in \{1, \dots, s-1\}; \ 0 \neq \int u^s K(u)du < \infty.$$

The d-dimensional kernel is a product of d univariate kernels. We omit the notation of K here when it is sufficiently clear from the context.

(C3) The bandwidth $h = O(n^{-\delta})$ for $(4s)^{-1} < \delta < (2d)^{-1}$.

(C4) All the moments $E[\{\mathbf{m}(\boldsymbol{\beta}^\top \mathbf{x})\}^\top \{\mathbf{m}(\boldsymbol{\beta}^\top \mathbf{x})\}]$, $E(\mathbf{x}^\top \mathbf{x})^\tau$, $E(\mathbf{y}^\top \mathbf{y})$ and $E\big[\{\mathbf{m}^{(1)}(\boldsymbol{\beta}^\top \mathbf{x})\}^\top \{\mathbf{m}^{(1)}(\boldsymbol{\beta}^\top \mathbf{x})\}\big]$ exist for some $\tau > 1$, where $\mathbf{m}^{(1)}(\boldsymbol{\beta}^\top \mathbf{x})$ is the first derivative of $\mathbf{m}(\boldsymbol{\beta}^\top \mathbf{x})$ with respect to $\boldsymbol{\beta}^\top \mathbf{x}$, and is continuous.

(C5) The selection probability functions $\pi_k(\mathbf{x})$ have a bounded continuous second derivatives almost surely and $\inf_{\mathbf{x}} \pi_k(\mathbf{x}) > 0$ for $k \in \{1, \dots, q\}$.

(C6) The MLEs $\widehat{\boldsymbol{\gamma}}_k$ of $\boldsymbol{\gamma}_k$, $k \in \{1, \dots, q\}$, in model (2.1) are root-n consistent.

Remark 2.1. These conditions are generally regarded as mild. In particular, the smoothness condition (C1) imposes on the mean and density functions and allows us to implement local smoothers such as kernel and local polynomial regressions (see [7]). Condition (C2) states that an sth order kernel function is used. Condition (C3) is typical for deriving a convergence rate when the nonparametric estimation is employed. We assume the moment conditions in (C4) to establish the asymptotic normality. Similar conditions are also assumed in [15,16]. Condition (C5) is a necessary condition in missing data analysis and is also assumed in [10,19], among others. Condition (C6), which can be satisfied easily under mild conditions (see Section 10.6.2 of [2] and [20]), is employed to simplify the proof procedure.

The following Theorem 2.1 indicates that the estimator $\widehat{\boldsymbol{\beta}}_{-d}$ of $\boldsymbol{\beta}_{-d}$ is root-n consistent and asymptotically normal. Define $\mathbf{A}(\mathbf{x}) = \operatorname{diag}\Big\{1/\sqrt{\pi_1(\mathbf{x})}, \dots, 1/\sqrt{\pi_q(\mathbf{x})}\Big\}$, $\widetilde{\mathbf{x}}_{-d} = \mathbf{x}_{-d} - E(\mathbf{x}_{-d}|\boldsymbol{\beta}^\top \mathbf{x})$,

$$\boldsymbol{\Omega}_1 = E\Big[\big\{\mathbf{m}^{(1)^\top}(\boldsymbol{\beta}^\top \mathbf{x}) \otimes \widetilde{\mathbf{x}}_{-d}\big\}\big\{\mathbf{m}^{(1)}(\boldsymbol{\beta}^\top \mathbf{x}) \otimes \widetilde{\mathbf{x}}_{-d}^\top\big\}\Big], \quad \boldsymbol{\Sigma} = \operatorname{cov}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top), \ \ \boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{m}(\boldsymbol{\beta}^\top \mathbf{x})$$

and $\boldsymbol{\Omega}_2 = E\Big[\{\mathbf{m}^{(1)^\top}(\boldsymbol{\beta}^\top \mathbf{x}) \otimes \widetilde{\mathbf{x}}_{-d}\}\mathbf{A}(\mathbf{x})\boldsymbol{\Sigma}\mathbf{A}(\mathbf{x})\{\mathbf{m}^{(1)}(\boldsymbol{\beta}^\top \mathbf{x}) \otimes \widetilde{\mathbf{x}}_{-d}^\top\}\Big]$, where $\otimes$ denotes the Kronecker product. Further, for a matrix $\mathbf{D}$, $\operatorname{vec}(\mathbf{D})$ denotes the vector of the stacked columns of $\mathbf{D}$, starting with the first one.

Theorem 2.1. *Suppose that (C1)–(C6) hold. Then* $n^{1/2}\big\{\operatorname{vec}(\widehat{\boldsymbol{\beta}}_{-d}) - \operatorname{vec}(\boldsymbol{\beta}_{-d})\big\} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_1^{-1}\boldsymbol{\Omega}_2\boldsymbol{\Omega}_1^{-1})$.

The asymptotic distribution of $\widehat{\boldsymbol{\beta}}_{-d}$ in Theorem 2.1 is based on the assumption that $\pi_k(\mathbf{x}_i)$ is logistic regression function. Actually, similar result can be derived if $\pi_k(\mathbf{x}_i)$ is any purely parametric function. In order to make statistical inference on $\boldsymbol{\beta}_{-d}$ based on Theorem 2.1, a consistent estimator for the asymptotic covariance matrix is given in the following Theorem.

Let $\widehat{\boldsymbol{\beta}} = (\mathbf{I}_{d\times d}, \widehat{\boldsymbol{\beta}}_{-d}^\top)^\top$, $\widehat{\Sigma} = n^{-1}\sum_{i=1}^n \{\mathbf{y}_i - \widehat{\mathbf{m}}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i)\}\widehat{\mathbf{W}}_i\{\mathbf{y}_i - \widehat{\mathbf{m}}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i)\}^\top$,

$$\widehat{\widetilde{\mathbf{x}}}_{-d,ik} = \mathbf{x}_{-d,i} - \sum_{j=1,j\neq i}^{n} \delta_{jk}K_h(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_j - \widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i)\mathbf{x}_{-d,i} \Big/ \sum_{j=1,j\neq i}^{n} \delta_{jk}K_h(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_j - \widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i),$$

$$\widehat{\Omega}_1 = n^{-1}\sum_{i=1}^{n} \mathbf{B}_i\widehat{\mathbf{W}}_i\mathbf{B}_i^\top \quad \text{and} \quad \widehat{\Omega}_2 = n^{-1}\sum_{i=1}^{n} \mathbf{B}_i\widehat{\mathbf{W}}_i\widehat{\Sigma}\widehat{\mathbf{W}}_i\mathbf{B}_i^\top, \tag{2.4}$$

where $\mathbf{B}_i = \left\{\widehat{\mathbf{m}}_1^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i) \otimes \widehat{\widetilde{\mathbf{x}}}_{-d,i1}, \ldots, \widehat{\mathbf{m}}_q^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i) \otimes \widehat{\widetilde{\mathbf{x}}}_{-d,iq}\right\}$ is a $(p-d)d \times q$ matrix.

Theorem 2.2. *Suppose that (C1)–(C6) hold. Then* $\widehat{\Omega}_1 \xrightarrow{p} \Omega_1$, $\widehat{\Omega}_2 \xrightarrow{p} \Omega_2$, *and hence* $\widehat{\Omega}_1^{-1}\widehat{\Omega}_2\widehat{\Omega}_1^{-1} \xrightarrow{p} \Omega_1^{-1}\Omega_2\Omega_1^{-1}$, *where* $\widehat{\Omega}_1$ *and* $\widehat{\Omega}_2$ *are given in (2.4).*

Testing whether $\mathbf{x}_i$ is important in predicting $\mathbf{y}$ amounts to testing whether all components of the ith row of $\boldsymbol{\beta}$ are simultaneously zero. In a general context, we are interested in the following hypothesis testing problem:

$$H_0 : \mathbf{Q}\boldsymbol{\beta}_{-d} = \mathbf{q}_0 \quad \text{versus} \quad H_1 : \mathbf{Q}\boldsymbol{\beta}_{-d} \neq \mathbf{q}_0,$$

where $\boldsymbol{\beta}_{-d}$ is a $(p-d)\times d$ matrix composed of the last $(p-d)$ rows of $\boldsymbol{\beta} = (\mathbf{I}_{d\times d}, \boldsymbol{\beta}_{-d}^\top)^\top$, $\mathbf{Q}$ is a user-specified $q_0 \times (p-d)$ matrix and $\mathbf{q}_0$ is another user-specified $q_0 \times d$ matrix. A Wald chi-square test statistic is defined as,

$$T_n = n\left\{(\mathbf{I}_{d\times d}\otimes \mathbf{Q})\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{-d}) - \mathrm{vec}(\mathbf{q}_0)\right\}^\top \left\{(\mathbf{I}_{d\times d}\otimes \mathbf{Q})\widehat{\Omega}_1^{-1}\widehat{\Omega}_2\widehat{\Omega}_1^{-1}(\mathbf{I}_{d\times d}\otimes \mathbf{Q}^\top)\right\}^{-1}\left\{(\mathbf{I}_{d\times d}\otimes \mathbf{Q})\mathrm{vec}(\widehat{\boldsymbol{\beta}}_{-d}) - \mathrm{vec}(\mathbf{q}_0)\right\}.$$

From Theorem 2.1, the following corollary can be established directly.

Corollary 2.1. *Under (C1)–(C6) and* H_0, *we have* $T_n \xrightarrow{d} \chi^2(q_0 d)$, *where* $\chi^2(q_0 d)$ *stands for the central chi-square distribution with* $q_0 d$ *degrees of freedom.*

Another important component in the dimension reduction is to decide the structural dimension d of the span$(\boldsymbol{\beta})$ based on the incomplete data. For the sliced inverse regression method, Li [12] proposed a sequential chi-squared test procedure and Bura and Cook [1] suggested a general weighted chis-squared sequential test to determine the dimension. But the retained dimension in these methods relies heavily on the significance level. In this paper, based on the profile least squares technique, we suggest a BIC-type criterion (see [23,26]) to estimate the structural dimension of the span$(\boldsymbol{\beta})$. The procedure is easy to implement and the consistency of the estimator can be established. Specifically, for a working dimension d, we define

$$\mathcal{L}(d) = \sum_{i=1}^{n}\left\{\mathbf{y}_i - \widehat{\mathbf{m}}(\widehat{\boldsymbol{\beta}}_d^\top \mathbf{x}_i)\right\}^\top \widehat{\mathbf{W}}_i\left\{\mathbf{y}_i - \widehat{\mathbf{m}}(\widehat{\boldsymbol{\beta}}_d^\top \mathbf{x}_i)\right\} + pd\lambda_n\left\{\sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}})^\top(\mathbf{y}_i - \bar{\mathbf{y}})\right\}^{1/2},$$

where $\widehat{\boldsymbol{\beta}}_d = (\mathbf{I}_{d\times d}, \widehat{\boldsymbol{\beta}}_{-d})^\top$, $\bar{\mathbf{y}} = n^{-1}\sum_{i=1}^n \mathbf{y}_i$, the second term is the penalty term and λ_n is a penalty constant. The estimated structural dimension is then given by

$$\widehat{d} = \underset{1\le d\le p}{\mathrm{argmin}}\, \mathcal{L}(d). \tag{2.5}$$

Theorem 2.3. *Suppose that (C1)–(C6) hold, and that* $n^{-1/2}\lambda_n \to 0$ *and* $\lambda_n/\ln n \to \infty$. *If the sth derivative of* $\mathbf{m}(\cdot)$ *is bounded, then* $\Pr(\widehat{d} = d_0) \to 1$.

Remark 2.2.

(a) It can be concluded from Theorem 2.3 that the BIC-type criterion can select the true structural dimensional of the span$(\boldsymbol{\beta})$ consistently. Although the penalty term λ_n is allowed to vary from $\ln n$ to $n^{1/2}$, the BIC-type criterion may overestimate d if λ_n is too small and underestimate it if λ_n is too large. How to choose an optimal penalty λ_n in a data-driven manner is a challenging work. We have tried several values of $\lambda_n = \alpha n^{\kappa}$ for different α and κ, and found that $(\alpha, \kappa) = (1, 0.05)$ performs better overall than some other choices.
(b) We now outline the algorithm for estimating $\boldsymbol{\beta}$. Firstly, we get the maximum likelihood estimators of $\boldsymbol{\gamma}_k$ and start with a working dimension d and an initial value of $\boldsymbol{\beta}$.

   (1) Estimate $\mathbf{m}$ based on (2.3) for a given $\boldsymbol{\beta}$.
   (2) Estimate $\boldsymbol{\beta}$ based on (2.2) for given $\mathbf{m}$.
   (3) Repeat the above two steps until convergence. The derived estimator, denoted by $\widehat{\boldsymbol{\beta}}_d = (\mathbf{I}_{d\times d}, \widehat{\boldsymbol{\beta}}_{-d}^\top)^\top$, is referred to as the profile least squares estimator for a working dimension d.
   (4) Varying the working dimension d from 1 through p and repeat the above three steps. The estimated structure dimension $\widehat{d}$ is given in (2.5).

Table 1
The choices of $\boldsymbol{\gamma}_k$, $k \in \{1, 2, 3\}$ to attain different missing rates (MRs) for Models I and II given in (3.1) and (3.2).

Model I	MR	$\boldsymbol{\gamma}$							
	10%	$\boldsymbol{\gamma}_1$	3.4	1.5	1.1	0.6	−1.0	0.5	1.0
		$\boldsymbol{\gamma}_2$	1.5	2.4	−1.5	1.4	1.2	−0.6	1.1
		$\boldsymbol{\gamma}_3$	−0.5	0.9	1.1	0.8	1.0	1.2	0.5
	30%	$\boldsymbol{\gamma}_1$	0.5	−1.0	1.2	1.0	−1.8	1.0	0.5
		$\boldsymbol{\gamma}_2$	2.4	1.0	−3.2	2.1	1.0	−1.9	0.8
		$\boldsymbol{\gamma}_3$	1.2	1.1	−1.1	1.0	−2.2	0.7	1.0
Model II	**MR**	$\boldsymbol{\gamma}$							
	10%	$\boldsymbol{\gamma}_1$	1.2	2.5	0.8	−0.6	1.0	0.6	
		$\boldsymbol{\gamma}_2$	0.5	1.5	1.6	−1.0	1.3	0.9	
		$\boldsymbol{\gamma}_3$	1.5	−0.9	1.3	0.9	3.6	0.9	
	30%	$\boldsymbol{\gamma}_1$	1.5	1.0	−1.2	−1.6	1.3	0.7	
		$\boldsymbol{\gamma}_2$	1.3	2.1	1.2	−2.1	1.2	−1.5	
		$\boldsymbol{\gamma}_3$	2.2	1.5	−1.1	0.9	0.8	−2.2	

(5) Set $d = \widehat{d}$ in the first two steps and repeat the first two steps until convergence. The final estimator is given by $\widehat{\boldsymbol{\beta}} = (\mathbf{I}_{d\times d}, \widehat{\boldsymbol{\beta}}_{-d})$.

3. Numerical study

3.1. Simulation experiments

In this section, we use several simulation examples to illustrate the finite sample performance of the proposed IPW method for dealing with missing and multivariate response data. In the simulation below, the covariate $\mathbf{x}$ was drawn from multivariate normal distribution $\mathcal{N}(\mathbf{1}_{p\times 1}, \boldsymbol{\Sigma}_{\mathbf{x}})$, where $\boldsymbol{\Sigma}_{\mathbf{x}} = (\sigma_{k,l})$ with $\sigma_{k,l} = \mathrm{cov}(X_k, X_l) = 0.5^{|k-l|}$ for $1 \le k, l \le p$. To illustrate the performance of the IPW method, we compare our proposal with three methods: the omniscient method (Omni for short) which uses the true selection probability functions π_k in (2.2), the CC method which deletes the missing values naively and imputation approach (IMP for short) which imputes a plausible value for each missing datum and then analyze the results as if they are complete. The following two simulated models were considered:

$$\text{Model I:} \quad \begin{cases} Y_1 = (\boldsymbol{\beta}_1^\top \mathbf{x})/\{0.5 + (\boldsymbol{\beta}_2^\top \mathbf{x} + 1.5)^2\} + \varepsilon_1, \\ Y_2 = \sin(\boldsymbol{\beta}_1^\top \mathbf{x}) + \cos(\boldsymbol{\beta}_2^\top \mathbf{x}) + \varepsilon_2, \\ Y_3 = 2\boldsymbol{\beta}_1^\top \mathbf{x} - \boldsymbol{\beta}_2^\top \mathbf{x} + \varepsilon_3, \end{cases} \tag{3.1}$$

$$\text{Model II:} \quad \begin{cases} Y_1 = (\boldsymbol{\beta}_1^\top \mathbf{x})/\{0.5 + (\boldsymbol{\beta}_2^\top \mathbf{x} + \boldsymbol{\beta}_3^\top \mathbf{x} + 1.5)^2\} + \varepsilon_1, \\ Y_2 = (\boldsymbol{\beta}_1^\top \mathbf{x})^2 + \sin(\boldsymbol{\beta}_2^\top \mathbf{x}) + \boldsymbol{\beta}_3^\top \mathbf{x} + \varepsilon_2, \\ Y_3 = \boldsymbol{\beta}_1^\top \mathbf{x} - 2\boldsymbol{\beta}_2^\top \mathbf{x} + \cos(\boldsymbol{\beta}_3^\top \mathbf{x}) + \varepsilon_3. \end{cases} \tag{3.2}$$

We set $p = 7$, $q = 3$, $\boldsymbol{\beta}_1 = (1, 0, 0.8, -0.6, 0.4, -0.2, 0)^\top$ and $\boldsymbol{\beta}_2 = (0, 1, -0.8, 0.6, -0.4, 0.2, 0)^\top$ in Model I, and set $p = 6$, $q = 3$, $\boldsymbol{\beta}_1 = (1, 0, 0, 0.8, -0.4, 0.2)^\top$, $\boldsymbol{\beta}_2 = (0, 1, 0, -0.2, 0.4, -0.8)^\top$ and $\boldsymbol{\beta}_3 = (0, 0, 1, 0.3, 0.5, 0.7)^\top$ in Model II. The model error $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3)^\top$ was drawn from the normal distribution with mean zero and covariance matrix $\mathrm{cov}(\varepsilon_k, \varepsilon_l) = \left(0.5^{|k-l|}\right)$ for $1 \le k, l \le 3$ in Models I and II. To compare the influence of different missing rates (MRs), the missing mechanisms were chosen from the following logistic models $\pi_k(\mathbf{x}_i, \boldsymbol{\gamma}_k) = 1/\{1 + \exp(-\boldsymbol{\gamma}_k^\top \mathbf{x}_i)\}$, for $k \in \{1, 2, 3\}$, where the values of $\boldsymbol{\gamma}_k$ are given in Table 1. In the proposed IPW approach, the MLE for the parameters $\boldsymbol{\gamma}_k$ was used. The sample sizes n were chosen to be 300 and 500, respectively. The simulation results were based on $N = 1000$ replicates. We took the Gaussian kernel with the bandwidth $h = (4/3n)^{1/(d+4)}s$, where s is the median of the robust estimators of the standard deviation of $\boldsymbol{\beta}^\top \mathbf{x}$.

In Tables 2–5, we report the average bias of the estimators ("bias"), the Monte Carlo standard deviation ("std"), the average of the estimated standard deviation ("$\widehat{\text{std}}$") based on the theoretical calculation, and the empirical coverage probability ("cvp") at the nominal 95% confidence level for all free parameters in models I–II with different settings. In order to show the comprehensive performance intuitively for different parameters, we give the mean square error (MSE) of the estimators in Table 6, where the MSE is defined as $\mathrm{MSE}(\widehat{\boldsymbol{\beta}}) = N^{-1}\sum_{i=1}^{p}\sum_{j=1}^{d}\sum_{k=1}^{N}\left\{\widehat{\beta}_{ij}(k) - \beta_{ij}(k)\right\}^2$. For the imputation method, we only give the bias, MSE and Monte Carlo standard deviation, since the theoretical distribution for the estimator of $\boldsymbol{\beta}$ based on imputation method is unknown. In all, the bias of all the estimators are very small, and they get smaller as the sample size increases. This finding implies that the four estimators are consistent. By comparing the MSE, the omniscient method gives the best performance, while the CC method offers the worst performance. This phenomenon is well-understood, since the omniscient method uses the true selection probability function and the complete case method

Table 2
Simulation results for Model I with $n = 300$: the average bias of the estimators ("bias"), the Monte Carlo standard deviation ("std"), the average of the estimated standard deviation ("$\widehat{\text{std}}$") based on the theoretical calculation, and the empirical coverage probability ("cvp") at the nominal 95% confidence level. All simulation results reported below are multiplied by 100. The methods Omni, IPW and CC are described in Section 3.1. The parameters $\widehat{\beta}_{ij}$ are defined in (2.2).

Method		$\widehat{\beta}_{13}$	$\widehat{\beta}_{14}$	$\widehat{\beta}_{15}$	$\widehat{\beta}_{16}$	$\widehat{\beta}_{17}$	$\widehat{\beta}_{23}$	$\widehat{\beta}_{24}$	$\widehat{\beta}_{25}$	$\widehat{\beta}_{26}$	$\widehat{\beta}_{27}$
		0.8	−0.6	0.4	−0.2	0	−0.8	0.6	−0.4	0.2	0
		Case 1: MR = 10%									
Omni	bias	0.64	−0.30	0.32	−0.19	0.20	−0.51	0.71	−0.40	0.12	0.17
	std	4.66	4.62	3.92	3.61	3.07	5.45	5.40	4.95	4.31	3.61
	$\widehat{\text{std}}$	5.13	5.05	4.35	3.84	3.29	6.27	6.19	5.34	4.72	4.05
	cvp	96.00	95.80	96.00	95.10	95.80	97.10	96.90	95.00	96.10	97.30
IPW	bias	0.63	−0.25	0.32	−0.15	0.15	−0.53	0.78	−0.49	0.08	0.09
	std	4.78	4.72	3.94	3.66	3.15	5.59	5.56	5.02	4.42	3.71
	$\widehat{\text{std}}$	5.23	5.12	4.42	3.89	3.33	6.45	6.35	5.49	4.85	4.16
	cvp	95.90	96.10	97.10	95.40	95.80	97.60	96.80	95.50	96.50	97.10
CC	bias	0.92	−0.60	0.45	−0.20	−0.20	−0.21	0.68	−0.76	0.26	−0.07
	std	5.66	5.49	4.87	4.23	3.66	7.31	7.24	6.71	5.92	5.04
	$\widehat{\text{std}}$	5.58	5.43	4.78	4.18	3.59	7.39	7.24	6.41	5.63	4.86
	cvp	94.20	94.50	94.20	95.20	94.60	95.30	95.40	93.70	93.10	92.90
		Case 2: MR = 30%									
Omni	bias	0.67	−0.38	0.30	−0.12	−0.02	−0.17	0.35	−0.32	0.09	0.01
	std	4.91	4.66	4.03	3.56	3.23	5.13	4.98	4.36	3.93	3.35
	$\widehat{\text{std}}$	5.62	5.52	4.80	4.19	3.59	5.75	5.65	4.90	4.31	3.70
	cvp	95.20	96.80	97.30	96.90	96.30	95.70	96.10	96.30	95.80	96.20
IPW	bias	0.67	−0.31	0.22	−0.08	−0.02	−0.29	0.53	−0.47	0.19	−0.01
	std	4.95	4.78	4.19	3.69	3.26	5.23	5.15	4.56	4.17	3.47
	$\widehat{\text{std}}$	5.83	5.72	5.00	4.34	3.71	6.04	5.96	5.19	4.53	3.88
	cvp	96.60	97.10	97.20	97.00	96.70	96.80	97.10	96.40	96.00	96.70
CC	bias	1.88	−1.13	0.68	0.01	−0.18	0.19	0.86	−1.09	0.45	−0.08
	std	8.13	8.01	7.51	6.32	5.48	11.35	11.65	11.24	9.44	7.65
	$\widehat{\text{std}}$	8.12	7.95	7.23	5.95	4.91	11.13	10.90	9.99	8.23	6.80
	cvp	95.70	94.50	94.40	93.80	91.40	93.50	94.00	91.50	91.50	91.50

only uses the information of the observed data. Also, our proposed inverse probability weighted approach performs better than the imputation method. When the missing rate becomes higher, the complete case method performs worse quickly while the omniscient and IPW methods are not affected so much. The Monte Carlo standard deviations are very close to the average of the estimated standard deviations which imply that the standard deviations were estimated precisely and confirm the consistency results stated in Theorem 2.2. The empirical coverage probabilities of the 95% confidence interval for three estimators are very close to the pre-specified nominal level, which suggest that our inferential results are reliable.

To illustrate the performance of the proposed Wald test approach, we tested whether $\mathbf{x}_7$ is important in predicting $\mathbf{y}$ in Model I. We choose $\mathbf{Q} = (0, \ldots, 0, 1)_{1\times 5}$, $\mathbf{q}_0 = \mathbf{0}_{1\times 2}$ in our testing problem. The power performance was investigated by simulation runs with different alternatives through varying the values of the last row of $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$. Specifically, we change the last row of $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$ to (τ, τ), where τ is from 0 to 0.16 and the step length is 0.02. Obviously, $\tau = 0$ corresponds to the case that $\mathbf{x}_7$ is not important covariate. The power curves with different sample sizes and missing rates are displayed in Fig. 1. The empirical sizes of T_n with three different estimation methods are very close to the theoretical level 0.05. Under the alternative hypothesis $\tau \neq 0$, the power increases quickly as τ increases, that is, the tests are very sensitive to the alternatives. Also, the power with $n = 500$ is higher than that with $n = 300$, and that with a small MR is higher than that with a large MR. Besides, the power performance based on the omniscient and IPW methods is significantly better than that based on the CC method.

We now further assess efficacy of the proposed BIC-type criterion in estimating the structural dimension of the span($\boldsymbol{\beta}$) for incomplete data. The estimation results of the structural dimension based on Models I and II with $(\alpha, \kappa) = (1, 0.05)$ are summarized in Table 7. The structural dimension of Model I is $d = 2$ and that of Model II is $d = 3$. From Table 7 we can see that the BIC-type criterion gives very good performance for all three methods. When the MR is high, the complete case method offers poorer performance than the omniscient and IPW approaches.

3.2. Application to hypertension study

We further illustrate our proposed method by applying to analyzing a primary hypertension data collected in the Inner Mongolia Autonomous Region of P. R. China in 2002. The aim of this study was to understand environmental risk factors

Table 3
Simulation results for Model I with $n = 500$: the average bias of the estimators ("bias"), the Monte Carlo standard deviation ("std"), the average of the estimated standard deviation ("$\widehat{\text{std}}$") based on the theoretical calculation, and the empirical coverage probability ("cvp") at the nominal 95% confidence level. All simulation results reported below are multiplied by 100. The methods Omni, IPW and CC are described in Section 3.1. The parameters $\widehat{\beta}_{ij}$ are defined in (2.2).

Method		$\widehat{\beta}_{13}$	$\widehat{\beta}_{14}$	$\widehat{\beta}_{15}$	$\widehat{\beta}_{16}$	$\widehat{\beta}_{17}$	$\widehat{\beta}_{23}$	$\widehat{\beta}_{24}$	$\widehat{\beta}_{25}$	$\widehat{\beta}_{26}$	$\widehat{\beta}_{27}$
		0.8	−0.6	0.4	−0.2	0	−0.8	0.6	−0.4	0.2	0
		Case 1: MR = 10%									
Omni	bias	0.35	−0.15	0.19	−0.05	0.27	−0.42	0.79	−0.19	0.07	0.29
	std	3.75	3.67	3.09	2.74	2.45	4.25	4.23	3.80	3.26	2.76
	$\widehat{\text{std}}$	3.98	3.91	3.36	2.97	2.55	4.86	4.79	4.13	3.66	3.14
	cvp	95.40	96.40	96.00	95.90	95.60	97.30	96.70	95.40	96.80	96.50
IPW	bias	0.34	−0.13	0.18	−0.04	0.28	−0.48	0.86	−0.25	0.11	0.23
	std	3.81	3.70	3.15	2.83	2.45	4.29	4.26	3.84	3.34	2.79
	$\widehat{\text{std}}$	4.04	3.96	3.41	3.02	2.58	4.99	4.91	4.24	3.75	3.21
	cvp	95.50	96.10	96.70	95.70	95.30	97.70	97.20	96.20	96.60	97.30
CC	bias	0.95	−0.51	0.22	−0.08	−0.16	0.00	0.44	−0.40	0.34	−0.51
	std	4.39	4.25	3.72	3.42	2.97	5.87	5.69	5.05	4.75	4.07
	$\widehat{\text{std}}$	4.52	4.39	3.85	3.37	2.88	6.08	5.93	5.25	4.59	3.95
	cvp	95.90	95.10	94.70	95.50	94.10	95.70	95.80	95.90	94.30	93.60
		Case 2: MR = 30%									
Omni	bias	0.24	−0.14	0.07	−0.12	0.15	−0.17	0.30	−0.27	0.14	0.03
	std	3.82	3.72	3.24	2.98	2.49	4.07	4.04	3.38	3.05	2.55
	$\widehat{\text{std}}$	4.05	3.97	3.43	3.03	2.57	4.22	4.12	3.55	3.14	2.67
	cvp	94.60	96.20	95.50	94.00	94.70	95.00	95.30	95.10	95.20	95.40
IPW	bias	0.22	−0.12	0.04	−0.08	0.15	−0.19	0.35	−0.34	0.21	0.03
	std	3.94	3.82	3.31	3.02	2.58	4.18	4.20	3.57	3.18	2.67
	$\widehat{\text{std}}$	4.17	4.09	3.55	3.12	2.65	4.41	4.31	3.73	3.29	2.79
	cvp	94.70	95.90	96.00	94.70	95.30	95.60	94.70	95.10	95.80	96.20
CC	bias	1.40	−0.64	0.06	0.16	0.07	0.32	1.09	−1.33	0.78	−0.24
	std	6.43	6.39	5.68	4.82	4.05	9.04	9.05	8.47	7.01	6.01
	$\widehat{\text{std}}$	6.61	6.45	5.84	4.82	3.95	9.43	9.19	8.37	6.91	5.69
	cvp	95.20	95.40	96.00	94.90	93.40	95.60	94.50	95.50	94.20	94.00

related to the primary hypertension. The hypertension data set contains 1051 subjects that aged 20 years or older and had been living in the rural areas in the region at least for three generations. Both the systolic (Y_1) and the diastolic blood pressures (Y_2) are recorded for each subject. Fifteen candidate environmental risk factors denoted as $X_1, \ldots, X_{15}$ were collected: height, weight, age, waist, gender, glucose, triglyceride (tg for short), total cholesterol (tc for short), high density lipoprotein cholesterol (hdlc for short), low density lipoprotein cholesterol (ldlc for short), Apolipoprotein A1 (apoa for short), Apolipoprotein B (apob for short), blood urea nitrogen (bun for short), cellulase gene glone (cre for short) and hip. Since the data have no missing data, in order to illustrate our proposed method, we used the following two selection probability functions to remove some responses,

$$\pi_1(\mathbf{x}) = 1/[1 + \exp\{-0.2(X_2 + X_3) + 3.3(X_7 + X_8)\}], \quad \pi_2(\mathbf{x}) = 1/[1 + \exp\{0.4(X_5 + X_6) - 1.2(X_{10} + X_{12})\}].$$

The MRs are about 0.2331 and 0.3720 for $\pi_1(\mathbf{x})$ and $\pi_2(\mathbf{x})$ respectively. The BIC-type criterion yielded $\widehat{d} = 2$ for all three methods. The profile least squares estimates, along with their standard deviations and the p-values are given in Table 8 from which we can see that the omniscient method gives the smallest standard deviations and the complete case approach offers the biggest standard deviations. The major risk factors predisposing to essential hypertension in Mongolian population include age, obesity, gender, total cholesterol, triglyceride, Apolipoprotein A1, etc. The potential causes for controlling essential hypertension could be weight control and healthy diet habits. Thus doing exercises and keeping the balance of wide variety of chemicals in the diet are very important to control essential hypertension. Our analysis provides scientific foundations for prevention and control of essential hypertension.

4. Concluding remarks

In practice, nowadays many output and input data are easily collected. For example, as demonstrated in the real data analysis of hypertension studies, both the diastolic and the systolic blood pressures, as well as many environmental factors are typically recorded simultaneously. In such studies, both the responses and predictors are multivariate and further the response data may be missing at random. To reduce the dimension of the predictors, we estimate the central mean subspace via the profile least squares method with the help of the inverse probability weighted approach. The proposed estimation method is different from the existing SDR methods, since our proposed estimation method does not require any

Table 4
Simulation results for Model II with $n = 300$: the average bias of the estimators ("bias"), the Monte Carlo standard deviation ("std"), the average of the estimated standard deviation ("$\widehat{\text{std}}$") based on the theoretical calculation, and the empirical coverage probability ("cvp") at the nominal 95% confidence level. All simulation results reported below are multiplied by 100. The methods Omni, IPW and CC are described in Section 3.1. The parameters $\widehat{\beta}_{ij}$ are defined in (2.2).

Method		$\widehat{\beta}_{14}$	$\widehat{\beta}_{15}$	$\widehat{\beta}_{16}$	$\widehat{\beta}_{24}$	$\widehat{\beta}_{25}$	$\widehat{\beta}_{26}$	$\widehat{\beta}_{34}$	$\widehat{\beta}_{35}$	$\widehat{\beta}_{36}$
		0.8	−0.4	0.2	−0.2	0.4	−0.8	0.3	0.5	0.7
		Case 1: MR = 10%								
Omni	bias	−0.63	−0.85	−0.46	−1.38	−1.79	0.50	3.30	4.23	2.61
	std	1.93	1.77	1.58	3.59	3.22	2.98	7.01	6.19	5.64
	$\widehat{\text{std}}$	2.38	2.24	2.04	4.35	3.92	3.80	7.62	6.85	6.74
	cvp	97.00	97.40	96.70	96.90	94.70	97.90	94.80	90.90	94.90
IPW	bias	−0.63	−0.85	−0.45	−1.41	−1.83	0.51	3.27	4.23	2.55
	std	1.91	1.79	1.59	3.60	3.31	3.01	7.07	6.32	5.68
	$\widehat{\text{std}}$	2.38	2.24	2.03	4.33	3.88	3.77	7.60	6.81	6.68
	cvp	97.20	96.90	96.80	96.70	94.20	98.10	94.10	90.90	95.10
CC	bias	−0.88	−1.01	−0.58	−1.74	−2.27	0.81	5.36	5.23	3.81
	std	1.99	1.90	1.70	3.97	3.78	3.43	7.98	7.72	6.96
	$\widehat{\text{std}}$	2.05	1.97	1.75	4.11	3.92	3.49	8.61	8.30	7.44
	cvp	94.50	92.70	94.80	93.50	92.00	95.20	93.10	92.60	94.40
		Case 2: MR = 30%								
Omni	bias	−0.38	−0.65	−0.28	−0.49	−1.34	0.67	1.81	3.15	0.85
	std	1.87	1.95	1.60	3.99	3.88	3.14	7.04	7.26	5.78
	$\widehat{\text{std}}$	2.24	2.37	1.84	4.19	4.18	3.23	6.98	7.10	5.59
	cvp	94.90	95.70	94.60	93.80	94.00	92.80	92.70	91.00	90.20
IPW	bias	−0.35	−0.62	−0.27	−0.47	−1.35	0.66	1.78	2.97	0.88
	std	1.92	1.92	1.59	4.10	3.87	3.08	7.59	7.38	5.75
	$\widehat{\text{std}}$	2.22	2.35	1.83	4.17	4.18	3.21	6.89	7.05	5.53
	cvp	95.10	96.20	94.40	95.10	94.30	92.50	91.90	91.90	89.90
CC	bias	−0.46	−1.21	−0.35	−0.89	−2.73	1.31	2.88	7.22	2.23
	std	2.51	2.34	2.36	5.50	4.96	5.12	9.99	9.62	10.23
	$\widehat{\text{std}}$	2.48	2.30	2.31	5.39	4.92	5.05	10.65	9.77	10.06
	cvp	95.10	91.60	95.00	94.10	92.70	93.10	95.80	90.80	93.70

distributional assumption on the covariates, and hence allows for categorical predictors and facilitates statistical inference on the central mean subspace.

Acknowledgments

The authors thank the Editor-in-Chief, an Associate Editor and two anonymous referees for helpful comments and suggestions which led to an improved version of this paper. We also thank Professor Suojin Wang for detailed comments and suggestions. This research was supported by the National Natural Science Foundation of China (11671299, 11401006).

Appendix A. Some additional simulations

In this Appendix, we present additional simulation results to show the performance of our proposals with large p and q. Specifically, we considered the following model,

$$\text{Model III:}\quad \begin{cases} Y_1 = (\boldsymbol{\beta}_1^\top \mathbf{x})/\{0.5 + (\boldsymbol{\beta}_2^\top \mathbf{x} + 1.5)^2\} + \varepsilon_1, \\ Y_2 = \sin(\boldsymbol{\beta}_1^\top \mathbf{x}) - 2\boldsymbol{\beta}_2^\top \mathbf{x} + \varepsilon_2, \\ Y_3 = 2\boldsymbol{\beta}_1^\top \mathbf{x} - \cos(\boldsymbol{\beta}_2^\top \mathbf{x}) + \varepsilon_3, \\ Y_4 = (\boldsymbol{\beta}_1^\top \mathbf{x})^2 - \boldsymbol{\beta}_2^\top \mathbf{x} + \varepsilon_4, \\ Y_5 = 4\boldsymbol{\beta}_1^\top \mathbf{x} + 2(\boldsymbol{\beta}_2^\top \mathbf{x})^2 + \varepsilon_5, \\ Y_6 = (\boldsymbol{\beta}_2^\top \mathbf{x})/\{0.5 + (\boldsymbol{\beta}_1^\top \mathbf{x} + 1.5)^2\} + \varepsilon_6, \end{cases} \tag{A.1}$$

where we set $p = 12$, $q = 6$, $\boldsymbol{\beta}_1 = (1, 0, 0.8, -0.6, 0.4, -0.2, 0, -0.8, 0.6, -0.4, 0.2, 0)^\top$ and $\boldsymbol{\beta}_2 = (0, 1, -0.8, 0.6, -0.4, 0.2, 0, 0.8, -0.6, 0.4, -0.2, 0)^\top$. The values of $\boldsymbol{\gamma}_k$ in (2.1) are given in Table 9 that resulted in approximately 10% and 30% MRs. The bias, std, $\widehat{\text{std}}$ and the cvp at the nominal 95% confidence level for all free parameters are presented in Tables 10–13 for different settings. The MSEs for four different estimators with different sample sizes and missing rates are put in Table 14. It can be seen from Tables 10–14 that our proposed method still works well when the dimensions

Table 5
Simulation results for Model II with $n = 500$: the average bias of the estimators ("bias"), the Monte Carlo standard deviation ("std"), the average of the estimated standard deviation ("$\widehat{\text{std}}$") based on the theoretical calculation, and the empirical coverage probability ("cvp") at the nominal 95% confidence level. All simulation results reported below are multiplied by 100. The methods Omni, IPW and CC are described in Section 3.1. The parameters $\widehat{\beta}_{ij}$ are defined in (2.2).

Method		$\widehat{\beta}_{14}$	$\widehat{\beta}_{15}$	$\widehat{\beta}_{16}$	$\widehat{\beta}_{24}$	$\widehat{\beta}_{25}$	$\widehat{\beta}_{26}$	$\widehat{\beta}_{34}$	$\widehat{\beta}_{35}$	$\widehat{\beta}_{36}$
		0.8	−0.4	0.2	−0.2	0.4	−0.8	0.3	0.5	0.7
		Case 1: MR = 10%								
Omni	bias	−0.61	−0.75	−0.50	−1.48	−1.56	0.24	3.11	3.85	2.65
	std	1.47	1.39	1.33	2.83	2.73	2.44	5.66	5.04	4.71
	$\widehat{\text{std}}$	1.83	1.71	1.54	3.38	2.99	2.88	6.00	5.25	5.17
	cvp	96.10	96.90	95.70	95.50	92.40	97.40	92.20	89.10	92.40
IPW	bias	−0.61	−0.74	−0.49	−1.46	−1.55	0.26	3.13	3.80	2.60
	std	1.47	1.39	1.34	2.84	2.71	2.44	5.64	5.00	4.75
	$\widehat{\text{std}}$	1.82	1.70	1.54	3.37	2.98	2.87	5.98	5.25	5.14
	cvp	96.50	96.40	95.30	95.70	92.70	96.90	92.20	90.10	92.60
CC	bias	−0.86	−0.95	−0.61	−1.59	−2.07	0.63	5.27	5.04	3.84
	std	1.59	1.55	1.42	3.15	3.13	2.84	6.61	6.57	5.70
	$\widehat{\text{std}}$	1.68	1.61	1.43	3.39	3.24	2.87	7.28	7.05	6.24
	cvp	93.40	92.80	92.80	94.30	91.40	95.20	92.40	91.50	93.00
		Case 2: MR = 30%								
Omni	bias	−0.24	−0.45	−0.23	−0.53	−1.15	0.63	1.10	2.42	0.74
	std	1.57	1.59	1.37	3.11	3.09	2.52	6.54	6.05	4.90
	$\widehat{\text{std}}$	1.70	1.76	1.40	3.30	3.20	2.50	5.38	5.39	4.22
	cvp	94.10	94.80	92.20	94.80	93.60	91.20	91.70	90.40	91.10
IPW	bias	−0.24	−0.43	−0.24	−0.54	−1.12	0.62	1.07	2.30	0.79
	std	1.64	1.62	1.38	3.11	3.05	2.57	6.91	6.30	5.06
	$\widehat{\text{std}}$	1.70	1.76	1.39	3.28	3.18	2.47	5.37	5.39	4.20
	cvp	94.70	94.30	93.10	95.40	94.30	91.30	91.80	90.10	91.40
CC	bias	−0.37	−1.11	−0.37	−0.98	−2.54	0.98	2.70	6.47	2.43
	std	2.04	1.79	2.05	4.43	3.94	4.21	8.64	7.90	8.49
	$\widehat{\text{std}}$	2.08	1.90	1.91	4.56	4.14	4.24	9.23	8.37	8.68
	cvp	95.60	92.90	92.00	95.00	93.90	94.80	96.10	92.10	93.80

Table 6
The MSE results for Models I and II with different sample sizes and missing rates (MRs). All simulation results reported below are multiplied by 100. The methods Omni, IPW, IMP and CC are described in Section 3.1.

Model	n	Case 1: MR = 10%				Case 2: MR = 30%			
		Omni	IPW	IMP	CC	Omni	IPW	IMP	CC
I	300	2.3709	2.3780	2.8719	3.2486	2.8011	2.8072	4.4831	8.0105
	500	1.3703	1.3775	1.8531	1.9774	1.5430	1.6666	2.8897	4.9815
II	300	2.4900	2.5301	2.6046	3.5130	2.7569	2.7959	3.4639	5.7921
	500	1.8489	1.8662	1.9600	2.4480	2.1583	2.2603	2.8026	4.1918

p and q increase, and the simulation results are similar to that in Section 3.1. For example, by the MSE, the omniscient method gives the best performance, and the inverse probability weighted method performs better than its competitors: the complete case and imputation methods. The omniscient and IPW methods are not affected so much by missing rates, while the complete case method performs worse quickly when the missing rate gets higher. Also, the Monte Carlo standard deviations are very close to the average of the estimated standard deviations and the empirical coverage probabilities of the 95% confidence interval are very close to the pre-specified nominal level.

Further, we evaluated the efficacy of the proposed BIC-type criterion in determining the structural dimension of span($\boldsymbol{\beta}$) in Model III for incomplete data. The structural dimension of the span($\boldsymbol{\beta}$) is $d = 2$ and the penalty term is chosen to be $(\alpha, \kappa) = (1, 0.05)$. The percentages for each estimated dimension are charted in Table 15 from which we can see that the BIC-type criterion has satisfactory performance.

In addition, to show the influence of the penalty term on the BIC-type criterion, we have tried several values of $\lambda_n = \alpha n^{\kappa}$ and give some typical results of (α, κ) for Models I–III with $n = 500$ and MR = 30%. The frequencies based on three methods of the estimated structural dimension $\widehat{d}$ are summarized in Table 16. It can be concluded from Table 16 that the choice of $(\alpha, \kappa) = (1, 0.05)$ gives the best performance. Further, when the penalty term gets larger, the proportion of underestimate dimension gradually gets higher, while the percentage of overestimate dimension gets higher as the penalty term decreases. For example, when $(\alpha, \kappa) = (1, 0.3)$ in Model III, the percentages of $\widehat{d} = 1$ (which underestimated) for all three methods are very high. The proportions of overestimated dimension for Model II with $(\alpha, \kappa) = (0.35, 0.05)$ based

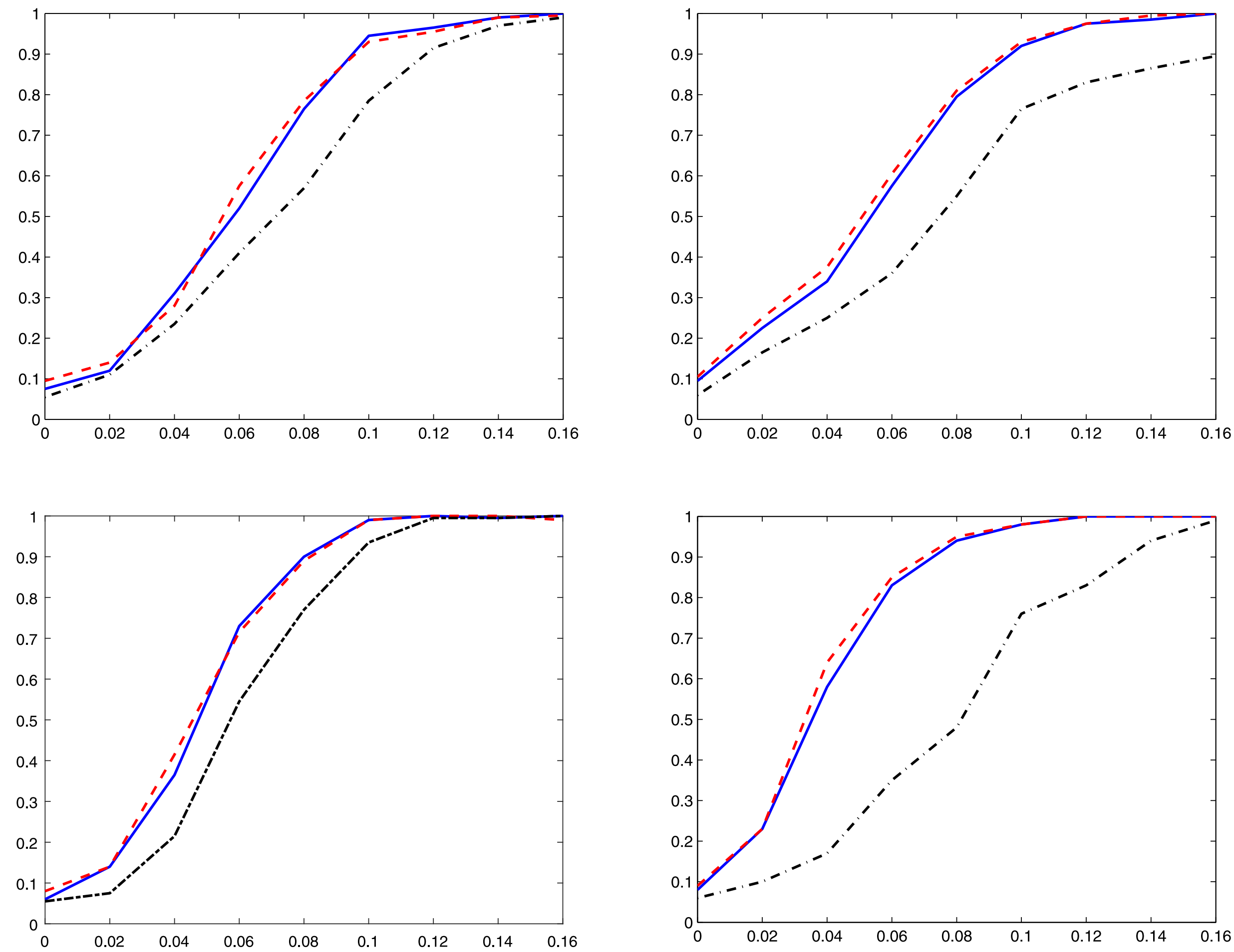

Fig. 1. The power curves of the Wald test for MR = 10% (left panel) and MR = 30% (right panel) with $n = 300$ (top) and $n = 500$ (bottom) based on Omni (solid line), IPW (dashed line) and CC (dot dash line) methods. The methods Omni, IPW and CC are described in Section 3.1.

on the three methods are very high too. Besides, it is also interesting to note that the structural dimension estimation of Model I is more sensitive to penalty term than the other two models. To sum up, we recommend to use $(\alpha, \kappa) = (1, 0.05)$ in the penalty term in practice.

Appendix B. Proof of Theorem 2.1

Lemma B.1 ([18]). *Let $\{\zeta_1, \ldots, \zeta_n\}$ be independent and identically distributed random variables with $E\zeta_1 = 0$ and $E|\zeta_1|^r \leq C < \infty$ for some $r > 1$. Suppose that $\{a_{ij}, 1 \leq i, j \leq n\}$ is a series of real numbers such that $\max_{1\leq j\leq n} \sum_{i=1}^n |a_{ij}| \leq C < \infty$. Set $d_n = \max_{1\leq i,j\leq n} |a_{ij}|$. Then $\max_{1\leq j\leq n} \left|\sum_{i=1}^n a_{ij}\zeta_i\right| = O\left\{(n^{1/r} d_n \vee d_n^{1/2}) \ln n\right\}$, a.s.*

Let $\widehat{\pi}_{ik} = \pi(\mathbf{x}_i, \widehat{\boldsymbol{\gamma}}_k)$, $\pi_{ik} = \pi(\mathbf{x}_i, \boldsymbol{\gamma}_k)$, $\mathbf{W}_i = \operatorname{diag}\left\{W_{i1}, \ldots, W_{iq}\right\} = \operatorname{diag}\left\{\delta_{i1}/\pi_1(\mathbf{x}_i), \ldots, \delta_{iq}/\pi_q(\mathbf{x}_i)\right\}$,

$$\boldsymbol{\Gamma}_k = E\left[\pi_k(\mathbf{x}_i)\{1 - \pi_k(\mathbf{x}_i)\}\mathbf{x}_i\mathbf{x}_i^\top\right],\ \widehat{E}(\boldsymbol{\varepsilon}_i|\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i) = \left\{\widehat{E}_1(\varepsilon_{i1}|\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i), \ldots, \widehat{E}_q(\varepsilon_{iq}|\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i)\right\}^\top,$$
$$\widehat{E}(\mathbf{x}_{-d,i}|\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i) = \left\{\widehat{E}_1(\mathbf{x}_{-d,i}|\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i), \ldots, \widehat{E}_q(\mathbf{x}_{-d,i}|\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i)\right\}$$

and $\widehat{\widetilde{\mathbf{x}}}_{-d,i} = \left\{\mathbf{x}_{-d,i} - \widehat{E}_1(\mathbf{x}_{-d,i}|\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i), \ldots, \mathbf{x}_{-d,i} - \widehat{E}_q(\mathbf{x}_{-d,i}|\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i)\right\}$, where

$$\widehat{E}_k(\mathbf{x}_{-d,i}|\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i) = \sum_{j\neq i} \delta_{jk} K_h(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_j - \widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i)\mathbf{x}_{-d,j} \Big/ \sum_{j\neq i} \delta_{jk} K_h(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_j - \widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i),$$
$$\widehat{E}_k(\varepsilon_{ik}|\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i) = \sum_{j\neq i} \delta_{jk} K_h(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_j - \widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i)\varepsilon_{jk} \Big/ \sum_{j\neq i} \delta_{jk} K_h(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_j - \widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i)$$

Table 7
The frequencies (%) of the estimated structural dimension $\widehat{d}$ for Models I and II with different sample sizes and missing rates (MRs). The methods Omni, IPW and CC are described in Section 3.1.

Molde	n	MR	Method	$\widehat{d}=1$	$\widehat{d}=2$	$\widehat{d}=3$	$\widehat{d}\geq 4$
Model I	300	10%	Omni	3.30	96.70	0.00	0.00
			IPW	4.20	95.80	0.00	0.00
			CC	17.60	82.40	0.00	0.00
		30%	Omni	5.20	94.80	0.00	0.00
			IPW	4.30	95.70	0.00	0.00
			CC	44.30	55.70	0.00	0.00
	500	10%	Omni	0.00	100.00	0.00	0.00
			IPW	0.00	100.00	0.00	0.00
			CC	0.00	100.00	0.00	0.00
		30%	Omni	0.00	100.00	0.00	0.00
			IPW	0.00	100.00	0.00	0.00
			CC	8.60	91.40	0.00	0.00
Model II	300	10%	Omni	0.00	0.00	100.00	0.00
			IPW	0.00	0.00	100.00	0.00
			CC	0.00	0.00	100.00	0.00
		30%	Omni	0.00	0.00	100.00	0.00
			IPW	0.00	0.00	100.00	0.00
			CC	0.00	0.00	100.00	0.00
	500	10%	Omni	0.00	0.00	100.00	0.00
			IPW	0.00	0.00	100.00	0.00
			CC	0.00	0.00	100.00	0.00
		30%	Omni	0.00	0.00	100.00	0.00
			IPW	0.00	0.00	100.00	0.00
			CC	0.00	0.00	100.00	0.00

Table 8
The estimated coefficients, the standard errors along with the p-values based on the Ommi, IPW and CC methods for the hypertension study data. The methods Omni, IPW and CC are described in Section 3.1.

	age	waist	gender	glucose	tg	tc	hdlc	ldlc	apoa	apob	bun	cre	hip
Omni													
coef	0.4589	0.2033	0.4144	0.9194	1.5558	−2.5903	2.1748	2.4774	9.2123	0.3360	−0.5714	0.0903	0.4138
	0.4587	0.2068	0.4172	0.8991	1.5837	−2.6258	2.2905	2.5476	9.1465	0.2816	−0.5759	0.0912	0.4117
std	0.0029	0.0059	0.0787	0.0313	0.0313	0.0561	0.2021	0.0483	0.1731	0.1639	0.0177	0.0040	0.0090
	0.0021	0.0050	0.1157	0.0165	0.0202	0.0405	0.1392	0.0382	0.1473	0.2020	0.0226	0.0027	0.0083
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0403	0.0000	0.0000	0.0000
	0.0000	0.0000	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1633	0.0000	0.0000	0.0000
IPW													
coef	0.4590	0.2032	0.4163	0.9174	1.5599	−2.5937	2.1899	2.4757	9.2003	0.3504	−0.5711	0.0901	0.4138
	0.4587	0.2065	0.4165	0.9016	1.5823	−2.6249	2.2862	2.5421	9.1489	0.2938	−0.5762	0.0912	0.4116
std	0.0030	0.0061	0.0805	0.0319	0.0343	0.0576	0.2183	0.0503	0.1739	0.1673	0.0180	0.0040	0.0092
	0.0021	0.0051	0.1185	0.0171	0.0204	0.0411	0.1392	0.0379	0.1506	0.2070	0.0231	0.0027	0.0084
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0362	0.0000	0.0000	0.0000
	0.0000	0.0000	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1557	0.0000	0.0000	0.0000
CC													
coef	0.4646	0.1784	0.5480	1.2437	2.1166	−4.1440	6.9894	4.3364	7.6726	−0.0062	−0.8023	0.1624	0.4114
	0.4957	0.2746	−1.9330	1.1240	2.6016	−2.0561	2.2061	0.5530	9.7228	−1.3318	−0.8624	0.1731	0.5029
std	0.0163	0.0369	0.5952	0.1695	0.3530	0.5061	0.9872	0.6101	1.1822	1.6311	0.1224	0.0206	0.0484
	0.0170	0.0390	0.5985	0.1902	0.3275	0.4249	0.8719	0.5234	1.0091	1.3993	0.1179	0.0192	0.0555
p-value	0.0000	0.0000	0.3572	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9970	0.0000	0.0000	0.0000
	0.0000	0.0000	0.0012	0.0000	0.0000	0.0000	0.0114	0.2907	0.0000	0.3412	0.0000	0.0000	0.0000

for $k \in \{1,\ldots,q\}$. Note that $\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}^\top\mathbf{x}_i) = \widehat{E}(\mathbf{y}_i|\widehat{\boldsymbol\beta}^\top\mathbf{x}_i)$, which together with $\mathbf{y}_i = \mathbf{m}(\boldsymbol\beta^\top\mathbf{x}_i) + \boldsymbol\varepsilon_i$, gives that $\mathbf{y}_i - \widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}^\top\mathbf{x}_i) = \mathbf{m}(\boldsymbol\beta^\top\mathbf{x}_i) + \boldsymbol\varepsilon_i - \widehat{E}\{\mathbf{m}(\boldsymbol\beta^\top\mathbf{x}_i) + \boldsymbol\varepsilon_i|\widehat{\boldsymbol\beta}^\top\mathbf{x}_i\}$. It follows from Taylor expansion that $\mathbf{m}(\boldsymbol\beta^\top\mathbf{x}_i) = \mathbf{m}(\widehat{\boldsymbol\beta}^\top\mathbf{x}_i) + \mathbf{m}^{(1)}(\widehat{\boldsymbol\beta}^\top\mathbf{x}_i)(\boldsymbol\beta_{-d} - \widehat{\boldsymbol\beta}_{-d})^\top\mathbf{x}_{-d,i} + o_p(\|\widehat{\boldsymbol\beta}_{-d} - \boldsymbol\beta_{-d}\|)$. Hence

$$\mathbf{y}_i - \widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}^\top\mathbf{x}_i) = \mathscr{D}\left\{\mathbf{m}^{(1)}(\widehat{\boldsymbol\beta}^\top\mathbf{x}_i)(\boldsymbol\beta_{-d} - \widehat{\boldsymbol\beta}_{-d})^\top\widetilde{\widehat{\mathbf{x}}}_{-d,i}\right\} + \{\boldsymbol\varepsilon_i - \widehat{E}(\boldsymbol\varepsilon_i|\widehat{\boldsymbol\beta}^\top\mathbf{x}_i)\} + o_p(\|\widehat{\boldsymbol\beta}_{-d} - \boldsymbol\beta_{-d}\|),$$

Table 9
The choices of γ_k, $k \in \{1, \ldots, 6\}$ to attain different missing rates (MRs) for Model III that given in (A.1).

MR	γ												
10%	γ_1	2.1	−2.0	1.5	2.3	2.5	−2.2	−0.2	1.1	−1.3	1.6	2.1	0.2
	γ_2	1.4	2.2	2.4	1.7	−2.4	1.9	1.3	−1.1	1.6	−2.5	1.3	1.0
	γ_3	1.6	−1.0	1.6	1.6	−1.2	2.5	2.8	−1.5	−1.1	−1.2	1.8	1.9
	γ_4	1.5	1.1	2.7	−1.1	0.6	−1.9	1.8	0.3	−0.7	0.3	0.3	1.5
	γ_5	2.2	−1.1	0.3	0.6	2.7	2.2	0.4	0.9	−2.7	2.9	−2.1	1.8
	γ_6	2.1	−2.1	0.5	2.6	0.5	−2.4	0.9	1.4	−1.3	2.0	1.3	1.9
30%	γ_1	0.5	0.5	1.2	−1.6	0.2	−1.7	3.3	1.1	−2.1	−2.7	1.1	3.3
	γ_2	−2.2	−1.5	1.0	2.1	−1.6	1.5	−2.3	2.4	2.1	−0.2	1.4	0.3
	γ_3	2.1	1.7	−2.6	1.5	1.3	0.4	−1.2	1.3	−1.5	−3.1	2.5	0.5
	γ_4	1.9	−1.5	1.4	−1.3	−1.3	0.4	−1.6	1.4	−2.3	2.8	0.1	2.9
	γ_5	2.3	−2.9	1.6	0.7	2.2	0.3	1.1	−2.7	1.5	1.2	−2.0	−0.5
	γ_6	1.9	1.2	1.6	−2.6	1.4	2.0	−2.6	1.5	−1.9	0.7	−2.9	2.7

Table 10
Simulation results for Model III with $n = 300$ and MR = 10%: the average bias of the estimators ("bias"), the Monte Carlo standard deviation ("std"), the average of the estimated standard deviation ("$\widehat{\text{std}}$") based on the theoretical calculation, and the empirical coverage probability ("cvp") at the nominal 95% confidence level. All simulation results reported below are multiplied by 100. The methods Omni, IPW and CC are described in Section 3.1. The parameters $\widehat{\beta}_{ij}$ are defined in (2.2).

Method	$\boldsymbol{\beta}_1$	$\widehat{\beta}_{1,3}$	$\widehat{\beta}_{1,4}$	$\widehat{\beta}_{1,5}$	$\widehat{\beta}_{1,6}$	$\widehat{\beta}_{1,7}$	$\widehat{\beta}_{1,8}$	$\widehat{\beta}_{1,9}$	$\widehat{\beta}_{1,10}$	$\widehat{\beta}_{1,11}$	$\widehat{\beta}_{1,12}$
		0.8	−0.6	0.4	−0.2	0	−0.8	0.6	−0.4	0.2	0
Omni	bias	−0.49	0.44	−0.14	0.22	−0.11	0.54	−0.38	0.39	−0.20	0.13
	std	2.50	2.52	2.22	1.77	1.86	2.80	2.60	2.06	1.78	1.59
	$\widehat{\text{std}}$	2.38	2.37	2.07	1.82	1.76	2.77	2.35	2.04	1.83	1.57
	cvp	91.90	92.30	93.30	94.20	94.20	93.30	90.50	94.70	95.70	94.20
IPW	bias	−0.44	0.42	−0.13	0.20	−0.09	0.47	−0.32	0.37	−0.21	0.14
	std	2.47	2.49	2.22	1.78	1.85	2.78	2.64	2.05	1.79	1.60
	$\widehat{\text{std}}$	2.36	2.35	2.05	1.80	1.75	2.74	2.32	2.02	1.82	1.55
	cvp	92.80	93.30	93.30	92.80	94.20	94.20	90.40	95.20	96.60	94.70
IMP	bias	−0.56	0.52	−0.59	0.11	−0.05	0.55	−0.62	0.42	−0.04	−0.10
	std	2.60	2.53	2.22	1.98	1.85	2.93	2.50	2.25	2.01	1.67
CC	bias	−0.43	0.43	−0.08	0.24	−0.11	0.48	−0.39	0.28	−0.18	0.02
	std	3.14	3.35	2.90	2.45	2.42	3.54	3.39	2.78	2.26	2.14
	$\widehat{\text{std}}$	2.89	2.94	2.43	2.24	2.11	3.33	2.90	2.53	2.19	1.87
	cvp	92.30	89.10	88.10	92.80	90.50	92.50	90.40	92.30	92.80	92.80
Method	$\boldsymbol{\beta}_2$	$\widehat{\beta}_{2,3}$	$\widehat{\beta}_{2,4}$	$\widehat{\beta}_{2,5}$	$\widehat{\beta}_{2,6}$	$\widehat{\beta}_{2,7}$	$\widehat{\beta}_{2,8}$	$\widehat{\beta}_{2,9}$	$\widehat{\beta}_{2,10}$	$\widehat{\beta}_{2,11}$	$\widehat{\beta}_{2,12}$
		−0.8	0.6	−0.4	0.2	0	0.8	−0.6	0.4	−0.2	0
Omni	bias	0.46	−0.33	0.24	−0.11	0.00	−0.26	0.20	−0.14	0.05	0.03
	std	1.83	1.74	1.47	1.41	1.32	1.95	1.72	1.42	1.33	1.04
	$\widehat{\text{std}}$	1.80	1.77	1.55	1.36	1.32	2.06	1.75	1.52	1.37	1.16
	cvp	91.90	95.20	94.20	93.30	91.40	94.20	93.30	96.10	92.80	95.20
IPW	bias	0.49	−0.39	0.27	−0.12	−0.01	−0.28	0.21	−0.15	0.05	0.02
	std	1.84	1.74	1.48	1.42	1.31	2.01	1.76	1.46	1.36	1.08
	$\widehat{\text{std}}$	1.78	1.75	1.53	1.35	1.30	2.04	1.73	1.50	1.36	1.15
	cvp	90.90	94.20	94.70	94.20	91.90	92.30	92.80	94.50	92.30	95.20
IMP	bias	0.58	−0.53	0.20	−0.21	−0.07	−0.71	0.68	−0.55	0.24	−0.10
	std	1.78	1.65	1.41	1.37	1.28	1.88	1.72	1.56	1.40	1.23
CC	bias	0.62	−0.57	0.25	−0.19	−0.01	−0.43	0.36	−0.20	0.09	−0.07
	std	2.13	2.13	1.77	1.67	1.61	2.35	2.06	1.78	1.60	1.38
	$\widehat{\text{std}}$	2.00	2.03	1.67	1.53	1.45	2.28	1.99	1.73	1.50	1.28
	cvp	92.80	92.80	92.30	92.80	92.30	93.30	92.30	93.40	93.00	94.20

where $\mathscr{D}(A) = (a_{11}, a_{22}, \ldots, a_{qq})^\top$ for any matrix $A = (a_{ij})_{q\times q}$. Then Eq. (2.2) can be rewritten as

$$\sum_{i=1}^{n}\{\mathbf{y}_i - \widehat{\mathbf{m}}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i)\}^\top \widehat{\mathbf{W}}_i\{\mathbf{y}_i - \widehat{\mathbf{m}}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i)\} \approx \sum_{i=1}^{n}\Big[\mathscr{D}\big\{\mathbf{m}^{(1)}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i)(\boldsymbol{\beta}_{-d} - \widehat{\boldsymbol{\beta}}_{-d})^\top \widehat{\widetilde{\mathbf{x}}}_{-d,i}\big\}$$
$$+ \{\boldsymbol{\varepsilon}_i - \widehat{E}(\boldsymbol{\varepsilon}_i|\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i)\}\Big]^\top \widehat{\mathbf{W}}_i\Big[\mathscr{D}\big\{\mathbf{m}^{(1)}(\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i)(\boldsymbol{\beta}_{-d} - \widehat{\boldsymbol{\beta}}_{-d})^\top \widehat{\widetilde{\mathbf{x}}}_{-d,i}\big\} + \{\boldsymbol{\varepsilon}_i - \widehat{E}(\boldsymbol{\varepsilon}_i|\widehat{\boldsymbol{\beta}}^\top \mathbf{x}_i)\}\Big],$$

Table 11
Simulation results for Model III with $n = 500$ and MR $=$ 10%: the average bias of the estimators ("bias"), the Monte Carlo standard deviation ("std"), the average of the estimated standard deviation ("$\widehat{\text{std}}$") based on the theoretical calculation, and the empirical coverage probability ("cvp") at the nominal 95% confidence level. All simulation results reported below are multiplied by 100. The methods Omni, IPW and CC are described in Section 3.1. The parameters $\widehat{\beta}_{ij}$ are defined in (2.2).

Method	$\boldsymbol{\beta}_1$	$\widehat{\beta}_{1,3}$	$\widehat{\beta}_{1,4}$	$\widehat{\beta}_{1,5}$	$\widehat{\beta}_{1,6}$	$\widehat{\beta}_{1,7}$	$\widehat{\beta}_{1,8}$	$\widehat{\beta}_{1,9}$	$\widehat{\beta}_{1,10}$	$\widehat{\beta}_{1,11}$	$\widehat{\beta}_{1,12}$
		0.8	−0.6	0.4	−0.2	0	−0.8	0.6	−0.4	0.2	0
Omni	bias	−0.14	0.12	0.12	−0.01	−0.01	0.30	−0.35	0.23	−0.12	0.19
	std	1.87	1.88	1.56	1.42	1.34	2.32	1.89	1.50	1.48	1.31
	$\widehat{\text{std}}$	1.86	1.84	1.61	1.42	1.36	2.15	1.83	1.59	1.43	1.22
	cvp	93.30	93.50	94.70	97.60	95.70	91.90	93.80	96.10	95.20	92.30
IPW	bias	−0.14	0.12	0.13	−0.02	0.02	0.28	−0.36	0.24	−0.12	0.19
	std	1.88	1.90	1.58	1.44	1.35	2.32	1.88	1.52	1.48	1.33
	$\widehat{\text{std}}$	1.85	1.83	1.60	1.41	1.35	2.14	1.82	1.58	1.42	1.21
	cvp	93.30	92.80	93.80	96.10	95.70	92.30	92.80	96.10	94.20	92.80
IMP	bias	−0.33	0.26	−0.22	−0.01	−0.24	0.37	−0.07	0.12	−0.13	−0.08
	std	1.88	1.81	1.67	1.50	1.57	2.18	1.90	1.61	1.57	1.22
CC	bias	−0.23	0.14	0.02	0.01	−0.10	0.38	−0.46	0.34	−0.07	0.21
	std	2.46	2.37	1.96	1.76	1.74	2.79	2.26	1.98	1.89	1.67
	$\widehat{\text{std}}$	2.29	2.32	1.94	1.78	1.65	2.64	2.32	2.03	1.74	1.49
	cvp	93.30	93.30	94.70	97.10	93.50	95.70	94.20	96.10	91.90	90.90
Method	$\boldsymbol{\beta}_2$	$\widehat{\beta}_{2,3}$	$\widehat{\beta}_{2,4}$	$\widehat{\beta}_{2,5}$	$\widehat{\beta}_{2,6}$	$\widehat{\beta}_{2,7}$	$\widehat{\beta}_{2,8}$	$\widehat{\beta}_{2,9}$	$\widehat{\beta}_{2,10}$	$\widehat{\beta}_{2,11}$	$\widehat{\beta}_{2,12}$
		−0.8	0.6	−0.4	0.2	0	0.8	−0.6	0.4	−0.2	0
Omni	bias	0.24	−0.21	0.14	−0.11	0.07	−0.31	0.27	−0.15	0.08	−0.07
	std	1.29	1.21	1.07	1.01	0.94	1.40	1.30	1.17	1.09	0.86
	$\widehat{\text{std}}$	1.39	1.37	1.19	1.05	1.01	1.59	1.36	1.17	1.06	0.91
	cvp	97.00	95.70	97.10	95.70	95.70	97.10	95.20	94.20	90.90	96.10
IPW	bias	0.25	−0.22	0.13	−0.11	0.07	−0.32	0.29	−0.18	0.09	−0.08
	std	1.31	1.22	1.08	1.01	0.93	1.40	1.31	1.18	1.08	0.86
	$\widehat{\text{std}}$	1.39	1.37	1.19	1.05	1.00	1.59	1.35	1.17	1.05	0.90
	cvp	96.10	95.70	96.60	95.70	96.10	96.60	93.80	92.30	91.90	95.20
IMP	bias	0.41	−0.32	0.18	−0.25	0.09	−0.58	0.42	−0.26	0.18	−0.10
	std	1.44	1.47	1.13	1.05	1.01	1.61	1.47	1.23	1.05	0.87
CC	bias	0.37	−0.26	0.19	−0.23	0.08	−0.44	0.41	−0.24	0.12	−0.09
	std	1.44	1.47	1.24	1.18	1.18	1.66	1.54	1.35	1.29	1.06
	$\widehat{\text{std}}$	1.54	1.56	1.30	1.19	1.12	1.77	1.55	1.35	1.17	1.00
	cvp	97.10	96.50	96.10	94.20	93.80	97.60	95.70	95.20	93.30	94.20

which yields

$$\begin{aligned}\mathrm{vec}(\boldsymbol{\beta}_{-d}) - \mathrm{vec}(\widehat{\boldsymbol{\beta}}_{-d}) &= \Bigg[\sum_{i=1}^{n}\Big\{\mathbf{m}_1^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widehat{\mathbf{x}}}_{-d,i1},\ldots,\mathbf{m}_q^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widehat{\mathbf{x}}}_{-d,iq}\Big\}\widehat{\mathbf{W}}_i\\ &\quad\times\Big\{\mathbf{m}_1^{(1)}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widehat{\mathbf{x}}}_{-d,i1}^\top,\ldots,\mathbf{m}_q^{(1)}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widehat{\mathbf{x}}}_{-d,iq}^\top\Big\}\Bigg]^{-1}\cdot\Bigg[\sum_{i=1}^{n}\Big\{\mathbf{m}_1^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widehat{\mathbf{x}}}_{-d,i1},\\ &\quad\ldots,\mathbf{m}_q^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widehat{\mathbf{x}}}_{-d,iq}\Big\}\widehat{\mathbf{W}}_i\Big\{\boldsymbol{\varepsilon}_i - \widehat{E}(\boldsymbol{\varepsilon}_i|\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\Big\}\Bigg] + o_p(\|\boldsymbol{\beta}_{-d}-\widehat{\boldsymbol{\beta}}_{-d}\|).\end{aligned}\tag{B.1}$$

Let $\widehat{W}_{ik}$ be the (k,k)th element of $\widehat{\mathbf{W}}_i$. Note that $\widehat{W}_{ik}$ can be decomposed as $\delta_{ik}/\widehat{\pi}_{ik} = \delta_{ik}/\pi_{ik} - (\widehat{\pi}_{ik}-\pi_{ik})\delta_{ik}/\pi_{ik}^2 + (\widehat{\pi}_{ik}-\pi_{ik})^2\delta_{ik}/(\widehat{\pi}_{ik}\pi_{ik}^2)$ for $k \in \{1,\ldots,q\}$. Because $\pi_{ik} = \pi_k(\mathbf{x}_i, \boldsymbol{\gamma}_k)$ comes from the logistic regression model and the MLEs $\widehat{\boldsymbol{\gamma}}_k$ are root-n consistent estimators of $\boldsymbol{\gamma}_k$ for $k \in \{1,\ldots,q\}$, by Taylor expansion, we can obtain that $\widehat{\pi}_{ik} - \pi_{ik} = \pi_{ik}(1-\pi_{ik})\mathbf{x}_i^\top(\widehat{\boldsymbol{\gamma}}_k - \boldsymbol{\gamma}_k)\{1+o_p(1)\}$, and further

$$\widehat{W}_{ik} - W_{ik} = \delta_{ik}(1-\pi_{ik})\mathbf{x}_i^\top(\widehat{\boldsymbol{\gamma}}_k-\boldsymbol{\gamma}_k)/\pi_{ik}\{1+o_p(1)\},\tag{B.2}$$

where $\widehat{W}_{ik} = \delta_{ik}/\widehat{\pi}_{ik}$ and $W_{ik} = \delta_{ik}/\pi_{ik}$ for $k \in \{1,\ldots,q\}$. Write

$$\begin{aligned}&\sum_{i=1}^{n}\Big\{\mathbf{m}_1^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widehat{\mathbf{x}}}_{-d,i1},\ldots,\mathbf{m}_q^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widehat{\mathbf{x}}}_{-d,iq}\Big\}\widehat{\mathbf{W}}_i\Big\{\boldsymbol{\varepsilon}_i-\widehat{E}(\boldsymbol{\varepsilon}_i|\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\Big\}\\ &\quad=\sum_{i=1}^{n}\Big\{\mathbf{m}_1^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widehat{\mathbf{x}}}_{-d,i1},\ldots,\mathbf{m}_q^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widehat{\mathbf{x}}}_{-d,iq}\Big\}\mathbf{W}_i\Big\{\boldsymbol{\varepsilon}_i-\widehat{E}(\boldsymbol{\varepsilon}_i|\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\Big\}\end{aligned}$$

Table 12
Simulation results for Model III with $n = 300$ and MR = 30%: the average bias of the estimators ("bias"), the Monte Carlo standard deviation ("std"), the average of the estimated standard deviation ("$\widehat{\text{std}}$") based on the theoretical calculation, and the empirical coverage probability ("cvp") at the nominal 95% confidence level. All simulation results reported below are multiplied by 100. The methods Omni, IPW and CC are described in Section 3.1. The parameters $\widehat{\beta}_{ij}$ are defined in (2.2).

Method	$\boldsymbol{\beta}_1$	$\widehat{\beta}_{1,3}$	$\widehat{\beta}_{1,4}$	$\widehat{\beta}_{1,5}$	$\widehat{\beta}_{1,6}$	$\widehat{\beta}_{1,7}$	$\widehat{\beta}_{1,8}$	$\widehat{\beta}_{1,9}$	$\widehat{\beta}_{1,10}$	$\widehat{\beta}_{1,11}$	$\widehat{\beta}_{1,12}$
		0.8	−0.6	0.4	−0.2	0	−0.8	0.6	−0.4	0.2	0
Omni	bias	−0.14	−0.04	−0.07	−0.04	−0.12	0.58	−0.53	0.26	0.05	−0.10
	std	2.45	2.25	1.98	1.84	1.74	3.04	2.42	2.07	1.76	1.48
	$\widehat{\text{std}}$	2.63	2.54	2.23	2.04	1.92	2.96	2.55	2.20	1.99	1.69
	cvp	93.50	95.20	94.70	95.20	96.10	92.30	93.80	90.90	95.70	95.20
IPW	bias	−0.12	−0.09	−0.04	−0.08	−0.13	0.50	−0.45	0.27	0.08	−0.08
	std	2.48	2.25	2.01	1.85	1.77	3.05	2.46	2.13	1.80	1.50
	$\widehat{\text{std}}$	2.59	2.54	2.20	2.02	1.90	2.94	2.50	2.18	1.95	1.68
	cvp	93.80	95.20	94.70	95.20	96.10	93.30	92.80	92.30	94.20	94.70
IMP	bias	−0.81	0.91	−1.24	0.39	0.00	1.06	−0.27	−0.17	0.37	−0.37
	std	3.16	3.69	3.41	2.70	3.17	4.28	3.80	3.65	3.55	3.34
CC	bias	−1.19	0.86	−0.94	1.03	−0.44	0.43	−0.45	0.57	−0.23	−0.64
	std	8.38	8.65	8.18	6.53	8.22	9.05	8.63	8.23	7.65	6.82
	$\widehat{\text{std}}$	6.27	6.52	5.51	5.11	5.71	7.12	6.51	6.41	5.62	5.24
	cvp	86.10	84.70	80.90	84.20	84.20	87.60	83.80	87.10	85.20	83.30
Method	$\boldsymbol{\beta}_2$	$\widehat{\beta}_{2,3}$	$\widehat{\beta}_{2,4}$	$\widehat{\beta}_{2,5}$	$\widehat{\beta}_{2,6}$	$\widehat{\beta}_{2,7}$	$\widehat{\beta}_{2,8}$	$\widehat{\beta}_{2,9}$	$\widehat{\beta}_{2,10}$	$\widehat{\beta}_{2,11}$	$\widehat{\beta}_{2,12}$
		−0.8	0.6	−0.4	0.2	0	0.8	−0.6	0.4	−0.2	0
Omni	bias	0.02	0.08	0.06	0.02	−0.08	−0.18	0.21	−0.05	−0.25	−0.01
	std	2.02	1.85	1.63	1.46	1.35	2.19	1.96	1.84	1.55	1.21
	$\widehat{\text{std}}$	2.15	2.09	1.84	1.69	1.61	2.36	2.07	1.82	1.60	1.42
	cvp	93.30	93.80	94.00	92.30	92.80	92.80	93.80	89.50	92.40	94.20
IPW	bias	0.08	0.03	0.07	0.01	−0.06	−0.24	0.20	−0.08	−0.24	0.01
	std	2.03	1.80	1.60	1.50	1.34	2.19	1.93	1.87	1.53	1.21
	$\widehat{\text{std}}$	2.17	2.09	1.81	1.72	1.60	2.37	2.07	1.80	1.58	1.42
	cvp	93.50	96.60	93.30	91.00	93.80	91.90	90.90	88.10	91.90	93.80
IMP	bias	1.59	−1.84	0.45	−0.78	−0.57	−1.51	0.85	−0.92	0.75	0.54
	std	3.60	3.96	2.87	2.95	2.98	3.63	3.43	3.52	3.37	2.59
CC	bias	0.84	−1.29	1.36	−0.76	−0.16	−0.86	0.38	−1.57	0.47	0.49
	std	6.49	7.29	6.92	5.57	6.43	7.21	6.49	7.48	5.96	5.88
	$\widehat{\text{std}}$	5.36	5.63	4.72	4.38	4.85	6.13	5.57	5.46	4.83	4.48
	cvp	90.00	83.30	86.60	89.10	90.00	90.90	91.40	88.10	89.50	89.50

$$
\begin{aligned}
&+\sum_{i=1}^{n}\left\{\mathbf{m}_1^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widetilde{\mathbf{x}}}_{-d,i1},\ldots,\mathbf{m}_q^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widetilde{\mathbf{x}}}_{-d,iq}\right\}(\widehat{\mathbf{W}}_i-\mathbf{W}_i)\left\{\boldsymbol{\varepsilon}_i-\widehat{E}(\boldsymbol{\varepsilon}_i|\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\right\}\\
&:=A_{1n}+A_{2n}. \qquad \text{(B.3)}
\end{aligned}
$$

Applying the same approaches used in [28], one can derive that

$$
n^{-1/2}A_{1n}=n^{-1/2}\sum_{i=1}^{n}\left\{\mathbf{m}_1^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widetilde{\mathbf{x}}}_{-d,i1},\ldots,\mathbf{m}_q^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widetilde{\mathbf{x}}}_{-d,iq}\right\}\mathbf{W}_i\boldsymbol{\varepsilon}_i+o_p(1)\xrightarrow{d}\mathcal{N}(\mathbf{0},\boldsymbol{\Omega}_2).
$$

It is easy to see that

$$
\begin{aligned}
A_{2n}=\sum_{i=1}^{n}\Bigg[&\left\{\mathbf{m}_1^{(1)}(\boldsymbol{\beta}^\top\mathbf{x}_i)+\mathbf{m}_1^{(1)}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)-\mathbf{m}_1^{(1)}(\boldsymbol{\beta}^\top\mathbf{x}_i)\right\}^\top\otimes\Big\{\mathbf{x}_{-d,i}-E(\mathbf{x}_{-d,i}|\boldsymbol{\beta}^\top\mathbf{x}_i)\\
&+E(\mathbf{x}_{-d,i}|\boldsymbol{\beta}^\top\mathbf{x}_i)-\widehat{E}_1(\mathbf{x}_{-d,i}|\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\Big\},\ldots,\left\{\mathbf{m}_q^{(1)}(\boldsymbol{\beta}^\top\mathbf{x}_i)+\mathbf{m}_q^{(1)}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)-\mathbf{m}_q^{(1)}(\boldsymbol{\beta}^\top\mathbf{x}_i)\right\}^\top\\
&\otimes\left\{\mathbf{x}_{-d,i}-E(\mathbf{x}_{-d,i}|\boldsymbol{\beta}^\top\mathbf{x}_i)+E(\mathbf{x}_{-d,i}|\boldsymbol{\beta}^\top\mathbf{x}_i)-\widehat{E}_q(\mathbf{x}_{-d,i}|\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\right\}\Bigg](\widehat{\mathbf{W}}_i-\mathbf{W}_i)\\
&\cdot\left\{\boldsymbol{\varepsilon}_i-\widehat{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{\beta}^\top\mathbf{x}_i)+\widehat{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{\beta}^\top\mathbf{x}_i)-\widehat{E}(\boldsymbol{\varepsilon}_i|\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\right\}.
\end{aligned}
$$

Table 13
Simulation results for Model III with $n = 500$ and MR $= 30\%$: the average bias of the estimators ("bias"), the Monte Carlo standard deviation ("std"), the average of the estimated standard deviation ("$\widehat{\text{std}}$") based on the theoretical calculation, and the empirical coverage probability ("cvp") at the nominal 95% confidence level. All simulation results reported below are multiplied by 100. The methods Omni, IPW and CC are described in Section 3.1. The parameters $\widehat{\beta}_{ij}$ are defined in (2.2).

Method	$\boldsymbol{\beta}_1$	$\widehat{\beta}_{1,3}$	$\widehat{\beta}_{1,4}$	$\widehat{\beta}_{1,5}$	$\widehat{\beta}_{1,6}$	$\widehat{\beta}_{1,7}$	$\widehat{\beta}_{1,8}$	$\widehat{\beta}_{1,9}$	$\widehat{\beta}_{1,10}$	$\widehat{\beta}_{1,11}$	$\widehat{\beta}_{1,12}$
		0.8	−0.6	0.4	−0.2	0	−0.8	0.6	−0.4	0.2	0
Omni	bias	−0.21	0.35	−0.19	0.12	−0.23	0.66	−0.48	0.32	−0.13	0.04
	std	1.87	1.78	1.45	1.28	1.33	2.27	1.92	1.50	1.29	1.05
	$\widehat{\text{std}}$	1.99	1.95	1.65	1.50	1.42	2.23	1.90	1.60	1.47	1.26
	cvp	93.80	95.70	98.10	96.10	95.20	91.90	93.30	94.00	96.60	97.60
IPW	bias	−0.36	0.29	−0.08	0.13	−0.14	0.43	−0.31	0.26	−0.24	−0.06
	std	2.01	2.04	1.66	1.43	1.36	2.13	1.84	1.44	1.31	1.21
	$\widehat{\text{std}}$	1.96	1.92	1.63	1.50	1.42	2.21	1.89	1.60	1.46	1.24
	cvp	92.80	90.40	95.20	95.70	95.20	94.20	94.00	95.70	96.10	95.70
IMP	bias	−0.49	0.30	−0.45	−0.14	−0.09	0.47	0.14	−0.40	0.24	−0.46
	std	2.31	2.66	2.33	2.04	2.28	3.04	2.82	2.66	2.30	2.01
CC	bias	−0.05	0.46	−0.21	0.24	−0.70	1.28	−0.65	0.60	−0.60	−0.05
	std	6.24	6.90	5.67	5.10	5.69	6.97	6.22	6.48	5.84	5.21
	$\widehat{\text{std}}$	5.15	5.32	4.46	4.11	4.44	5.88	5.43	5.12	4.51	4.03
	cvp	90.40	86.60	88.50	89.50	86.10	90.00	90.40	86.10	87.10	86.10
Method	$\boldsymbol{\beta}_2$	$\widehat{\beta}_{2,3}$	$\widehat{\beta}_{2,4}$	$\widehat{\beta}_{2,5}$	$\widehat{\beta}_{2,6}$	$\widehat{\beta}_{2,7}$	$\widehat{\beta}_{2,8}$	$\widehat{\beta}_{2,9}$	$\widehat{\beta}_{2,10}$	$\widehat{\beta}_{2,11}$	$\widehat{\beta}_{2,12}$
		−0.8	0.6	−0.4	0.2	0	0.8	−0.6	0.4	−0.2	0
Omni	bias	−0.08	0.09	0.16	−0.08	0.13	−0.19	0.15	−0.03	−0.02	0.00
	std	1.55	1.47	1.23	1.12	1.06	1.78	1.44	1.38	1.15	0.97
	$\widehat{\text{std}}$	1.67	1.66	1.36	1.22	1.16	1.88	1.61	1.41	1.24	1.08
	cvp	93.40	92.80	93.30	91.90	91.90	93.30	92.30	91.00	91.90	95.20
IPW	bias	0.29	−0.18	0.09	−0.08	0.19	−0.50	0.33	−0.14	0.04	0.06
	std	1.52	1.52	1.33	1.14	1.07	1.82	1.45	1.28	1.08	0.92
	$\widehat{\text{std}}$	1.61	1.59	1.33	1.22	1.16	1.82	1.56	1.36	1.20	1.04
	cvp	94.20	92.80	92.30	94.20	94.70	91.40	92.80	92.30	92.30	94.70
IMP	bias	1.53	−1.69	0.35	−0.36	−0.32	−1.46	0.86	−0.77	0.86	0.35
	std	2.57	2.81	1.91	1.74	1.98	2.57	2.22	2.18	2.08	1.83
CC	bias	0.62	−0.43	0.21	−0.37	0.11	−0.91	0.76	−0.79	0.74	0.18
	std	4.54	4.95	3.93	3.75	3.95	4.93	4.59	4.55	4.31	3.57
	$\widehat{\text{std}}$	4.05	4.21	3.47	3.23	3.51	4.58	4.24	4.03	3.58	3.24
	cvp	93.30	92.80	94.70	90.90	91.40	92.30	92.80	92.80	89.50	90.00

Table 14
The MSE results for Model III with different sample sizes and missing rates (MRs). All simulation results reported below are multiplied by 100. The methods Omni, IPW, IMP and CC are described in Section 3.1.

Model	n	Case 1: MR = 10%				Case 2: MR = 30%			
		Omni	IPW	IMP	CC	Omni	IPW	IMP	CC
III	300	0.7251	0.7304	0.7600	1.1785	0.7609	0.7702	2.3225	10.8675
	500	0.4153	0.4190	0.4502	0.6309	0.4389	0.4594	1.0977	5.5542

Together with Lemma B.1, $E(\boldsymbol{\varepsilon}_i|\mathbf{x}_i) = \mathbf{0}$ and (B.2), one can derive that $\left\| \sum_{i=1}^n \{\mathbf{m}^{(1)}(\boldsymbol{\beta}^\top\mathbf{x}_i)\otimes\mathbf{x}_{-d,i}\}(\widehat{\mathbf{W}}_i-\mathbf{W}_i)\boldsymbol{\varepsilon}_i \right\| = O_p(\ln n) = o_p(n^{1/2})$. By Taylor expansion, it follows that

$$\sum_{i=1}^n \{\mathbf{m}^{(1)^\top}(\boldsymbol{\beta}^\top\mathbf{x}_i)\otimes\mathbf{x}_{-d,i}\}(\widehat{\mathbf{W}}_i-\mathbf{W}_i)\{\widehat{E}(\boldsymbol{\varepsilon}_i|\boldsymbol{\beta}^\top\mathbf{x}_i) - \widehat{E}(\boldsymbol{\varepsilon}_i|\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\}$$
$$= \sum_{i=1}^n \{\mathbf{m}^{(1)^\top}(\boldsymbol{\beta}^\top\mathbf{x}_i)\otimes\mathbf{x}_{-d,i}\}(\widehat{\mathbf{W}}_i-\mathbf{W}_i)\widehat{E}^{(1)}(\boldsymbol{\varepsilon}_i|\boldsymbol{\beta}^{*^\top}\mathbf{x}_i)(\widehat{\boldsymbol{\beta}}_{-d}-\boldsymbol{\beta}_{-d})^\top\mathbf{x}_{-d,i} = o_p(n^{1/2}).$$

Similarly, one can derive the other terms in A_{2n} to be $o_p(n^{1/2})$. Thus, by (B.3), we have

$$n^{-1/2}\{\mathbf{m}_1^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widetilde{\mathbf{x}}}_{-d,i1},\ldots,\mathbf{m}_q^{(1)^\top}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\otimes\widehat{\widetilde{\mathbf{x}}}_{-d,iq}\}\widehat{\mathbf{W}}_i\{\boldsymbol{\varepsilon}_i - \widehat{E}(\boldsymbol{\varepsilon}_i|\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_2).$$

Analogously, one can show easily that n^{-1} times the quantity in the first square brackets in (B.1) converges in probability to $\boldsymbol{\Omega}_1$. Then by the Slutsky Theorem, the result of Theorem 2.1 is obtained. □

Table 15
The frequency (%) of the estimated structural dimension $\widehat{d}$ for Model III with different sample sizes and missing rates (MRs). The methods Omni, IPW and CC are described in Section 3.1.

n	MR	Method	$\widehat{d}=1$	$\widehat{d}=2$	$\widehat{d}=3$	$\widehat{d}\geq 4$
300	10%	Omni	0.00	100.00	0.00	0.00
		IPW	0.00	100.00	0.00	0.00
		CC	0.00	100.00	0.00	0.00
	30%	Omni	0.00	100.00	0.00	0.00
		IPW	0.00	100.00	0.00	0.00
		CC	0.50	95.50	0.00	0.00
500	10%	Omni	0.00	100.00	0.00	0.00
		IPW	0.00	100.00	0.00	0.00
		CC	0.00	100.00	0.00	0.00
	30%	Omni	0.00	100.00	0.00	0.00
		IPW	0.00	100.00	0.00	0.00
		CC	0.00	100.00	0.00	0.00

Table 16
The frequency (%) of the estimated structural dimension $\widehat{d}$ for Models I–III with $n = 500$ and 30% missing rate for different (α, κ). The methods Omni, IPW and CC are described in Section 3.1.

Model	(α,κ)	Omni				IPW				CC			
		$\widehat{d}=1$	$\widehat{d}=2$	$\widehat{d}=3$	$\widehat{d}\geq 4$	$\widehat{d}=1$	$\widehat{d}=2$	$\widehat{d}=3$	$\widehat{d}\geq 4$	$\widehat{d}=1$	$\widehat{d}=2$	$\widehat{d}=3$	$\widehat{d}\geq 4$
I	(0.60, 0.05)	0.00	86.30	0.00	13.70	0.00	87.20	0.00	12.80	0.00	83.60	2.70	13.70
	(1.00, 0.05)	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00	8.60	91.40	0.00	0.00
	(1.00, 0.08)	2.40	97.60	0.00	0.00	1.40	98.60	0.00	0.00	47.60	52.40	0.00	0.00
	(1.00, 0.12)	27.10	72.90	0.00	0.00	27.60	72.40	0.00	0.00	94.80	5.20	0.00	0.00
II	(0.35, 0.05)	0.00	0.00	31.80	68.20	0.00	0.00	35.50	64.50	0.00	0.00	20.00	80.00
	(0.40, 0.05)	0.00	0.00	74.50	25.50	0.00	0.00	75.40	24.60	0.00	0.00	63.70	36.30
	(0.50, 0.05)	0.00	0.00	100.00	0.00	0.00	0.00	99.10	0.90	0.00	0.00	98.20	1.80
	(1.00, 0.05)	0.00	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	100.00	0.00
	(1.00, 0.08)	0.00	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	100.00	0.00
	(1.00, 0.12)	0.00	0.50	99.50	0.00	0.00	0.90	99.10	0.00	0.00	0.50	99.50	0.00
	(1.00, 0.15)	0.00	23.70	76.30	0.00	0.00	29.10	70.10	0.00	0.00	16.40	83.60	0.00
III	(0.20, 0.05)	0.00	91.80	0.00	8.20	0.00	93.60	0.00	6.40	0.00	73.60	14.60	11.80
	(0.50, 0.05)	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00
	(1.00, 0.05)	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00
	(1.00, 0.08)	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00
	(1.00, 0.12)	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00
	(1.00, 0.15)	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00
	(1.00, 0.20)	0.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00	4.60	95.40	0.00	0.00
	(1.00, 0.28)	12.70	87.30	0.00	0.00	15.40	84.60	0.00	0.00	34.50	65.50	0.00	0.00
	(1.00, 0.30)	35.50	64.50	0.00	0.00	31.80	68.20	0.00	0.00	48.10	51.90	0.00	0.00

Appendix C. Proof of Theorem 2.2

Observe that $\widehat{\boldsymbol{\Omega}}_1 = n^{-1}\sum_{i=1}^n(\mathbf{B}_{i1},\ldots,\mathbf{B}_{iq})(\mathbf{W}_i+\widehat{\mathbf{W}}_i-\mathbf{W}_i)(\mathbf{B}_{i1},\ldots,\mathbf{B}_{iq})^\top$, where $\mathbf{B}_{ik}$ is the kth column of $\mathbf{B}_i$ which can be rewritten as

$$\begin{aligned}\mathbf{B}_{ik} =&\Big[\mathbf{m}_k^{(1)^\top}(\boldsymbol{\beta}^\top\mathbf{x}_i)+\widehat{\mathbf{m}}_k^{(1),\top}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)-\mathbf{m}_k^{(1)^\top}(\boldsymbol{\beta}^\top\mathbf{x}_i)\Big]\otimes\Big[\widetilde{\mathbf{x}}_{-d,i}+\Big\{E(\mathbf{x}_{-d,i}|\boldsymbol{\beta}^\top\mathbf{x}_i)\\ &-\widehat{E}_k(\mathbf{x}_{-d,i}|\boldsymbol{\beta}^\top\mathbf{x}_i)\Big\}+\Big(\widehat{E}_k(\mathbf{x}_{-d,i}|\boldsymbol{\beta}^\top\mathbf{x}_i)-\widehat{E}_k(\mathbf{x}_{-d,i}|\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)\Big)\Big].\end{aligned}$$

It can be concluded from the weak law of large numbers that $n^{-1}\sum_{i=1}^n\big\{\mathbf{m}^{(1)^\top}(\boldsymbol{\beta}^\top\mathbf{x}_i)\otimes\widetilde{\mathbf{x}}_{-d,i}\big\}\mathbf{W}_i\big\{\mathbf{m}^{(1)}(\boldsymbol{\beta}^\top\mathbf{x}_i)\otimes\widetilde{\mathbf{x}}_{-d,i}^\top\big\} = \boldsymbol{\Omega}_1+o_p(1)$. Following the proof of Lemma A.1 in [16], applying Taylor expansion to $\widehat{E}(\mathbf{x}_{-d,i}|\boldsymbol{\beta}^\top\mathbf{x}_i)-\widehat{E}(\mathbf{x}_{-d,i}|\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)$, and by $|\widehat{\mathbf{m}}^{(1)}(\widehat{\boldsymbol{\beta}}^\top\mathbf{x}_i)-\mathbf{m}^{(1)}(\boldsymbol{\beta}^\top\mathbf{x}_i)| = o_p(1)$ and (B.2), we can obtain that the rest terms of $\widehat{\boldsymbol{\Omega}}_1$ are all $o_p(1)$. Similarly, one can prove $\widehat{\boldsymbol{\Omega}}_2 \xrightarrow{p} \boldsymbol{\Omega}_2$. Thus, the proof of Theorem 2.2 is completed. □

Appendix D. Proof of Theorem 2.3

Let $\mathcal{L}_1(d)=\sum_{i=1}^n\{\mathbf{y}_i-\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_d^\top\mathbf{x}_i)\}^\top\widehat{\mathbf{W}}_i\{\mathbf{y}_i-\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_d^\top\mathbf{x}_i)\}$. By definition, we write

$$\begin{aligned}\mathcal{L}_1(d)-\mathcal{L}_1(d_0)&=\sum_{i=1}^n\{\mathbf{y}_i-\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_d^\top\mathbf{x}_i)\}^\top\widehat{\mathbf{W}}_i\{\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_{d_0}^\top\mathbf{x}_i)-\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_d^\top\mathbf{x}_i)\}+\sum_{i=1}^n\{\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_{d_0}^\top\mathbf{x}_i)-\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_d^\top\mathbf{x}_i)\}^\top\widehat{\mathbf{W}}_i\{\mathbf{y}_i-\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_{d_0}^\top\mathbf{x}_i)\}\\&:=\Lambda_1+\Lambda_2.\end{aligned}$$

Note that $E(\mathbf{y}_i|\boldsymbol\beta_d^\top\mathbf{x}_i)\neq E(\mathbf{y}_i|\boldsymbol\beta_{d_0}^\top\mathbf{x})$ if $d<d_0$ and $E(\mathbf{y}_i|\boldsymbol\beta_d^\top\mathbf{x}_i)=E(\mathbf{y}_i|\boldsymbol\beta_{d_0}^\top\mathbf{x}_i)$ otherwise. Then by nonparametric regression and Condition (C2), it can be verified that

$$\mathbf{m}(\widehat{\boldsymbol\beta}_{d_0}^\top\mathbf{x}_i)-\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_{d_0}^\top\mathbf{x}_i)=O_p[h^s+\{\ln n/(nh^d)\}^{1/2}]. \tag{D.1}$$

If $d<d_0$, $\Lambda_1=\sum_{i=1}^n\{\mathbf{m}(\boldsymbol\beta_{d_0}^\top\mathbf{x}_i)-\mathbf{m}(\boldsymbol\beta_d^\top\mathbf{x}_i)\}^\top\mathbf{W}_i\{\mathbf{m}(\boldsymbol\beta_{d_0}^\top\mathbf{x}_i)-\mathbf{m}(\boldsymbol\beta_d^\top\mathbf{x}_i)\}+o_p(n)$. The first term of Λ_1 is $O_p(n)$ and is positive. For Λ_2, by (D.1), we can derive

$$\begin{aligned}\Lambda_2&=\sum_{i=1}^n\{\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_{d_0}^\top\mathbf{x}_i)-\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_d^\top\mathbf{x}_i)\}^\top\widehat{\mathbf{W}}_i\big[\{\mathbf{m}(\widehat{\boldsymbol\beta}_{d_0}^\top\mathbf{x}_i)-\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_{d_0}^\top\mathbf{x}_i)\}+\boldsymbol\varepsilon_i\big]\\&=\sum_{i=1}^n\{\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_{d_0}^\top\mathbf{x}_i)-\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_d^\top\mathbf{x}_i)\}^\top\widehat{\mathbf{W}}_i\boldsymbol\varepsilon_i+\sum_{i=1}^n\{\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_{d_0}^\top\mathbf{x}_i)-\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_d^\top\mathbf{x}_i)\}^\top\widehat{\mathbf{W}}_i\{\mathbf{m}(\widehat{\boldsymbol\beta}_{d_0}^\top\mathbf{x}_i)-\widehat{\mathbf{m}}(\widehat{\boldsymbol\beta}_{d_0}^\top\mathbf{x}_i)\}\\&=\sum_{i=1}^n\{\mathbf{m}(\boldsymbol\beta_{d_0}^\top\mathbf{x}_i)-\mathbf{m}(\boldsymbol\beta_d^\top\mathbf{x}_i)\}^\top\mathbf{W}_i\boldsymbol\varepsilon_i+o_p(n)=o_p(n).\end{aligned}$$

Hence from $\lambda_n/\sqrt{n}\to 0$ and Condition (C4), if $d<d_0$,

$$\mathcal{L}(d)-\mathcal{L}(d_0)=\{\mathcal{L}_1(d)-\mathcal{L}_1(d_0)\}+p(d-d_0)\lambda_n\Big\{\sum_{i=1}^n(\mathbf{y}_i-\bar{\mathbf{y}})^\top(\mathbf{y}_i-\bar{\mathbf{y}})\Big\}^{1/2}>0,\ \text{in probability}.$$

Analogously, by Conditions (C3)–(C4) and $\lambda_n/\ln n\to\infty$, when $d>d_0$,

$$\mathcal{L}(d)-\mathcal{L}(d_0)=O_p\{nh^{2s}+h^{-d}\ln n\}+p(d-d_0)\lambda_n\Big\{\sum_{i=1}^n(\mathbf{y}_i-\bar{\mathbf{y}})^\top(\mathbf{y}_i-\bar{\mathbf{y}})\Big\}^{1/2}>0,\ \text{in probability}.$$

Therefore, $\Pr(\widehat{d}=d_0)\to 1$ and the proof is completed. $\square$

References

[1] E. Bura, R.D. Cook, Extending sliced inverse regression: the weighted chi-squared test, J. Amer. Statist. Assoc. 96 (2001) 996–1003.
[2] G. Casella, R. Berger, Statistical Inference, Duxbury Press, Pacific Grove, 2002.
[3] Q. Cheng, L. Zhu, On relative efficiency of principal Hessian directions, Statist. Probab. Lett. 126 (2017) 108–113.
[4] R.D. Cook, Regression Graphics, Wiley, New York, 1998.
[5] R.D. Cook, B. Li, Dimension reduction for conditional mean in regression, Ann. Statist. 30 (2002) 455–474.
[6] J. Deng, Q. Wang, Dimension reduction estimation for probability density with data missing at random when covariables are present, J. Statist. Plann. Inference 181 (2017) 11–29.
[7] J. Fan, I. Gijbels, Local Polynomial Modelling and Its Applications, Chapman and Hall, London, 1996.
[8] G.L. Fan, L. Zhu, S. Ma, Nonlinear interaction detection through model-based sufficient dimension reduction, Statist. Sinica 29 (2019) 917–937.
[9] X. Guo, T. Wang, W. Xu, L. Zhu, Dimension reduction with missing response at random, Comput. Statist. Data Anal. 69 (2014) 228–242.
[10] X. Guo, W. Xu, Goodness-of-fit tests for general linear models with covariates missed at random, J. Statist. Plann. Inference 142 (2012) 2047–2058.
[11] Z. Hu, D.A. Follmann, J. Qin, Semiparametric dimension reduction estimation for mean response with missing data, Biometrika 97 (2010) 305–319.
[12] K.C. Li, Sliced inverse regression for dimension reduction, J. Amer. Statist. Assoc. 86 (1991) 316–327.
[13] K.C. Li, On principal hessian directions for data visualization and dimension reduction: another application of Stein's lemma, J. Amer. Statist. Assoc. 87 (1992) 1025–1039.
[14] S.G. Liao, Y. Lin, D.D. Kang, D. Chandra, J. Bon, N. Kaminski, F.C. Sciurba, G.C. Tseng, Missing value imputation in high-dimensional phenomic data: imputable or not, and how? BMC Bioinformatics 15 (2014) 346.
[15] Y. Ma, L. Zhu, A semiparametric approach to dimension reduction, J. Amer. Statist. Assoc. 107 (2012) 168–179.
[16] Y. Ma, L. Zhu, Efficiency loss and the linearity condition in dimension reduction, Biometrika 100 (2013) 371–383.
[17] Y. Ma, L. Zhu, On estimation efficiency of the central mean subspace, J. R. Stat. Soc. Ser. B Stat. Methodol. 76 (2014) 885–901.
[18] J. Shi, T.S. Lau, Empirical likelihood for partially linear models, J. Multivariate Anal. 72 (2000) 132–148.
[19] Z.H. Sun, Q.H. Wang, Checking the adequacy of a general linear model with responses missing at random, J. Statist. Plann. Inference 139 (2009) 3588–3604.
[20] L. Tang, Z. Zhou, Weighted local linear CQR for varying-coefficient models with missing covariates, Test 24 (2015) 583–604.
[21] Y. Wei, Y. Tang, P.D. McNicholas, Flexible high-dimensional unsupervised learning with missing data, 2017, arXiv:1706.06185.
[22] Y. Xia, H. Tong, W. Li, L.X. Zhu, An adaptive estimation of dimension reduction space, J. R. Stat. Soc. Ser. B Stat. Methodol. 64 (2002) 363–410.

[23] K. Xu, W. Guo, M. Xiong, L. Zhu, L. Jin, An estimating equation approach to dimension reduction for longitudinal data, Biometrika 103 (2016) 189–203.
[24] X. Yang, Q. Wang, Sufficient dimension reduction under dimension-reduction-based imputation with predictors missing at random, Statist. Sinica (2019) http://dx.doi.org/10.5705/ss.202017.0288.
[25] Y. Zhao, Q. Long, Multiple imputation in the presence of high-dimensional data, Stat. Methods Med. Res. 25 (2016) 2021–2035.
[26] L. Zhu, B. Miao, H. Peng, On sliced inverse regression with high-dimensional covariates, J. Amer. Statist. Assoc. 101 (2006) 630–643.
[27] L. Zhu, T. Wang, L. Zhu, Sufficient dimension reduction in regressions with missing predictors, Statist. Sinica 22 (2012) 1611–1637.
[28] L. Zhu, W. Zhong, Estimation and inference on central mean subspace for multivariate response data, Comput. Statist. Data Anal. 92 (2015) 68–83.