Journal of Statistical Theory and Applications
Vol. **18(1)**, March 2019, *pp.* 46–64
DOI: https://doi.org/10.2991/jsta.d.190306.006; ISSN 1538-7887
https://www.atlantis-press.com/journals/jsta

On Partially Linear Single-Index Models with Missing Response and Error-in-Variable Predictors

Tsung-Lin Cheng[1], Yin-Ying Lin[2], Xuewen Lu[3], and Radhey Singh[4*]

[1]*Department of Mathematics, National Changhua University of Education, Changhua city, Taiwan*
[2]*Institute of Statistics and Information, National Changhua University of Education, Changhua city, Taiwan*
[3]*Department of Mathematics and Statistics, University of Calgary, Calgary, AB, Canada*
[4]*Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada*

ARTICLE INFO

Article History

Received 20 January 2017
Accepted 15 August 2018

ABSTRACT

In this paper, we consider a partially linear single-index model when missing responses and nonlinear regressors with measurement error are taken into account. Utilizing data imputation for missing values and regression calibration for error-prone regressors, we not only estimate the parameters in the linear part as well as the single-index part, but also estimate the nonparametric link function by local linear fit. Under normalization, all the proposed estimators for the regression coefficients and the link function are proven to be asymptotically normal, and some illustrative simulations are provided to justify our methods.

1. INTRODUCTION

To avoid the so called "curse of dimensionality" in the nonparametric or semiparametric regression analysis, partially linear single-index models (PLSIM) emerged as an effective device for dimension reduction; see, for example, Härdle and Stoker [1], Powell *et al.* [2], Newey and Stoker [3], Ichimura [4], Carroll *et al.* [5], Xia and Härdle [6], Lu and Cheng [7], and many others. Served as an effective way of modelling a nonlinear relationship between several covariates and their response, PLSIM, however, might obtain biased estimations when the covariates and/or their response are not complete.

When one collects data (e.g., survival data), due to many practical problems, he may obtain an incomplete data set which, to such an extent, may lead to a biased estimation. Therefore, the augmentation of the missing data becomes more and more important in the data demanded world. In general, missing data might emerge in both of the responses and covariates, while in this paper we will mainly focus on the case when solely the response is missing. According to the nature of missing data, Little and Rubin [8] firstly classified the types of missingness into three categories—missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). In the present paper, we will consider the MAR mechanism (see, e.g., Wang *et al.* [9], Yun *et al.* [10]) which kicks in when the probability that a response is missing dose not depend on the unobserved measurements. A very important type of missingness is censoring; in particular, for the censoring case in PLSIM, Lu and Cheng [7] adopted a Kaplan–Meier-like transformation to overcome the biasedness of the estimation of the coefficients and link function. Besides, Cheng *et al.* [11] considers a more difficult problem concerning the estimation of the parameters and nonparametric function for a PLSIM with censored response and covariates having measurement error. However, for general missingness of the responses in PLSIM, there's not a paper studying on it.

Another important issue concerning incomplete data is about measurement error. Measurement error models have been largely studied in the literature, for example, Fuller [12], Carroll [13], Carroll and Stefanski [14], Carroll and Li [15], Lue [16], and Fan and Troung [17], among others. It was indicated by Carroll *et al.* [18] that, there are three effects caused by measurement errors: first, it causes bias in parameter estimation for statistical models; second, it leads to a loss of power, sometimes profound, for detecting interesting relationship among variables; finally, it makes the features of the data, making graphical model analysis difficult. Especially, the effects of biasedness of the parameters become severer especially when the relationship between the covariates and responses appear to be nonlinear.

In this paper, we consider the following PLSIM

$$Y = \beta_0^T V + \lambda_0\left(\alpha_0^T X\right) + \sigma\left(V, X\right)\varepsilon, \qquad \|\alpha_0\| = 1, \tag{1}$$

*Corresponding author. Email: rssingh@uwaterloo.ca

where Y is the response variable, $X = \left(X_1, \ldots, X_p\right)^T$ and $V = \left(V_1, \ldots, V_q\right)^T$ are predictors, α_0 and β_0 are parameters to be estimated, $\lambda_0(\cdot)$ is an unknown smooth function, and $\sigma(\cdot,\cdot)$ denotes the conditional variance representing the possible heteroscedasticity. Throughout this paper, $\|\cdot\|$ denotes the Euclidean norm. The restriction $\|\alpha_0\| = 1$ assures identifiability. Suppose that (V_i, X_i) and ε_i are independent, and ε_i are assumed to have mean $E(\varepsilon_i) = 0$ and variance $Var(\varepsilon_i) = 1$, for $i = 1, \ldots, n$. Suppose that we obtain a random sample of incomplete data

$$(Y_i,\ \delta_i,\ V_i,\ X_i), \quad i = 1, 2, \ldots, n$$

from model Eq. (1), where $\delta_i = 0$ if Y_i is missing, otherwise $\delta_i = 1$. The MAR assumption implies that δ and Y are conditionally independent given V and X, that is, $p\left(\delta = 1|Y, V, X\right) = p\left(\delta = 1|V, X\right)$.

Among the wide variety of procedures to handle missing data, data imputation is an important step. By imputing a plausible value for each missing datum, under mild conditions, the problem can be dealt with as if they were complete. Different categories of imputations can be found in Schulte Nordholt [19]. The first classification, roughly speaking, comprises the deterministic as well as the stochastic imputations [20]. The second classification is a distinction between naive and principled approaches. The naive imputations, mainly based on analyzing complete cases (listwise or pairwise), are a quick option. For example, the imputation of an unconditional mean is a naive approach. It might lead to a biased estimate even if the data are randomly missing. Little and Rubin ([8] Chapter 3) indicated that the obvious corrections of this biasedness will obtain the same estimates as found with available case procedures. The principled approaches adopt models for both the observed and missing data on which the imputations are based.

Besides, there is a distinction between imputations according to "explicit" and "implicit" models [21, 22]. Examples can be referred to the hot-deck procedures [23], in which missing values are imputed with donor cases from the set of completely observed cases. There are still many other imputation methods, for example, linear regression imputation [24], multiple imputation [20, 25], nonparametric kernel regression imputation[26, 27]), nearest neighbor imputation [28], ratio imputation [29], regression calibration [30], and semiparametric regression imputation [9], and so on. Wang and Sun [31] adopted semiparametric imputation, semiparametric regression surrogate and inverse marginal probability weighted (IMPW) approaches, separately, to estimate β and $g(\cdot)$ simultaneously in the partial linear model

$$Y = X^\tau \beta + g(T) + \varepsilon, \tag{2}$$

where Y is a scalar response MAR, X and T are complete covariates, β is a unknown regression parameter, $g(\cdot)$ is an unknown measurable function, and ε is the prediction error independent of X and T. As mentioned above, Wang and Sun adopted three imputation methods to the partial linear model. When we consider PLSIM, it is found that the third imputation approach, that is, the IMPW approach, doesn't work good according to our simulation study. Therefore, we drop the IMPW approach and adopt the other two approaches in the PLSIM setting. In this paper, we consider not only the missing responses, but also the regressors with measurement errors. Suppose that we can't observe the real covariate X but its contaminants W instead. In a general framework, the relationship between X and W can be described as below:

$$W = \gamma + \Gamma X + e, \tag{3}$$

where Γ is a $q \times p$ matrix, $p \le q$, which may be known, unknown, or partly known. An important case is when Γ equals the identity matrix I. No additional assumption is made except that δ, which has mean zero and constant covariance matrix $\sum_\delta$, is independent of (X, V, ε). When X is a scalar and $\alpha_0 = 1$, model Eq. (1) is a partially linear model. Partially linear model has many applications, in which Engle *et al.* [32] is the first to consider this kind of models. A more general case than model Eq. (1), was studied by Carroll *et al.* [5] in which model Eq. (1) is replaced by $g^{-1}\left\{E\left(Y|X, V\right)\right\}$, with a known link function g. Model Eq. (1) reduces to that of Carroll *et al.* [5] when the link function g becomes identity. Recently, partially linear single-index model with measurement error was studied by Liang and Wang [33]. They assumed the linear predictor V to be subjected to measurement errors, while in our setting not only the response is MAR but also the nonlinear regressor X has measurement errors. The paper is organized as follows: in Section 2, we depict the estimation procedures for model Eq. (1); Section 3 states the results on the asymptotic properties of our estimators; in Section 4, we present some illustrative simulations. All related proofs and theorems can be found in Appendix I. The estimation outcomes will be presented in Appendix II.

2. PROCEDURES OF ESTIMATIONS

2.1. Carroll and Li's Transformation

As mentioned in the introduction, how to calibrate the contaminated regressors to be unbiased is a very important issue. The Carroll and Li's [15] transformation, as stated in the following, is nothing more than a simple linear prediction of X by W,

$$U^* = LW = cov(X, W)\,\Sigma_W^{-1} W, \tag{4}$$

where Σ_W is the covariance matrix of W. Suppose that the individuals in a study are indexed by $i = 1, \ldots, n$, with the first m individuals being the validation sample, for which either the true X are observed in addition to the contaminated W or there are replicates of W. Conventionally, we refer the data consisting of i.i.d. sample (Y_i, W_i) $(i = m+1, \ldots, n)$ to be as the primary data. Typically, m is much less

than n. In general, L is unknown, and it can be estimated from a validation sample. Suppose that X and W are observed in Eq. (3) for a sample with validation data. L can be estimated by

$$\hat{L} = \hat{cov}(X, W)\,\hat{\Sigma}_{2W}^{-1},$$

where $\hat{cov}(X, W)$ is the sample covariance matrix between X and W and $\hat{\Sigma}_{2W}$ denotes the sample covariance matrix of W based on the validation sample $(X_i, W_i), i = 1, 2, \ldots, m$. Each row of $\hat{L}$ is the usual least squares regression slope of the corresponding coordinate of X against W with intercept included. Set $\hat{U}_i^* = \hat{L}W_i$ for $i = m+1, \ldots, n$ and define the associated sample covariance matrix $\hat{\Sigma}_{U^*} = \hat{L}\hat{\Sigma}_{1W}\hat{L}'$ based on the primary sample. Hereafter, the U_i^* can be replaced by $\hat{U}_i^*$ when L is unknown.

Suppose on the other hand that, we have a replicated data rather than a validation sample. As in Carroll and Li [15] and Lue [16], we consider an important special case when Γ is known and $p = q$. W.L.O.G. we take $\Gamma = I$. Let

$$W_{ij} = \gamma + X_i + e_{ij} \quad \left(j = 1, 2,\ i = 1, \ldots, m\right) \tag{5}$$

If Σ_e is the covariance matrix of e_{ij}, then $L = cov(X, W)\,\Sigma_W^{-1} = cov(X, \gamma + X + e)\,\Sigma_W^{-1} = \Sigma_X\Sigma_W^{-1} = (\Sigma_W - \Sigma_e)\,\Sigma_W^{-1} = I - \Sigma_e\Sigma_W^{-1}$. Let $\hat{\Sigma}_e$ and $\hat{\Sigma}_W - \frac{1}{2}\hat{\Sigma}_e$ be the sample covariance matrices of $(W_{i1} - W_{i2})/\sqrt{2}$ and $(W_{i1} + W_{i2})/2$, respectively; and let

$$\hat{L} = I - \hat{\Sigma}_e\hat{\Sigma}_W^{-1}.$$

With this choice of L, some similar results could be obtained.

2.2. Estimations of PLSIM with Missing Response and Error-Prone Predictors

Consider the PLSIM model defined by Eq. (1). In this section, we assume that we are given a data set with partially missing response and error-prone regressors in the nonlinear single-index term. In order to remedy the biasedness of estimations caused by missing and measurement error, we propose a modified quasi log-likelihood estimation procedure via an iterative minimization algorithm.

Let $\theta = (\alpha, \beta)$ be the vector of model parameters. If $\lambda(\cdot)$ were known and the data is free of measurement error and missing, the quasi log-likelihood estimator of $\theta_0 = (\alpha_0, \beta_0)$ and λ_0 is the one to minimize

$$\mathcal{L}_n(\theta, \lambda) = \sum_{i=m+1}^{n} \left[Y_i - \left\{\beta^T V_i + \lambda\left(\alpha^T X_i\right)\right\}\right]^2 \quad with\ \|\alpha\| = 1. \tag{6}$$

In the case when the data consists of MAR response variables and error-prone regressors, some auxiliary treatment of the data set is necessary. A difficulty common to single-index model is that, minimizing Eq. (6) involves the estimation of the nonparametric function λ. We partition Y into two parts $Y = (Y_{obs}, Y_{mis})$, with Y_{obs} indicating the observed values, and the $s \times 1$ vector Y_{mis}, indicating values that are missing. Assume that the observations are $\{(Y_i, \delta_i, V_i, W_i) : i = m+1, \ldots, n\}$, which is a random sample from the population $\{(Y, \delta, V, W)$ defined by Eqs. (1) and (2).

We denote the transformed U_i to be U_i^*, $i = m+1, \ldots, n$. Note that the transformed regressors $U_i^* = LW$, where $L = cov(X, W)\,\Sigma_W^{-1}$ if L is known and $\hat{L} = \hat{cov}(X, W)\,\hat{\Sigma}_W^{-1}$ otherwise. First, assume that Y is not missing and can be observed completely. For fixed $u \in \mathfrak{R}$ and v in a near neighborhood of u, one may approximate the unknown smooth function $\lambda(v)$ by

$$\lambda(v) \approx \lambda(u) + \lambda'(u)(v - u) \equiv a_0 + a_1(v - u), \tag{7}$$

which is called a "local linear fit." Thus, finding $\lambda(u)$ is tantamount to finding the intercept a_0 of the approximating regression line. Around u, model Eq. (1) is approximately becoming

$$Y = \beta^T V + \lambda(u) + \lambda'(u)\left(\alpha^T X - u\right) + \sigma(V, X)\,\varepsilon. \tag{8}$$

In order to claim that, when we replace X in model Eq. (1) by LW, the estimated α is unchanged, we reduce our problem to the following simple linear case. Consider the following model

$$Y^* = a_0 + a_1\left(\alpha^T X - u\right) + \varepsilon^*, \tag{9}$$

where $Y^* = Y - \beta^T V$, X is the same as that in Eq. (1) and is uncorrelated with $\varepsilon^* = \sigma(V, X)\,\varepsilon$, and a_0 and a_1 are constants. Let W satisfy Eq. (3), we may consider the following model

$$Y^* = a_0 + a_1\left(\alpha^T U^* - u\right) + \varepsilon^*, \tag{10}$$

where $U^* = LW$. It is clear that model Eqs. (9) and (10) have the same estimate of α when L is known. Moreover, even if L is unknown, utilizing the validation data to obtain $\hat{L} = \hat{cov}(X, W)\,\hat{\Sigma}_w^{-1}$, model Eqs. (9) and (10) still have approximately the same estimate of α.

Now we return to the case as Y is partially missing. Let $Z=(V,X)$, $\sigma^2(Z)=E\left(\varepsilon^2|Z\right)$ and $\Delta(z)=P\left(\delta=1|Z=z\right)$. Motivated by Wang and Sun [31], we have $Y_i^{[Im]}=\delta_i Y_i+(1-\delta_i)\left(\beta_0^T V_i+\lambda_0\left(\alpha_0^T X_i\right)\right)$, that is, $Y_i^{[Im]}=Y_i$ if $\delta_i=1$, otherwise, $Y_i^{[Im]}=\beta_0^T V_i+\lambda_0\left(\alpha_0^T X_i\right)$. By MAR assumption, we have $E\left[Y^{[Im]}|Z\right]=E\left[\delta Y+(1-\delta)\left(\beta_0^T V+\lambda_0\left(\alpha_0^T X\right)\right)|Z\right]=\beta_0^T V+\lambda_0\left(\alpha_0^T X\right)=E[Y|Z]$. But $Y_i^{[Im]}$ contains unknown α_0, β_0 and $\lambda_0(\cdot)$. Naturally, we might replace $Y_i^{[Im]}$ by

$$Y_i^{(I)}=\delta_i Y_i+(1-\delta_i)\left(\hat{\beta}_0^{(obs)T} V_i+\hat{\lambda}_0^{(obs)}\left(\hat{\alpha}_0^{(obs)T} X_i\right)\right) \tag{11}$$

where $\hat{\alpha}_0^{(obs)}$, $\hat{\beta}_0^{(obs)}$ and $\hat{\lambda}_0^{(obs)}$ are obtained by our estimation algorithm below by replacing Y^{**} with Y_{obs}. Similarly, we may define

$$Y_i^{(R)}=\hat{\beta}_0^{(obs)T} V_i+\hat{\lambda}_0^{(obs)}\left(\hat{\alpha}_0^{(obs)T} X_i\right) \tag{12}$$

to be as the semiparametric regression surrogate. Then we substitute these synthetic data, $Y^{(I)}$ and $Y^{(R)}$, into Step 1 to estimate both parametric component θ_0 and nonparametric function λ_0 by using the local linear fit and denote the corresponding estimator by $\hat{\theta}_0^{(I)}=\left(\hat{\alpha}_0^{(I)},\hat{\beta}_0^{(I)}\right)$, $\hat{\lambda}_0^{(I)}$, $\hat{\theta}_0^{(R)}=\left(\hat{\alpha}_0^{(R)},\hat{\beta}_0^{(R)}\right)$, $\hat{\lambda}_0^{(R)}$, respectively. With the local model Eq. (8), we may estimate $\lambda(\tilde{u})$ by minimizing the following modified local quasi-likelihood

$$\sum_{i=m+1}^{n}\left[Y^{**}-\left\{\beta^T V_i+a_0+a_1\left(\alpha^T U_i^*-\tilde{u}\right)\right\}\right]^2 K_h\left(\alpha^T U_i^*-\tilde{u}\right)$$

with respect to a_0 and a_1, where $K_h(\cdot)=h^{-1}K(\cdot/h)$, h is a suitable bandwidth, $\tilde{u}$ is a fixed real number, and Y^{**} may be $Y^{(I)}$ or $Y^{(R)}$ according to which augmentation is used. Fan and Gijbels [34] proposed a nonparametric estimator of $\lambda(\tilde{u})$, which is defined by

$$\hat{\lambda}(\tilde{u})=\sum_{i=m+1}^{n} w_i(\tilde{u})\, Y^{**}/\sum_{i=m+1}^{n} w_i(\tilde{u}),$$

with

$$w_i(\tilde{u})=K\left(\frac{\tilde{u}-\alpha^T U_i^*}{h}\right)\left[s_{n,2}-\left(\tilde{u}-\alpha^T U_i^*\right)s_{n,1}\right].$$

where

$$s_{n,l}=\sum_{i=m+1}^{n} K\left(\frac{\tilde{u}-\alpha^T U_i^*}{h}\right)\left(\tilde{u}-\alpha^T U^i\right)^l,\quad l=1,2.$$

Our estimation algorithm consists of the following steps:

- Step 1: Treat the synthetic data Y^{**} and U^* as complete data and obtain initial guess of $\theta_0=(\alpha_0,\beta_0)$ by Xia and Härdle's [6] algorithm. Let $\hat{\theta}=\left(\hat{\alpha},\hat{\beta}\right)$ be the initial guess of θ_0. Set $\|\hat{\alpha}\|=1$.
- Step 2: Find $\hat{\lambda}\left(\tilde{u};h,\hat{\theta}\right)=\hat{a}(\tilde{u})$ as a function of $\tilde{u}$ by minimizing

$$\sum_{i=m+1}^{n}\left[Y^{**}-\left\{\beta^T V_i+a_0+a_1\left(\alpha^T U_i^*-\tilde{u}\right)\right\}\right]^2 K_h\left(\alpha^T U_i^*-\tilde{u}\right). \tag{13}$$

- Step 3: Update $\hat{\theta}$ by minimizing

$$\sum_{i=m+1}^{n}\left[Y^{**}-\left\{\beta^T V_i+\hat{\lambda}\left(\alpha^T U_i^*;h,\hat{\theta}\right)\right\}\right]^2$$

  with respect to $\theta=(\alpha,\beta)$.
- Step 4: Iterate Steps 2 and 3 until convergence is achieved.

3. ASYMPTOTIC THEOREMS FOR THE ESTIMATORS

In this section, we will establish the asymptotic normality of the estimators of the parameters emerging in the PLSIM model. Condition A is given to ensure the asymptotic properties of the estimators to hold.

Condition A.

i. The kernel K is a symmetric function on [-1,1], and satisfies uniform Lipschitz condition of order 1 on R.

ii. The random vectors V and $U^* = L(X+e)$ are bounded.

iii. The marginal density $f(\tilde{u})$ of $\tilde{U} = \alpha_0^T U^*$ is positive, and has a continuous second derivative on its compact support $D \subset R$.

iv. The random vector $U^* = L(X+e)$ has a compact support $\aleph \subset R^p$, D_{λ_0} is an open interval containing $\cup\{\alpha^T u^* : \|\alpha\| = 1, u^* \in \aleph\}$. The second derivative of $\lambda_0(\tilde{u})$ exists, is continuous and bounded on D_{λ_0}.

v. The functions $E\{U^*|\tilde{U} = \tilde{u}\}$ and $E\{V|\tilde{U} = \tilde{u}\}$ are twice differentiable in $\tilde{u} \in D$, and their second derivatives satisfy Lipschitz condition of order 1. On the boundaries, the continuity and differentiability mean left or right continuity and differentiability.

vi. For a given $\hat{\lambda}$, assume that $\hat{\alpha} - \alpha_0$ and $\hat{\beta} - \beta_0 = O_p\left(n^{-1/2}\right)$, that is, the initial estimates are in a $\sqrt{n}$-neighborhood of the true parameter values in probability, respectively.

vii. Let

$$\Psi = \begin{pmatrix} U^* \lambda_0'\left(\tilde{U}\right) \\ V \end{pmatrix}, \; H = \Psi - E\left(\Psi|\tilde{U}\right), \; \epsilon^{**} = Y^{**} - \left\{\beta_0^T V + \lambda_0\left(\alpha_0^T U^*\right)\right\},$$

both $Q = E\left\{H^{\otimes^2}\right\}$ and $\Omega = E\left\{(H\epsilon^{**})^{\otimes^2}\right\}$ are positively definite, Y^{**} may be $Y^{(I)}$ or $Y^{(R)}$ and ϵ^{**} may be $\epsilon^{(I)}$ or $\epsilon^{(R)}$ according to which augmentation is used.

Theorem 1. Under Condition A and the following conditions on the bandwidth: $nh^4 \to 0$ and $nh^3 = O\left(logn\right)$, as $n \to \infty$, hold. Then, the estimator $\hat{\theta}_0^{(I)} = \left(\hat{\alpha}_0^{(I)}, \hat{\beta}_0^{(I)}\right)$ from the iterative algorithm satisfies

$$n^{1/2}\begin{pmatrix} \hat{\alpha}_0^{(I)} - \alpha_0 \\ \hat{\beta}_0^{(I)} - \beta_0 \end{pmatrix} \xrightarrow{D} N\left(0, Q^{-1}\Omega Q^{-1}\right),$$

where Q and Ω are defined in Condition$A$ (vii)," $\xrightarrow{D}$ " denotes convergence in distribution.

Theorem 2. Under the same conditions as given in Theorem 1, the estimator $\hat{\theta}_0^{(R)} = \left(\hat{\alpha}_0^{(R)}, \hat{\beta}_0^{(R)}\right)$ from the iterative algorithm satisfies

$$n^{1/2}\begin{pmatrix} \hat{\alpha}_0^{(R)} - \alpha_0 \\ \hat{\beta}_0^{(R)} - \beta_0 \end{pmatrix} \xrightarrow{D} N\left(0, Q^{-1}\Omega Q^{-1}\right).$$

It is interesting to note that $\left(\hat{\alpha}_0^{(I)}, \hat{\beta}_0^{(I)}\right)$ have the same asymptotic variance as $\left(\hat{\alpha}_0^{(R)}, \hat{\beta}_0^{(R)}\right)$, which has been shown by Wang and Sun [31]. All these related theorems are referred to Appendix I.

By the root-n consistency of $\left(\hat{\alpha}, \hat{\beta}\right)$ and the assumptions for the bandwidth h and the kernel function $K(\cdot)$, we may prove that $\hat{\lambda}\left(\tilde{u}; \hat{\alpha}, \hat{\beta}\right) - \hat{\lambda}\left(\tilde{u}; \alpha_0, \beta_0\right) = O_p\left(n^{-1/2}\right)$. When α_0 and β_0 are known, we can easily prove the asymptotic normality of $\hat{\lambda}\left(\tilde{u}; \alpha_0, \beta_0\right)$ using the results in Fan and Gijbels [34]. Therefore, the asymptotic normality for the local linear estimator $\hat{\lambda}\left(\tilde{u}; \hat{\alpha}, \hat{\beta}\right)$ with estimated parameters $\hat{\alpha}$ and $\hat{\beta}$ can be stated as follows:

Theorem 3. Let $f(\cdot)$ be the density function of $\tilde{U} = \alpha_0^T U*$. *If* $h = O\left(n^{-1/5}\right)$ and $K(\cdot)$ has third-order continuous derivatives and its third-order derivative is bounded on D, then under Condition A, on the covariates $\{\tilde{U}_1, \ldots, \tilde{U}_n\}$, for any interior point $\tilde{u} \in D$,

$$\sqrt{nh}\left(\hat{\lambda}\left(\tilde{u}; \hat{\alpha}, \hat{\beta}\right) - \lambda_0\left(\tilde{u}\right) - \lambda_0''\left(\tilde{u}\right) c_K h^2/2\right) \xrightarrow{D} N\left(0, d_K\right) \sigma^{*2}\left(\tilde{u}\right),$$

where $\sigma^*\left(\tilde{u}\right) = Var\left\{\left(Y^{**} - \beta_0^T V\right)|\tilde{U} = \tilde{u}\right\}$, $c_K = \int_{-\infty}^{+\infty} v^2 K(v)\, dv$ and $d_K = \int_{-\infty}^{+\infty} K^2(v)\, dv$.

4. SIMULATION

Example

In this example, we conduct some Monte Carlo simulations to estimate the regression coefficients for an partially single-index model with incomplete data, and $q = dim(V) = 2$ and $p = dim(X) = 2$. Let $X = (X_1, X_2)^T$, $X_1 \sim Uniform(-2, 2)$, $X_1' \sim Triangular(-2, 2)$, $X_2 = \left(\frac{1}{3}\right) X_1 + \left(\frac{2}{3}\right) X_1'$ and $V = (V_1, V_2)^T$, where V_1 and $V_2 \sim Bernoulli\left(p = 0.5\right)$ are independent. Assume in addition that the covariates V and X are independent. One would notice that X_1 and X_2 are dependent. Let the data be generated from the following model:

$$Y = \beta_0^T V + \lambda_0\left(\alpha_0^T X\right) + \varepsilon, \tag{14}$$

where $\varepsilon \sim N\left(0, \sigma_0^2 = 0.5^2\right)$, the true parameters are $\beta_0 = (-1, 2)^T$ and $\alpha_0 = \left(\sqrt{2}/2, \sqrt{2}/2\right)^T$ and the true unknown function is $\lambda_0(\tilde{u}) = (-1/2)\left(\tilde{u} - \sqrt{2}/2\right)^2 + 6$, $\tilde{u} = \alpha_0^T u^*$. First, we consider the case as Y is MAR. We generate, respectively, 300 replicates of random sample of size $n = 60, 120$, and 240 for the following three mechanisms:

Case 1: $\Delta_1(z) = P\left(\delta = 1 | V = (v_1, v_2), X = (x_1, x_2)\right) = 0.8 + 0.2\left(|v_1| + |v_2| + |x_1| + |x_2|\right)$ if $|v_1| + |v_2| + |x_1| + |x_2| \le 1$, and $= 0.90$ elsewhere.

Case 2: $\Delta_2(z) = P\left(\delta = 1 | V = (v_1, v_2), X = (x_1, x_2)\right) = 0.9 - 0.2\left(|v_1| + |v_2| + |x_1| + |x_2|\right)$ if $|v_1| + |v_2| + |x_1| + |x_2| \le 1.5$, and $= 0.80$ elsewhere.

Case 3: $\Delta_3(z) = P\left(\delta = 1 | V = (v_1, v_2), X = (x_1, x_2)\right) = 0.8 - 0.2\left(|v_1| + |v_2| + |x_1| + |x_2|\right)$ if $|v_1| + |v_2| + |x_1| + |x_2| \le 1$, and $= 0.50$ elsewhere.

By conducting Monte Carlo simulations, the mean response rates of the above three cases are $E\Delta_1(z) \approx 0.90$, $E\Delta_2(z) \approx 0.78$, and $E\Delta_3(z) \approx 0.51$, respectively. Accordingly, our missing proportions are about 10%, 22%, and 49%, respectively. Second, we focus on the case when the response Y is MAR and the covariate X of nonparametric part has a validation data concerning its contamination W and itself. We assume that the primary sample size is n' and the sample size of the validation data is m, $\gamma = 0$, $\Gamma = I$, and the distribution of e are normal with mean 0, variance $\sqrt{3}/4$.

In Table A.1 (resp. Table A.4), we report the results of $\left(\hat{\alpha}_0^{(I)}, \hat{\beta}_0^{(I)}\right)\left(\text{resp. } \left(\hat{\alpha}_0^{(R)}, \hat{\beta}_0^{(R)}\right)\right)$ when Y is MAR and X is without measurement error. In Table A.2 (resp. Table A.5), we consider the case as X has measurement error with σ_e^2 of e taken to be $\sqrt{3}/4$. After calibrating W into U^*, we report the results of $\left(\hat{\alpha}_0^{(I)}, \hat{\beta}_0^{(I)}\right)\left(\text{resp. } \left(\hat{\alpha}_0^{(R)}, \hat{\beta}_0^{(R)}\right)\right)$. While in Table A.3 (resp. Table A.6), the error-prone W is not calibrated and the other assumptions about missing are preserved. We conduct 300 simulations totally for each table. In these tables the sample mean (MEAN), standard derivation (SD), root-mean-square error (RMSE), and the median (MED) are represented as a function of the sample size n, primary size n', validation size m, and the missing proportion p. We use the well-known Epanechnikov kernel function $K(v) = \left(\frac{3}{4}\right)\left(1 - v^2\right) I\left[|v| \le 1\right]$ to do the kernel smoothing. Figures A.1–A.6 illustrate the true nonparametric curve and the fitted curve (dotted curve).

From Tables A.1 and A.4, all the proposed estimates of (α_0, β_0) have similar SD and RMSE. $\hat{\alpha}_0^{(I)}$ and $\hat{\alpha}_0^{(R)}$ perform similarly and $\hat{\beta}_0^{(I)}$ performs slightly better than $\hat{\beta}_0^{(R)}$. From Tables A.2, A.3, and Tables A.5, A.6, those estimates of (α_0, β_0) with calibrated outperform those with W uncalibrated. From Figures A.1 and , A.4 $\hat{\lambda}_0^{(I)}$ and $\hat{\lambda}_0^{(R)}$ perform similarly. From Figures A.2, A.3 and Figures A.5, A.6, both approaches work to relieve the effect upon missingness and measurement error.

REFERENCES

1. W. Härdle, T.M. Stoker, J. Am. Statist. Assoc. 84 (1989), 986–995.
2. J.L. Powell, J.H. Stock, T.M. Stoker, Econometrica. 57 (1989), 1403–1430.
3. W.K. Newey, T.M. Stoker, Econometrica. 61 (1993), 1199–1223.
4. H. Ichimura, J. Econometrics. 58 (1993), 71–129.
5. R.J. Carroll, J. Fan, I. Gijbels, M.P. Wand, J. Am. Statist. Assoc. 92 (1997), 477–489.
6. Y. Xia, W. Härdle, J. Multivariate Anal. 97 (2006), 1162–1184.
7. X.W. Lu, T.L. Cheng, J. Multivariate Anal. 98 (2007), 1895–1922.
8. R.J.A Little, D.B. Rubin, Statistical Analysis with Missing Data, John Wiley, New York, 2002.
9. Q.H. Wang, O. Lindon, W. Härdle, J. Am. Statist. Assoc. 99 (2004), 334–345.
10. S.C. Yun, Y. Lee, M.G. Kenward, Biometrika. 94(4) (2007), 905–919.
11. T.L. Cheng, H.H. Lue, X. Lu, J.W. Sun, J. Chinese Statist. Assoc. 48(2) (2010), 108–126.
12. W.A. Fuller, Measurement Error Models, Wiley, New York, 1987.
13. R.J. Carroll, Statist. Med. 8 (1989), 1075–1093.
14. R.J. Carroll, L.A. Stefanski, J. Am. Statist. Assoc. 85 (1990), 652–663.

15. R.J. Carroll, K.C. Li, J. Am. Statist. Assoc. 87 (1992), 1040–1050.
16. H.H. Lue, Biometrika. 91 (2004), 409–423.
17. J. Fan, Y.K. Troung, Ann. Statist. 21 (1993), 1900–1925.
18. D. Ruppert, M.P. Wand, R. Carroll, Semiparametric regression, Cambridge University Press, 2006.
19. E. Schulte Nordholt, Int. Stat. Rev. 66 (1998), 157–180.
20. D.B. Rubin, Multiple Imputation for Nonresponse in Surveys, John Wiley, New York, 1987.
21. R.J.A. Little, N. Schenker, in: G. Arminger, C.C. Clogg, M.E. Sobel, (Eds.), Handbook of Statistical Modeling for the Social and Behavioral Sciences, Plenum Press, New York, 1995, pp. 39–75.
22. J.L. Schafer, Analysis of Incomplete Multivariate Data, Chapman & Hall, London, 1997.
23. R.J.A. Little, D.B. Rubin, Statistical Analysis with Missing Data, second ed., John Wiley, New York, 2002.
24. M. Healy, M. Westmacott, Appl. Statist. 5 (1956), 203–206.
25. J.R. Carpenter, M.G. Kenward, I.R. White, Stat. Meth. Med. Res. 16(3) (2007), 259–275.
26. P.E. Cheng, J. Am. Statist. Assoc. 89 (1994), 81–87.
27. Q.H. Wang, J.N.K. Rao, Ann. Statist. 30 (2002), 345–358.
28. T. Nittner, Statist. Method & Appl., 12 (2003), 195–210.
29. J.N.K. Rao, J. Am. Statist. Assoc. 91 (1996), 499–520.
30. C.Y. Wang, L. Hsu, Z.D. Feng, R.L. Prentice, Biometrics. 53(1) (1997), 131–145.
31. Q.H. Wang, Z.H. Sun, J. Multivariate Anal. 98 (2007), 1470–1493.
32. R.F. Engle, C.W.J. Granger, J. Rice, A. Weiss, J. Am. Statist. Assoc. 81 (1986), 310–320.
33. H. Liang, N. Wang, Statistica Sinica. 15 (2005), 99–116.
34. J. Fan, I. Gijbels, J. Am. Statist. Assoc. 89 (1994), 560–570.

APPENDIX I

Proof of Theorem 1. and 2. The proof of Theorem 2 is just a part of arguments used in the proof of Theorem 1, therefore we omit it. Here, we give a detailed proof of Theorem 1 only.

Denote

$$\Psi = \begin{pmatrix} U^* \lambda_0' \left(\tilde{U}\right) \\ V \end{pmatrix}, \quad \Lambda = \begin{pmatrix} U^* \lambda_0' \left(\tilde{U}\right) & 0 \\ 0 & V \end{pmatrix},$$

and

$$\Omega = E\left[\left\{\left(\Psi - E\left\{\Psi|\tilde{U}\right\}\right)\epsilon^{(I)}\right\}\left\{\left(\Psi - E\left\{\Psi|\tilde{U}\right\}\right)\epsilon^{(I)}\right\}^T\right],$$

where $\tilde{U} = \alpha_0^T U^*$ and $\epsilon^{(I)} = Y^{(I)} - \left\{\lambda_0\left(\tilde{U}\right) + \beta_0^T V\right\}$. Let $Q = B_{\alpha_0,\beta_0} - A_{\alpha_0,\beta_0}$, with

$$A_{\alpha_0,\beta_0} = -E\left[\Psi\Psi^T\right], \quad B_{\alpha_0,\beta_0} = -E\left[E\left(\Psi|\tilde{U}\right)E\left(\Psi^T|\tilde{U}\right)\right].$$

The proof consists of two steps. The first step is to obtain an expansion for $\hat{\lambda}$. For simplicity, let $a_0 = a_0(\tilde{u}) = \lambda_0(\tilde{u})$, $a_1 = a_1(\tilde{u}) = h\lambda_0'(\tilde{u})$, $\epsilon_i^{(I)*} = Y_i^{(I)} - \left\{a_0 + a_1\left(\tilde{U}_i - \tilde{u}\right)/h + \beta_0^T V_i\right\}$. Without loss of generality, suppose that $D = [c, d]$ for $-\infty < c < d < \infty$, and define $D^0 = [c + h, d - h]$ and $D^1 = D \setminus D^0$, where h is the bandwidth. Let

$$\begin{aligned} L_n(\tilde{u}) = {} & n^{-1}\sum_{i=1}^n K_h(\tilde{U}_i - \tilde{u})\frac{\epsilon_i^{(I)*}}{f(\tilde{u})} - (\hat{\alpha}^T - \alpha_0^T)E\left\{U^*\lambda_0'(\tilde{U})|\tilde{U} = \tilde{u}\right\} \\ & - \left(\hat{\beta}^T - \beta_0^T\right)E\left\{V|\tilde{U} = \tilde{u}\right\}. \end{aligned} \tag{A.1}$$

We will show that

$$\begin{aligned} & \sup_{\tilde{u}\in|D^0}\left|\hat{\lambda}\left(\tilde{u};\hat{\alpha},\hat{\beta}\right) - \lambda_0(\tilde{u}) - L_n(\tilde{u})\right| = o_p\left(n^{-1/2}\right) + O_p\left(h^2\right), \\ & \sup_{\tilde{u}\in|D^0}\left|\hat{\lambda}\left(\tilde{u};\hat{\alpha},\hat{\beta}\right) - \lambda_0(\tilde{u}) - L_n(\tilde{u})\right| = o_p\left(n^{-1/2}\right) + O_p\left(h^2\right) + O_p(h). \end{aligned} \tag{A.2}$$

Denote the $k \times k$ identity matrix by I_k and P_{α_0} by

$$P_{\alpha_0} = \begin{bmatrix} I_p - \alpha_0\alpha_0^T & 0 \\ 0 & I_q \end{bmatrix}.$$

Then, we will obtain the following representation:

$$\begin{aligned} P_{\alpha_0} Q n^{1/2}\begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} &= n^{-1/2}\sum_{i=1}^n P_{\alpha_0}\left[\Psi_i - E\left\{\Psi_i|\tilde{U}_i\right\}\right]\epsilon_i^{(I)} + o_p(1) \\ &= S_n + o_p(1), \end{aligned} \tag{A.3}$$

where $\epsilon_i^{(I)} = Y_i^{(I)} - \left\{\lambda_0\left(\tilde{U}_i\right) + \beta_0^T V_i\right\}$. The second step is to show that the first term on the right-hand side of Eq. (A.3) has an asymptotic variance–covariance matrix $P_{\alpha_0}\Omega P_{\alpha_0}$. Therefore,

$$n^{1/2}\begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} = \left(P_{\alpha_0}Q\right)^-(S_n) + \left(P_{\alpha_0}Q\right)^- o_p(1),$$

where A^- denotes the generalized inverse of a square matrix A, $\left(P_{\alpha_0}Q\right)^-(S_n)$ has an asymptotic variance–covariance $\left(P_{\alpha_0}Q\right)^- P_{\alpha_0}(\Omega)P_{\alpha_0}\left\{\left(P_{\alpha_0}Q\right)^T\right\}^- = Q^-(\Omega)Q^- = Q^{-1}(\Omega)Q^{-1}$, $\left(P_{\alpha_0}Q\right)^- o_p(1) = o_p(1)$ since the elements of $\left(P_{\alpha_0}Q\right)^-$ are finite. To the end, Theorem 1 is proved by applying the central limit theorem. Now, we start to derive the desired results in each step.

Proof of (A.2). Let $a_0 = \lambda_0(\tilde{u})$, $a_1 = h\lambda_0'(\tilde{u})$. The local linear estimates of a_0 and a_1 are obtained from solving

$$0 = n^{-1}\sum_{i=1}^n K_h\left(\hat{\tilde{U}}_i - \tilde{u}\right)\begin{bmatrix} 1 \\ \left(\hat{\tilde{U}}_i - \tilde{u}\right)/h \end{bmatrix}\hat{\epsilon}_i^{(I)*},$$

where $\hat{\epsilon}_i^{(I)*} = Y_i^{(I)} - \left\{\hat{a}_0 + \hat{a}_1\left(\widehat{\tilde{U}}_i - \tilde{u}\right)/h + \hat{\beta}^T V_i\right\}$, $\hat{\cdot}$ indicates the estimated error and $\hat{\cdot}^*$ indicates a local version of the estimated error. By this convention, we define $\epsilon_i^{(I)*} = Y_i^{(I)} - \left\{\hat{\lambda}\left(\hat{\alpha}^T U_i^*; \hat{\alpha}, \hat{\beta}\right) + \hat{\beta}^T V_i\right\}$. Using a Taylor expansion approximately and eliminating higher order term, we get uniformly for $\tilde{u} \in D$,

$$0 = n^{-1}\sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right)\begin{bmatrix} 1 \\ \left(\tilde{U}_i - \tilde{u}\right)/h \end{bmatrix}\left\{\hat{\epsilon}_i^{(I)*} - (\hat{a}_0 - a_0) - \left(\left(\tilde{U}_i - \tilde{u}\right)/h\right)\times (\hat{a}_1 - a_1) - (a_1/h)\left(\hat{\alpha}^T - \alpha_0^T\right)U_i^* - \left(\hat{\beta}^T - \beta_0^T\right)V_i + o_p\left(n^{-1/2}\right) + O_p\left(h^2\right).$$

Solving the above equation for $\hat{a}_0 - a_0$, we have uniformly for $\tilde{u} \in D$,

$$\hat{a}_0 - a_0 = \left[1/\left\{n^{-1}\sum_{i=1}^{n} K_h(\tilde{U}_i - \tilde{u})\right\}\right]\left[n^{-1}\sum_{i=1}^{n} K_h(\tilde{U}_i - \tilde{u})\left\{\epsilon_i^{(I)*} - (a_1/h)\times \left(\hat{\alpha}^T - \alpha_0^T\right)U^i - \left(\hat{\beta}^T - \beta_0^T\right)V_i\right\} + o_p\left(n^{-1/2}\right) + O_p\left(h^2\right)\right]. \tag{A.4}$$

Let $\hat{f}(\tilde{u}) = n^{-1}\sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right)$ be the kernel estimator of $f(\tilde{u})$, we have the following results about the kernel density estimators (**proofs are put in Sections I.1 and I.2**):

$$\sup_{\tilde{u}\in D}\left|n^{-1}\sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right)(a_1/h)\,U_i^*/\hat{f}(\tilde{u}) - E\left\{U^*|\tilde{U} = \tilde{u}\right\}\lambda_0'(\tilde{u})\right| = O_p(h),$$

$$\sup_{\tilde{u}\in D}\left|n^{-1}\sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right)V_i/\hat{f}(\tilde{u}) - E\left\{V|\tilde{U} = \tilde{u}\right\}\right| = O_p(h), \tag{A.5}$$

$$\sup_{\tilde{u}\in D}\left|n^{-1}\sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right)\epsilon_i^{(I)*}/\hat{f}(\tilde{u}) - 0\right| = O_p(h), \tag{A.6}$$

and

$$\sup_{\tilde{u}\in D^0}\left|\hat{f}(\tilde{u}) - f(\tilde{u})\right| = O_p(h), \quad \sup_{\tilde{u}\in D^1}\left|\hat{f}(\tilde{u}) - f(\tilde{u})\right| = O_p(1). \tag{A.7}$$

Since

$$\frac{n^{-1}\sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right)\epsilon_i^{(I)*}}{n^{-1}\sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right)} - \frac{n^{-1}\sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right)\epsilon_i^{(I)*}}{f(\tilde{u})}$$

$$= \frac{n^{-1}\sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right)\epsilon_i^{(I)*}}{\hat{f}(\tilde{u})} \times \frac{f(\tilde{u}) - \hat{f}(\tilde{u})}{f(\tilde{u})},$$

by Eqs. (A.5) and (A.7), we obtain

$$\sup_{\tilde{u}\in D^0}\left|\frac{n^{-1}\sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right)\epsilon_i^{(I)*}}{n^{-1}\sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right)} - \frac{n^{-1}\sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right)\epsilon_i^{(I)*}}{f(\tilde{u})}\right| = O_p\left(h^2\right)$$

and

$$\sup_{\tilde{u}\in D^1}\left|\frac{n^{-1}\sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right)\epsilon_i^{(I)*}}{n^{-1}\sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right)} - \frac{n^{-1}\sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right)\epsilon_i^{(I)*}}{f(\tilde{u})}\right| = O_p(h).$$

Substituting the kernel terms in the linearized Eq. (A.4) by their asymptotic counterparts, we obtain Eq. (A.2).

Proof of (A.3). By a Taylor expansion, we have

$$\begin{aligned}
&\hat{\lambda}\left(\hat{\alpha}^T U_i^*;\hat{\alpha},\hat{\beta}\right)-\lambda_0\left(\alpha_0^T U_i^*\right)\\
&\quad=\hat{\lambda}\left(\hat{\alpha}^T U_i^*;\hat{\alpha},\hat{\beta}\right)-\hat{\lambda}\left(\alpha_0^T U_i^*;\hat{\alpha},\hat{\beta}\right)+\hat{\lambda}\left(\alpha_0^T U_i^*;\hat{\alpha},\hat{\beta}\right)-\lambda_0\left(\alpha_0^T U_i^*\right)\\
&\quad=\hat{\lambda}'\left(\hat{\alpha}_0^T U_i^*;\hat{\alpha},\hat{\beta}\right)\left(\hat{\alpha}^T-\alpha_o^T\right)U_i^*+\hat{\lambda}\left(\alpha_0^T U_i^*;\hat{\alpha},\hat{\beta}\right)-\lambda_0\left(\alpha_0^T U_i^*\right)+o_p\left(n^{-1/2}\right)\\
&\quad=\lambda_0'\left(\alpha_0^T U_i^*\right)\left(\hat{\alpha}^T-\alpha_o^T\right)U_i^*+\hat{\lambda}\left(\alpha_0^T U_i^*;\hat{\alpha},\hat{\beta}\right)-\lambda_0\left(\alpha_0^T U_i^*\right)+o_p\left(n^{-1/2}\right).
\end{aligned}\tag{A.8}$$

With ξ being the Lagrange multiplier, we know that $\left(\hat{\alpha},\hat{\beta}\right)$ is the solution to

$$0=\xi\begin{pmatrix}\hat{\alpha}\\0\end{pmatrix}+n^{-1/2}\sum_{i=1}^{n}\hat{\Lambda}_i D_i,$$

where

$$\hat{\Lambda}_i=\begin{pmatrix}U_i^*\hat{\lambda}'\left(\hat{\alpha}^T U_i^*;\hat{\alpha},\hat{\beta}\right) & 0\\ 0 & V_i\end{pmatrix},\quad D_i=\begin{pmatrix}\hat{\epsilon}_i^{(I)}\\ \hat{\epsilon}_i^{(I)},\end{pmatrix},$$

$$\hat{\epsilon}_i^{(I)}=Y_i^{(I)}-\left\{\hat{\lambda}\left(\hat{\alpha}^T U_i^*;\hat{\alpha},\hat{\beta}\right)+\hat{\beta}^T V_i\right\}.$$

Let

$$D_{0i}=\begin{pmatrix}\epsilon_i^{(I)}\\ \epsilon_i^{(I)},\end{pmatrix}.$$

By Taylor expansion, we obtain

$$\begin{aligned}
D_i&=D_{0i}+\begin{pmatrix}-1&-1\\-1&-1\end{pmatrix}\begin{pmatrix}\hat{\lambda}\left(\hat{\alpha}^T U_i^*;\hat{\alpha},\hat{\beta}\right)-\lambda_0\left(\alpha_0^T U_i^*\right)\\ V_i^T\left(\hat{\beta}-\beta_0\right)\end{pmatrix}+o_p\left(n^{-1/2}\right)\\
&=D_{0i}+\begin{pmatrix}-1\\-1\end{pmatrix}\left\{V_i^T\left(\hat{\beta}-\beta_0\right)\right\}\\
&\quad+\begin{pmatrix}-1\\-1\end{pmatrix}\left\{\hat{\lambda}\left(\hat{\alpha}^T U_i^*;\hat{\alpha},\hat{\beta}\right)-\lambda_0\left(\alpha_0^T U_i^*\right)\right\}+o_p\left(n^{-1/2}\right).
\end{aligned}$$

Since $\hat{\Lambda}_i=\Lambda_i+o_p(1)$, we have

$$\begin{aligned}
0&=\xi\begin{pmatrix}\hat{\alpha}\\0\end{pmatrix}+n^{-1/2}\sum_{i=1}^{n}\Lambda_i D_{0i}+n^{-1/2}\sum_{i=1}^{n}\Lambda_i\begin{pmatrix}-1\\-1\end{pmatrix}\left\{V_i^T\left(\hat{\beta}-\beta_0\right)\right\}\\
&\quad+n^{-1/2}\sum_{i=1}^{n}\Lambda_i\begin{pmatrix}-1\\-1\end{pmatrix}\left\{\hat{\lambda}\left(\hat{\alpha}^T U_i^*;\hat{\alpha},\hat{\beta}\right)-\lambda_0\left(\alpha_0^T U_i^*\right)\right\}+o_p(1).
\end{aligned}\tag{A.9}$$

By Eq. (A.8), we get

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^{n}\Lambda_i\begin{pmatrix}-1\\-1\end{pmatrix}\left\{\hat{\lambda}\left(\hat{\alpha}^T U_i^*;\hat{\alpha},\hat{\beta}\right)-\lambda_0\left(\alpha_0^T U_i^*\right)\right\}\\
&=n^{-1/2}\sum_{i=1}^{n}\Lambda_i\begin{pmatrix}-1\\-1\end{pmatrix}\lambda_0'\left(\alpha_0^T U_i^*\right)U_i^{*T}\left(\hat{\alpha}-\alpha_0\right)\\
&\quad+n^{-1/2}\sum_{i=1}^{n}\Lambda_i\begin{pmatrix}-1\\-1\end{pmatrix}\left\{\hat{\lambda}\left(\hat{\alpha}^T U_i^*;\hat{\alpha},\hat{\beta}\right)-\lambda_0\left(\alpha_0^T U_i^*\right)\right\}+o_p(1).
\end{aligned}$$

Plugging this into Eq. (A.9) gives

$$\begin{aligned}
0&=\xi\begin{pmatrix}\hat{\alpha}\\0\end{pmatrix}+n^{-1/2}\sum_{i=1}^{n}\hat{\Lambda}_i D_{0i}+n^{-1/2}\sum_{i=1}^{n}\hat{\Lambda}_i\begin{pmatrix}-1\\-1\end{pmatrix}\left\{V_i^T\left(\hat{\beta}-\beta_0\right)\right\}\\
&\quad+n^{-1/2}\sum_{i=1}^{n}\hat{\Lambda}_i\begin{pmatrix}-1\\-1\end{pmatrix}\lambda_0'\left(\alpha_0^T U_i^*\right)U_i^*\left(\hat{\alpha}-\alpha_0\right)\\
&\quad+n^{-1/2}\sum_{i=1}^{n}\hat{\Lambda}_i\begin{pmatrix}-1\\-1\end{pmatrix}\left\{\hat{\lambda}\left(\hat{\alpha}^T U_i^*;\hat{\alpha},\hat{\beta}\right)-\lambda_0\left(\alpha_0^T U_i^*\right)\right\}+o_p(1).
\end{aligned}$$

This leads to

$$0 = \xi \begin{pmatrix} \hat{\alpha} \\ 0 \end{pmatrix} + n^{-1/2} \sum_{i=1}^{n} \Lambda_i D_{0i} + n^{-1/2} \sum_{i=1}^{n} \Lambda_i \left[\begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \Lambda_i^T \right] \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix}$$
$$+ n^{-1/2} \sum_{i=1}^{n} \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} \left\{ \hat{\lambda} \left(\alpha_0^T U_i^*; \hat{\alpha}, \hat{\beta} \right) - \lambda_0 \left(\alpha_0^T U_i^* \right) \right\} + o_p(1).$$

Note that by using matrix notation, $L_n(\tilde{u})$ in Eq. (A.2) can be written as

$$L_n(\tilde{u}) = n^{-1} \sum_{i=1}^{n} K_h \left(\tilde{U}_i - \tilde{u} \right) \frac{\epsilon_i^{(I)*}}{f(\tilde{u})} + E\left[\left\{ \Lambda \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\}^T \Big| \tilde{U} = \tilde{u} \right] \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix}.$$

Then from Eq. (A.2) and the definition of A_{α_0,β_0}, we obtain

$$\begin{aligned} 0 = {} & \xi \begin{pmatrix} \hat{\alpha} \\ 0 \end{pmatrix} + n^{-1/2} \sum_{i=1}^{n} \Lambda_i D_{0i} + A_{\alpha_0,\beta_0} n^{1/2} \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} \\ & + n^{-1/2} \sum_{i=1}^{n} \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} E\left[\left\{ \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\}^T \Big| \tilde{U} = \tilde{U}_i \right] \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} \\ & + n^{-1/2} \sum_{i=1}^{n} \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} \left[\frac{n^{-1} \sum_{j=1}^{n} K_h \left(\tilde{U}_j - \tilde{U}_i \right) \epsilon_i^{(I)*}}{f\left(\tilde{U}_i \right)} \right] \\ & + n^{-1/2} \sum_{i=1}^{n} \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} \left\{ O_p\left(h^2\right) I\left[\tilde{U}_i \in D^0 \right] + O_p(h) I\left[\tilde{U}_i \in D^1 \right] \right. \\ & + o_p\left(n^{-1/2}\right) + O_p\left(h^2\right) + o_p(1). \end{aligned} \tag{A.10}$$

It is easy to see that the sixth term is Eq. (A.10) is $O_p\left(\sqrt{n}\, h^2\right) + o_p(1) = o_p(1)$. The fifth term in Eq. (A.10) is essentially the same as (a proof is given in Section I.3)

$$n^{-1/2} \sum_{i=1}^{n} E\left[\left\{ \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\} | \tilde{U} = \tilde{U}_i \right] \epsilon_i^{(I)} + o_p(1). \tag{A.11}$$

From

$$\begin{aligned} & -n^{-1} \sum_{i=1}^{n} \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} E\left[\left\{ \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\}^T | \tilde{U} = \tilde{U}_i \right] \\ & \xrightarrow{p} -E\left[E\left(\left\{ \Lambda \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\} | \tilde{U} \right) E\left(\left\{ \Lambda \begin{pmatrix} -1 \\ -1 \end{pmatrix} \right\}^T | \tilde{U} \right) \right] \\ & = B_{\alpha_0,\beta_0} \end{aligned}$$

and the definition of Q, Eq. (A.10) can be written as

$$\begin{aligned} 0 = {} & \xi \begin{pmatrix} \hat{\alpha} \\ 0 \end{pmatrix} + n^{-1/2} \sum_{i=1}^{n} \left[\Lambda_i \begin{pmatrix} \epsilon_i^{(I)} \\ \epsilon_i^{(I)} \end{pmatrix} + E\left\{ \Lambda_i \begin{pmatrix} -1 \\ -1 \end{pmatrix} | \tilde{U}_i \right\} \epsilon_i^{(I)} \right] \\ & -Q n^{-1/2} \begin{pmatrix} \hat{\alpha} - \alpha_0 \\ \hat{\beta} - \beta_0 \end{pmatrix} + o_p(1). \end{aligned}$$

Multiplying both sides by P_{α_o} and noticing that $\Lambda_i (1,1)^T = \Psi_i$, we obtain the first equality in Eq. (A.3). At the moment, we focus on those auxiliary results required to establish the first equality.

Section I.1. Proofs of (A.5) and (A.7)

Proof of (A.5). Let $\psi^*(\cdot i)$ denote the quantity $\psi\left(a_0(\tilde{u}) + a_1(\tilde{u})\left(\tilde{U}_i - \tilde{u}\right)/h + \beta_0^T V_i\right)$ and let $\psi(\cdot i)$ denote the similar quantity $\psi\left(a_0\left(\tilde{U}_i\right) + \beta_0^T V_i\right)$ for some differential and bounded function $\psi(\cdot)$ or one of the quantities V_i and U_i^* shown up in Eq. (A.5). We will show that

$$\sup_{\tilde{u} \in D} \left| n^{-1} \sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right) \psi^*(\cdot i) / \hat{f}(\tilde{u}) - n^{-1} \sum_{i=1}^{n} K_h\left(\tilde{U}_i - \tilde{u}\right) \psi(\cdot i) / \hat{f}(\tilde{u}) \right| = O_p(h) \tag{A.12}$$

and

$$\sup_{\tilde{u}\in D}\left|n^{-1}\sum_{i=1}^{n}K_h\left(\tilde{U}_i-\tilde{u}\right)\psi\left(\cdot i\right)/\hat{f}\left(\tilde{u}\right)-E\left\{\psi\left(\cdot\right)|\tilde{U}=\tilde{u}\right\}\right|=O_p\left(h\right). \tag{A.13}$$

Equation (A.12) will be used in the proof of Eq. (A.6). First, we assume that Eq. (A.13) holds and we prove Eq. (A.12). Let $\psi'(t)=\partial\psi(t)/\partial t$, then

$$\begin{aligned}\psi^*\left(\cdot i\right)-\psi\left(\cdot i\right) &= \psi\left(\lambda_0\left(\tilde{u}\right)+\lambda_0'\left(\tilde{u}\right)\left(\tilde{U}_i-\tilde{u}\right)+\beta_0^TV_i\right)-\psi\left(\lambda_0\left(\tilde{U}_i\right)+\beta_0^TV_i\right)\\ &= \psi'\left(\xi_i\left(\tilde{u}\right)\right)\left[\lambda_0\left(\tilde{u}\right)-\lambda_0\left(\tilde{U}_i\right)+\lambda_0'\left(\tilde{u}\right)\left(\tilde{U}_i-\tilde{u}\right)\right]\\ &= \psi'\left(\xi_i\left(\tilde{u}\right)\right)\left[-\left\{\lambda_0''\left(\xi_i\left(\tilde{u}\right)\right)/2\right\}\left(\tilde{U}_i-\tilde{u}\right)^2\right]\\ &= O_p\left(\left(\tilde{U}_i-\tilde{u}\right)^2\right),\end{aligned} \tag{A.14}$$

where $\xi_i(\tilde{u})$ is between $\lambda_0\left(\tilde{U}_i\right)+\beta_0^TV_i$ and $\lambda_0\left(\tilde{u}\right)+\lambda_0'\left(\tilde{u}\right)\left(\tilde{U}_i-\tilde{u}\right)+\beta_0^TV_i$, $\xi_i(\tilde{u})$ is between $\tilde{U}_i$ and $\tilde{u}$. Therefore

$$\begin{aligned}&\sup_{\tilde{u}\in D}\left|n^{-1}\sum_{i=1}^{n}K_h\left(\tilde{U}_i-\tilde{u}\right)\psi^*\left(\cdot i\right)/\hat{f}\left(\tilde{u}\right)-n^{-1}\sum_{i=1}^{n}K_h\left(\tilde{U}_i-\tilde{u}\right)\psi\left(\cdot i\right)/\hat{f}\left(\tilde{u}\right)\right|\\ &\le O_p\left(1\right)\sup_{\tilde{u}\in D}\left|n^{-1}\sum_{i=1}^{n}K_h\left(\tilde{U}_i-\tilde{u}\right)\left(\tilde{U}_i-\tilde{u}\right)^2/\hat{f}\left(\tilde{u}\right)\right|=O_p\left(h\right).\end{aligned}$$

using Eq. (A.13) by taking $\psi(\cdot i)=\left(\tilde{U}_i-\tilde{u}\right)^2$ and noticing that $E\left\{\left(\tilde{U}_i-\tilde{u}\right)^2|\tilde{U}_i=\tilde{u}\right\}=0$, this proves Eq. (A.12).

Now we prove Eq. (A.13). Let $\hat{r}_h(\tilde{u})=n^{-1}\sum_{i=1}^{n}K_h\left(\tilde{U}_i-\tilde{u}\right)\psi(\cdot i)$, then

$$\begin{aligned}&n^{-1}\sum_{i=1}^{n}K_h\left(\tilde{U}_i-\tilde{u}\right)\psi\left(\cdot i\right)/\hat{f}\left(\tilde{u}\right)-E\left\{\psi\left(\cdot\right)|\tilde{U}=\tilde{u}\right\}\\ &\quad=\frac{\left[\hat{r}_h\left(\tilde{u}\right)-E\left\{\hat{r}_h\left(\tilde{u}\right)\right\}\right]E\left\{\hat{f}\left(\tilde{u}\right)\right\}-\left[\hat{f}\left(\tilde{u}\right)-E\left\{\hat{f}\left(\tilde{u}\right)\right\}\right]E\left\{\hat{r}_h\left(\tilde{u}\right)\right\}}{\left[\hat{f}\left(\tilde{u}\right)-E\left\{\hat{f}\left(\tilde{u}\right)\right\}+E\left\{\hat{f}\left(\tilde{u}\right)\right\}\right]E\left\{\hat{f}\left(\tilde{u}\right)\right\}}\\ &\qquad+\left[\frac{E\left\{\hat{r}_h\left(\tilde{u}\right)\right\}}{E\left\{\hat{f}\left(\tilde{u}\right)\right\}}-E\left\{\psi\left(\cdot\right)|\tilde{U}=\tilde{u}\right\}\right]\\ &\quad\equiv I_1\left(\tilde{u}\right)+I_2\left(\tilde{u}\right).\end{aligned} \tag{A.15}$$

We consider $I_2(\tilde{u})$ first. Since

$$\begin{aligned}E\left\{\hat{r}_h\left(\tilde{u}\right)\right\} &= E\left\{K_h\left(\tilde{U}_i-\tilde{u}\right)\psi\left(\cdot i\right)\right\}=E\left[K_h\left(\tilde{U}_i-\tilde{u}\right)E\left\{\psi\left(\cdot i\right)|\tilde{U}_i\right\}\right]\\ &= \frac{1}{h}\int_c^d K\left(\frac{y-\tilde{u}}{h}\right)E\{\psi(\cdot)|\tilde{U}=y\}f(y)dy\\ &= \int_{\frac{c-\tilde{u}}{h}}^{\frac{d-\tilde{u}}{h}}K\left(t\right)E\left\{\psi\left(\cdot\right)|\tilde{U}=\tilde{u}+ht\right\}f\left(\tilde{u}+ht\right)dt\\ &= \left\{\int_{\frac{c-\tilde{u}}{h}}^{\frac{d-\tilde{u}}{h}}K\left(t\right)dt\right\}E\left\{\psi\left(\cdot\right)|\tilde{U}=\tilde{u}\right\}f\left(\tilde{u}\right)+O\left(h\right)\end{aligned}$$

and

$$E\left\{\hat{f}\left(\tilde{u}\right)\right\}=\left\{\int_{\frac{c-\tilde{u}}{h}}^{\frac{d-\tilde{u}}{h}}K\left(t\right)dt\right\}f\left(\tilde{u}\right)+O\left(h\right)$$

hold uniformly for $\tilde{u}\in D$, we have

$$\sup_{\tilde{u}\in D}\left|I_2\left(\tilde{u}\right)\right|=O\left(h\right).$$

To finish the proof, it suffices to show

$$\sup_{\tilde{u}\in D}\left|\hat{r}_h\left(\tilde{u}\right)-E\left\{\hat{r}_h\left(\tilde{u}\right)\right\}\right|=O_p\left(h\right), \tag{A.16}$$

$$\sup_{\tilde{u}\in D}|\hat{f}(\tilde{u})-E\left\{\hat{f}(\tilde{u})\right\}|=O_p(h). \tag{A.17}$$

We prove Eq. (A.16) but Eq. (A.17), since Eq. (A.17) is very easy to prove. We consider a more general case where $\psi(\cdot i)$ might be unbounded but $|\psi(\cdot)|\leq C_\psi C_T T_i^g$ for some constant C_ψ, C_T, $g>0$ and some i.i.d. random variables T_i for which $\sup_{i,\tilde{u}\in D}E\left\{T_i^{(2s+1)g}\middle|\tilde{U}_i=\tilde{u}\right\}<\infty$ and $\sup_i E\left\{T_i^{(2s+1)g}\right\}<\infty$ for some $s>1$. Taking $N_n=h^{-1/s}$ and writing

$$\begin{aligned}\hat{r}_h(\tilde{u}) &= n^{-1}\sum_{i=1}^n K_h(\tilde{U}_i-\tilde{u})\psi(\cdot i)I\left[|\psi(\cdot i)|\leq N_n\right]\\ &\quad+n^{-1}\sum_{i=1}^n K_h(\tilde{U}_i-\tilde{u})\psi(\cdot i)I\left[|\psi(\cdot i)|>N_n\right]\\ &\equiv J_1(\tilde{u})+J_2(\tilde{u}),\end{aligned}$$

it suffices to shown

$$\sup_{\tilde{u}\in D}|J_1(\tilde{u})-E\{J_1(\tilde{u})\}|=O_p(h) \tag{A.18}$$

and

$$\sup_{\tilde{u}\in D}|J_2(\tilde{u})-E\{J_2(\tilde{u})\}|=O_p(h). \tag{A.19}$$

When $\psi(\cdot i)$ is bounded, Eq. (A.19) is trivial.

Suppose that M_n intervals $\{\tilde{u}:|\tilde{u}-\tilde{u}_l|\leq\eta_n\}, l=1,2,\ldots,M_n$, cover the compact set D and the union of the these intervals equals D. Then, for any $\nabla>0$,

$$\begin{aligned}&P\left\{\sup_{\tilde{u}\in D}|J_1(\tilde{u})-E\{J_1(\tilde{u})\}|>\nabla h\right\}\\ &=P\left\{\sup_{l=1,\ldots,M_n}\sup_{|\tilde{u}-\tilde{u}_l|\leq\eta_n}|J_1(\tilde{u})-E\{J_1(\tilde{u})\}|>\nabla h\right\}\\ &\leq P\left\{\sup_{l=1,\ldots,M_n}|J_1(\tilde{u}_l)-E\{J_1(\tilde{u}_l)\}|>\frac{\nabla}{2}h\right\}\\ &\quad+P\left\{\sup_{l=1,\ldots,M_n}\sup_{|\tilde{u}-\tilde{u}_l|\leq\eta_n}|J_1(\tilde{u})-J_1(\tilde{u}_l)-(E\{J_1(\tilde{u})\}-E\{J_1(\tilde{u}_l)\})|>\frac{\nabla}{2}h\right\}.\end{aligned} \tag{A.20}$$

By Condition A(ii), there exists some constants $C_K>0$ and $C_L>0$ such that $|K(u^*)|\leq C_K$ and $|K(u_1^*)-K(u_2^*)|\leq C_L|u_1^*-u_2^*|$. Taking $M_n=O(n^2)$ and $\eta_n=O(n^{-2})$, when $|\tilde{u}-\tilde{u}_l|\leq\eta_n$, we have

$$\begin{aligned}|J_1(\tilde{u})-J_1(\tilde{u}_l)| &= |(nh)^{-1}\sum_{i=1}^n\left\{K\left(\frac{\tilde{U}_i-\tilde{u}}{h}\right)-K\left(\frac{\tilde{U}_i-\tilde{u}_l}{h}\right)\right\}\psi(\cdot i)I[\psi(\cdot i)\leq N_n]|\\ &\leq (nh)^{-1}C_L\left|\frac{\tilde{u}-\tilde{u}_l}{h}\right|nN_n\\ &= C_Lh^{-(2+1/s)}\eta_n=O\left(\left(nh^3\right)^{-2}\right)\cdot O\left(h^{4-1/s}\right)=o_p(h).\end{aligned}$$

Therefore, $\sup_{l=1,\ldots,M_n}\sup_{|\tilde{u}-\tilde{u}_l|\leq\eta_n}|J_1(\tilde{u})-J_1(\tilde{u}_l)|=o_p(h)$.

Similarly, $\sup_{l=1,\ldots,M_n}\sup_{|\tilde{u}-\tilde{u}_l|\leq\eta_n}|E\{J_1(\tilde{u})\}-E\{J_1(\tilde{u}_l)\}|=o(h)$. Hence, the second probability in Eq. (A.20) is negligible. Let $d_i(\tilde{u})=K\left(\frac{\tilde{U}_i-\tilde{u}}{h}\right)\psi(\cdot i)I[|\psi(\cdot i)|\leq N_n]$ and $S_n(\tilde{u})=\sum_{i=1}^n[d_i(\tilde{u})-E\{d_i(\tilde{u})\}]$. Then, $|d_i(\tilde{u})-E\{d_i(\tilde{u})\}|\leq 2C_KN_n$ and $\sigma_n^2=Var(S_n(\tilde{u}))=n\left[E\left\{K^2\left(\frac{\tilde{U}_i-\tilde{u}}{h}\right)\psi^2(\cdot i)I[|\psi(\cdot i)|\leq N_n]\right\}-\left\{E\left\{K\left(\frac{\tilde{U}_i-\tilde{u}}{h}\right)\psi(\cdot i)I[|\psi(\cdot i)|\leq N_n]\right\}\right\}^2\right]=O(nh)-O\left(nh^2\right)=O(nh)$, because $E\left\{\psi^2(\cdot i)|\tilde{U}_i=\tilde{u}\right\}\leq C_\psi^2C_T^2E\left\{T_i^{2g}|\tilde{U}_i=\tilde{u}\right\}<M<\infty$ for some constants $C_\psi>0$ and $M>0$ by the the preceding assumptions. Without loss of generality, we assume $\sigma_n^2=nh$. By Bernstein's inequality, for any $\omega>0$, we get

$$P\left(|S_n(\tilde{u}_l)|\geq\omega\sigma_n\right)\leq 2exp\left[-\frac{\omega^2}{2+\frac{2}{3}\frac{2C_KN_n}{\sigma_n}\omega}\right].$$

Taking $\omega = (\nabla h\sigma_n)/2$ and noticing $\sigma_n = \sqrt{nh}$ and $N_n = h^{-1/s}$, $s > 1$, we get

$$\begin{aligned} P\left(|S_n(\tilde{u}_l)| \geq (nh)(\nabla/2)h\right) &\leq 2exp\left[-\frac{\left(\frac{\nabla h\sigma_n}{2}\right)^2}{2+\frac{2}{3}\frac{2C_K N_n}{\sigma_n}\left(\frac{\nabla h\sigma_n}{2}\right)}\right] \\ &= 2exp\left[-\frac{\left(\frac{\nabla h\sigma_n}{2}\right)^2}{2+\frac{2}{3}(C_K\nabla)h^{1-1/s}}\right] \\ &= 2exp\left[-\frac{\left(\frac{\nabla}{2}\right)^2 O\left(nh^3\right)}{2+\frac{2}{3}(C_K\nabla)h^{1-1/s}}\right] \\ &\leq O\left(2n^{-(3/32)\nabla^2}\right)\ \left(Assume\, C_K\nabla h^{1-1/s} < 1\right). \end{aligned}$$

Since $M_n = O\left(n^2\right)$, when ∇ is large enough so that $\frac{3}{32}\nabla^2 > 2$, we get

$$P\left\{\sup_{l=1,\ldots,M_n}\left|\sum_{i=1}^{n}\left[d_i(\tilde{u}_l) - E\{d_i(\tilde{u}_l)\}\right]\right| \geq \left(\frac{\nabla h}{2}\right)\sigma_n^2\right\} \leq M_n O\left(2n^{-(3/32)\nabla^2}\right) \overset{n\to\infty}{\to} 0.$$

This implies

$$\sup_{l=1,\ldots,M_n}\left|\frac{1}{nh}\sum_{i=1}^{n}\left[d_i(\tilde{u}_l) - E\{d_i(\tilde{u}_l)\}\right]\right| = O_p(h). \tag{A.21}$$

Combining Eqs. (A.20) and (A.21) proves Eq. (A.18).

Now we prove Eq. (A.19). By Conditions A, $|\psi(\cdot i)| \leq C_\psi C_T T_i^g$ for some constant C_T, $C_\psi > 0$. From $|K(u^*) \leq C_K|$, we have

$$\sup_{\tilde{u}\in D}|J_2(\tilde{u}) - E\{J_2(\tilde{u})\}| = \frac{2C_\psi C_K C_T}{h}\frac{1}{n}\sum_{i=1}^{n} T_i^g I\left[T_i^g > N_n\right] \tag{A.22}$$

since $E\left\{T_i^g I\left[T_i^g > N_n\right]\right\} = \int_{t>N_n^{1/g}} t^g dF_T(t)$, where $F_T(t)$ is the c.d.f. of T, and $N_n = h^{-1/s}$, $s > 1$. Let $Q_n = N_n^{1/g} = h^{-1/(sg)}$, we get

$$\begin{aligned} \frac{\int_{t>Q_n} t^g dF_T(t)}{h^2} &= \frac{\int_{t>Q_n} t^g dF_T(t)}{Q_n^{-2sg}} \\ &\leq \int_{t>Q_n} t^{(2s+1)g} dF_T(t) \overset{n\to\infty}{\to} 0, \end{aligned}$$

because $E\left\{T^{(2s+1)g}\right\} < \infty$ by the preceding assumptions. This implies $(1/nh)\sum_{i=1}^{n} TigI\left[T_i^g > N_n\right] = O_p(h)$. Therefore, by Eq. (A.22), we obtain Eq. (A.19).

Proof of (A.7). Since

$$\sup_{\tilde{u}\in D^0}|\hat{f}(\tilde{u}) - f(\tilde{u})| \leq \sup_{\tilde{u}\in D^0}|\hat{f}(\tilde{u}) - E\left\{\hat{f}(\tilde{u})\right\}| + \sup_{\tilde{u}\in D^0}|E\left\{\hat{f}(\tilde{u})\right\} - f(\tilde{u})|,$$

$$\sup_{\tilde{u}\in D^1}|\hat{f}(\tilde{u}) - f(\tilde{u})| \leq \sup_{\tilde{u}\in D^1}|\hat{f}(\tilde{u}) - E\left\{\hat{f}(\tilde{u})\right\}| + \sup_{\tilde{u}\in D^1}|E\left\{\hat{f}(\tilde{u})\right\} - f(\tilde{u})|,$$

$$\sup_{\tilde{u}\in D^1}|E\left\{\hat{f}(\tilde{u})\right\} - f(\tilde{u})| \leq \sup_{\tilde{u}\in D^1}|E\left\{\hat{f}(\tilde{u})\right\}| + \sup_{\tilde{u}\in D^1}|f(\tilde{u})| = O(1),$$

using Eq. (A.17) and noticing that $\sup_{\tilde{u}\in D^0}|E\left\{\hat{f}(\tilde{u})\right\} - f(\tilde{u})| = O\left(h^2\right)$, we obtain Eq. (A.7).

Section I.2. Proof of (A.6)

Proof of (A.6).

Since $\epsilon_i^{(I)*} = Y_i^{(I)} - \{\lambda_0(\tilde{u}) + \lambda_0'(\tilde{u})(\tilde{U}_i - \tilde{u}) + \beta_0^T V_i\}$, it suffices to show

$$\sup_{\tilde{u}\in D} | n^{-1}\sum_{i=1}^{n} K_h(\tilde{U}_i - \tilde{u})\,\epsilon_i^{(I)*}/\hat{f}(\tilde{u}) - n^{-1}\sum_{i=1}^{n} K_h(\tilde{U}_i - \tilde{u})\,\epsilon_i^{(I)}/\hat{f}(\tilde{u}) \,| = O_p(h) \tag{A.23}$$

and

$$\sup_{\tilde{u}\in D} | n^{-1}\sum_{i=1}^{n} K_h(\tilde{U}_i - \tilde{u})\,\epsilon_i^{(I)}/\hat{f}(\tilde{u}) - 0 \,| = O_p(h). \tag{A.24}$$

The proof of Eq. (A.23) is similar to that of Eq. (A.12), we omit it. Now we prove Eq. (A.24). Note that $E\left\{n^{-1}\sum_{i=1}^{n} K_h(\tilde{U}_i - \tilde{u})\,\epsilon_i^{(I)}\right\} = 0$, by the same arguments used in the proof of Eq. (A.13), it suffices to show

$$\sup_{\tilde{u}\in D} | n^{-1}\sum_{i=1}^{n} K_h(\tilde{U}_i - \tilde{u})\,\epsilon_i^{(I)} - E\left\{n^{-1}\sum_{i=1}^{n} K_h(\tilde{U}_i - \tilde{u})\,\epsilon_i^{(I)}\right\} | = O_p(h). \tag{A.25}$$

By decomposition, it suffices to show

$$\sup_{\tilde{u}\in D} | n^{-1}\sum_{i=1}^{n} K_h(\tilde{U}_i - \tilde{u})\,\lambda_0(\tilde{U}_i) - E\left\{n^{-1}\sum_{i=1}^{n} K_h(\tilde{U}_i - \tilde{u})\,\lambda_0(\tilde{U}_i)\right\} | = O_p(h), \tag{A.26}$$

$$\sup_{\tilde{u}\in D} | n^{-1}\sum_{i=1}^{n} K_h(\tilde{U}_i - \tilde{u})\,\beta_0^T V_i - E\left\{n^{-1}\sum_{i=1}^{n} K_h(\tilde{U}_i - \tilde{u})\,\beta_0^T V_i\right\} | = O_p(h) \tag{A.27}$$

and

$$\sup_{\tilde{u}\in D} | n^{-1}\sum_{i=1}^{n} K_h(\tilde{U}_i - \tilde{u})\,Y_i^{(I)} - E\left\{n^{-1}\sum_{i=1}^{n} K_h(\tilde{U}_i - \tilde{u})\,Y_i^{(I)}\right\} | = o_p(1). \tag{A.28}$$

We shall apply the similar techniques used in the proof of Eq. (A.16) to prove the preceding three equalities. By Conditions A(ii) and (iv), $\lambda_0(\tilde{U}_i)$ and $\beta_0^T V_i$ are bounded random variables, the proofs of Eqs. (A.26) and (A.27) are straightforward. And we can obtain Eq. (A.28) by observing that

$$E\left\{n^{-1}\sum_{i=1}^{n}\left[K_h(\tilde{U}_i - \tilde{u})\,Y_i^{(I)} - E\left\{K_h(\tilde{U}_i - \tilde{u})\,Y_i^{(I)}\right\}\right]\right\}^2$$

$$= n^{-2}\sum_{i=1}^{n} E\left\{K_h(\tilde{U}_i - \tilde{u})\,Y_i^{(I)} - E\left\{K_h(\tilde{U}_i - \tilde{u})\,Y_i^{(I)}\right\}\right\}^2 \to 0. \tag{A.29}$$

Section I.3. Proof of (A.11)

Proof of Eq. (A.11).

To prove Eq. (A.11) is equivalent to prove the following equality:

$$\epsilon_i^{(I)*} = \epsilon_i^{(I)} + o_p(1). \tag{A.30}$$

Noting that

$$\epsilon_i^{(I)} = Y_i^{(I)} - E\left\{Y_i^{(I)} | V_i, U_i^*\right\} = Y_i^{(I)} - \{\lambda_0(\tilde{U}_i) + \beta_0^T V_i\}$$

and

$$\epsilon_i^{(I)*} = Y_i^{(I)} - \{\lambda_0(\tilde{u}) + \lambda_0'(\tilde{u})(\tilde{U}_i - \tilde{u}) + \beta_0^T V_i\},$$

we have

$$\begin{aligned}\epsilon_i^{(I)} - \epsilon_i^{(I)*} &= \lambda_0(\tilde{u}) + \lambda_0'(\tilde{u})\left(\tilde{U}_i - \tilde{u}\right) - \lambda_0\left(\tilde{U}_i\right) \\ &= O\left(\left(\tilde{U}_i - \tilde{u}\right)^2\right) = O\left(h^2\right).\end{aligned}$$

It implies that $n^{-1}\sum_{i=1}^n \epsilon_i^{(I)*} = n^{-1}\sum_{i=1}^n \epsilon_i^{(I)} + o_p(1)$, which leads to Eq. (A.30).

Now we return to the proof of Theorem 1. In Eq. (A.3), $\Psi_i - E\{\Psi_i|\tilde{U}_i\}$ is a vector with $p+q$ elements. Let $H_i = \Psi_i - E\{\Psi_i|\tilde{U}_i\}$ and suppose its elements are $H_{i,l} = H_{i,l}\left(U_i^*, V_i, \tilde{U}_i\right)$, $l = 1, 2, \ldots, p+q$, then we consider

$$M_{1n}^l = n^{-1/2}\sum_{i=1}^n H_{i,l}\epsilon_i^{(I)}, \quad l = 1, 2, \ldots, p+q. \tag{A.31}$$

Therefore, by Eq. (A.31), we have shown that

$$\lim_{n\to\infty} E\left[\{M_{1n}\}^{\otimes 2}\right] = \Omega, \tag{A.32}$$

where $M_{1n} = \left(M_{1n}^1, \ldots, M_{1n}^{p+q}\right)^T$. Theorem 1 is then proved by the central limit theorem for sums of independent random vectors.

APPENDIX II

Table A.1 | Descriptive statistics of $\left(\hat{\alpha}_0^{(I)}, \hat{\beta}_0^{(I)}\right)$ with missing response as a function of missing proportion p and sample sizes.

α_0	MEAN	SD	RMSE	MED	β_0	MEAN	SD	RMSE	MED
$p = 10\%$, $n = 60$									
0.707	0.710	0.061	0.061	0.707	−1	−0.998	0.148	0.147	−0.999
0.707	0.699	0.064	0.065	0.707	2	2.003	0.149	0.148	2.007
$p = 22\%$, $n = 120$									
0.707	0.709	0.044	0.044	0.707	−1	−0.995	0.107	0.107	−0.994
0.707	0.703	0.045	0.045	0.707	2	2.005	0.114	0.114	2.005
$p = 49\%$, $n = 240$									
0.707	0.704	0.038	0.038	0.707	−1	−1.002	0.096	0.096	−0.998
0.707	0.708	0.037	0.037	0.707	2	2.006	0.095	0.095	2.003

MED, median; SD, standard derivation ; RMSE, root-mean-square error.

Table A.2 | Descriptive statistics of $\left(\hat{\alpha}_0^{(I)}, \hat{\beta}_0^{(I)}\right)$ with missing response and error-prone $\left(\sigma_e^2 = \sqrt{3}/4\right)$ predictors when W calibrated and primary size n', validation size m, and missing proportion p.

α_0	MEAN	SD	RMSE	MED	β_0	MEAN	SD	RMSE	MED
$p = 10\%$, $n' = 60$, $m = 20$									
0.707	0.686	0.165	0.166	0.707	−1	−0.992	0.202	0.202	−1.009
0.707	0.682	0.192	0.193	0.707	2	2.011	0.194	0.194	2.011
$p = 22\%$, $n' = 120$, $m = 40$									
0.707	0.673	0.127	0.131	0.707	−1	−0.995	0.152	0.152	−0.989
0.707	0.718	0.122	0.122	0.707	2	2.007	0.149	0.149	2.014
$p = 49\%$, $n' = 240$, $m = 80$									
0.707	0.665	0.105	0.113	0.695	−1	−1.005	0.137	0.136	−1.001
0.707	0.733	0.095	0.098	0.719	2	1.993	0.127	0.127	1.993

MED, median; SD, standard derivation ; RMSE, root-mean-square error.

Table A.3 Descriptive statistics of $\left(\hat{\alpha}_0^{(I)}, \hat{\beta}_0^{(I)}\right)$ with missing response and error-prone $\left(\sigma_e^2 = \sqrt{3}/4\right)$ predictors when W, not calibrated and primary size, validation size and missing proportion.

α_0	MEAN	SD	RMSE	MED	β_0	MEAN	SD	RMSE	MED
$p = 10\%$, $n' = 60$, $m = 20$									
0.707	0.741	0.105	0.110	0.715	−1	−1.000	0.206	0.205	−1.003
0.707	0.651	0.129	0.140	0.699	2	1.973	0.203	0.204	1.964
$p = 22\%$, $n' = 120$, $m = 40$									
0.707	0.747	0.080	0.089	0.740	−1	−0.977	0.138	0.140	−0.981
0.707	0.653	0.097	0.111	0.673	2	1.995	0.150	0.150	1.987
$p = 49\%$, $n' = 240$, $m = 80$									
0.707	0.752	0.065	0.079	0.748	−1	−1.017	0.127	0.128	−1.014
0.707	0.651	0.077	0.095	0.663	2	1.984	0.127	0.128	1.985

MED, median; SD, standard derivation ; RMSE, root-mean-square error.

Table A.4 Descriptive statistics of $\left(\hat{\alpha}_0^{(R)}, \hat{\beta}_0^{(R)}\right)$ with missing response as a function of missing proportion and sample sizes n.

α_0	MEAN	SD	RMSE	MED	β_0	MEAN	SD	RMSE	MED
$p = 10\%$, $n = 60$									
0.707	0.709	0.046	0.046	0.707	−1	−1.002	0.144	0.144	−1.004
0.707	0.702	0.049	0.049	0.707	2	1.992	0.151	0.151	1.985
$p = 22\%$, $n = 120$									
0.707	0.710	0.040	0.040	0.707	−1	−0.989	0.108	0.108	−0.992
0.707	0.701	0.042	0.042	0.707	2	2.000	0.102	0.102	1.996
$p = 49\%$, $n = 240$									
0.707	0.710	0.038	0.039	0.707	−1	−0.994	0.097	0.097	−0.988
0.707	0.702	0.041	0.041	0.707	2	2.004	0.092	0.092	2.005

MED, median; SD, standard derivation ; RMSE, root-mean-square error.

Table A.5 Descriptive statistics of $\left(\hat{\alpha}_0^{(R)}, \hat{\beta}_0^{(R)}\right)$ with missing response and error-prone $\left(\sigma_e^2 = \sqrt{3}/4\right)$ predictors when calibrated and primary size, validation size and missing proportion .

α_0	MEAN	SD	RMSE	MED	β_0	MEAN	SD	RMSE	MED
$p = 10\%$, $n' = 60$, $m = 20$									
0.707	0.696	0.136	0.136	0.707	−1	−1.005	0.230	0.230	−0.985
0.707	0.689	0.151	0.151	0.707	2	2.002	0.183	0.183	2.003
$p = 22\%$, $n' = 120$, $m = 40$									
0.707	0.672	0.113	0.118	0.707	−1	−1.008	0.168	0.168	−1.001
0.707	0.724	0.109	0.110	0.707	2	1.999	0.154	0.154	2.002
$p = 49\%$, $n' = 240$, $m = 80$									
0.707	0.675	0.105	0.109	0.707	−1	−1.014	0.129	0.130	−1.019
0.707	0.724	0.102	0.103	0.707	2	1.988	0.133	0.133	1.995

MED, median; SD, standard derivation ; RMSE, root-mean-square error.

Table A.6 Descriptive statistics of $\left(\hat{\alpha}_0^{(R)}, \hat{\beta}_0^{(R)}\right)$ with missing response and error-prone $\left(\sigma_e^2 = \sqrt{3}/4\right)$ predictors when not calibrated and primary size n', validation size m and missing proportion.

α_0	MEAN	SD	RMSE	MED	β_0	MEAN	SD	RMSE	MED
$p = 10\%$, $n' = 60$, $m = 20$									
0.707	0.748	0.095	0.104	0.729	−1	−0.977	0.195	0.196	−0.965
0.707	0.645	0.125	0.140	0.684	2	2.021	0.210	0.210	2.021
$p = 22\%$, $n' = 120$, $m = 40$									
0.707	0.751	0.070	0.083	0.738	−1	−1.001	0.147	0.146	−0.998
0.707	0.651	0.087	0.103	0.674	2	2.008	0.161	0.161	2.011
$p = 49\%$, $n' = 240$, $m = 80$									
0.707	0.754	0.073	0.087	0.750	−1	−1.013	0.136	0.137	−1.016
0.707	0.647	0.087	0.106	0.661	2	1.984	0.134	0.135	1.996

MED, median; SD, standard derivation ; RMSE, root-mean-square error.

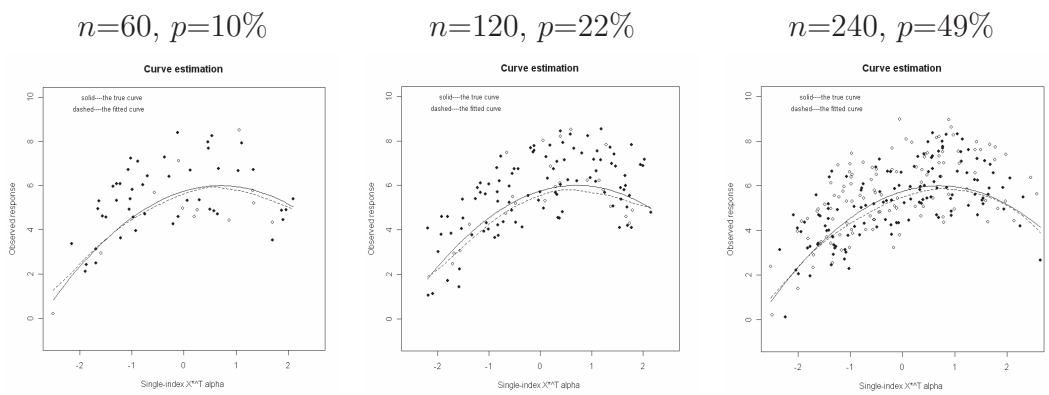

Figure A.1 | Simulated curves of $\hat{\lambda}_0^{(I)}$ with missing response, different sample sizes n and different missing proportions p (the title for the x axis: Single-index $\hat{\alpha}_0^{(I)T}X$, solid circle: the response is observed, circle: the response is missing, solid line: the true curve, dashed line: the fitted curve).

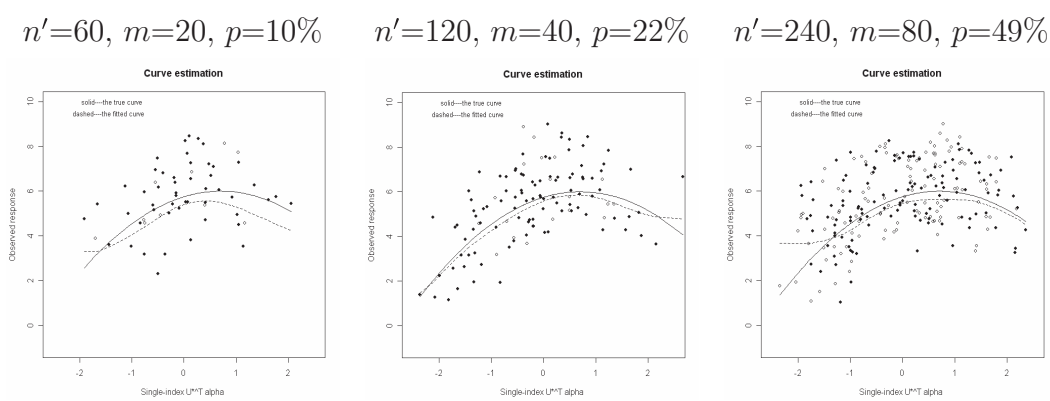

Figure A.2 | Simulated curves of $\hat{\lambda}_0^{(I)}$ with missing response and error-prone $\left(\sigma_e^2 = \sqrt{3}/4\right)$ predictors when W calibrated and primary size n', validation size m and missing proportion p (the title for the x axis: Single-index $\hat{\alpha}_0^{(I)T}U^*$, solid circle: the response is observed, circle: the response is missing, solid line: the true curve, dashed line: the fitted curve).

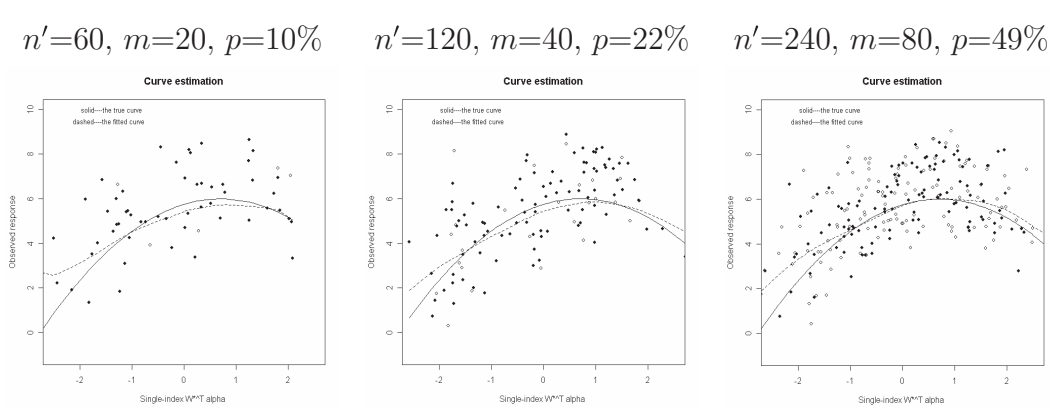

Figure A.3 | Simulated curves of $\hat{\lambda}_0^{(I)}$ with missing response and error-prone $\left(\sigma_e^2 = \sqrt{3}/4\right)$ predictors when W not calibrated and primary size n', validation size m and missing proportion p (the title for the x axis: Single-index $\hat{\alpha}_0^{(I)T}W$, solid circle: the response is observed, circle: the response is missing, solid line: the true curve, dashed line: the fitted curve).

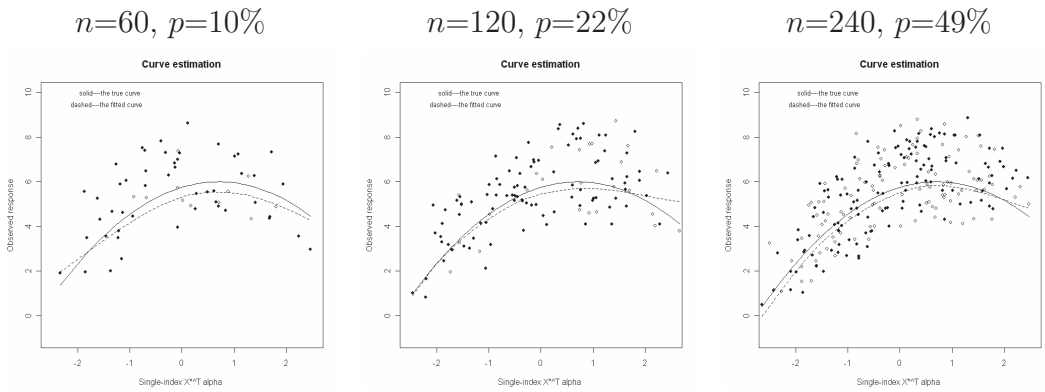

Figure A.4 | Simulated curves of $\hat{\lambda}_0^{(R)}$ with missing response, different sample sizes n and different missing proportions p (the title for the x axis: Single-index $\hat{\alpha}_0^{(R)T}X$, solid circle: the response is observed, circle: the response is missing, solid line: the true curve, dashed line: the fitted curve).

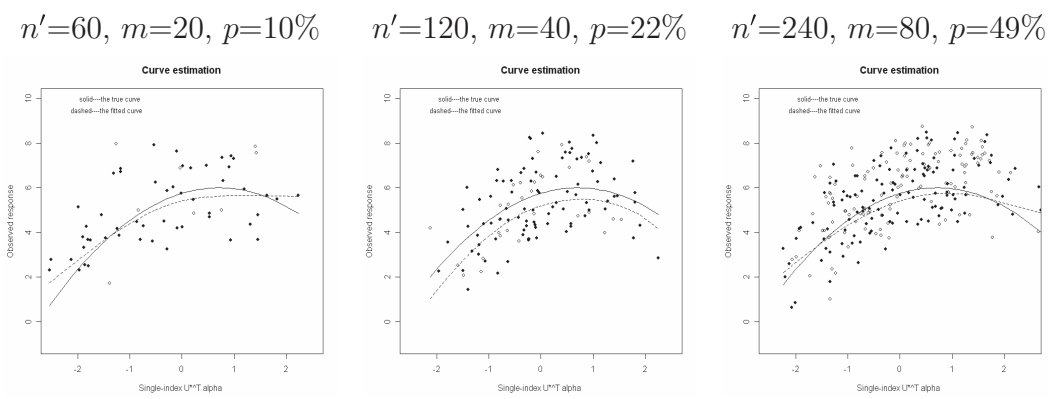

Figure A.5 | Simulated curves of $\hat{\lambda}_0^{(R)}$ with missing response and error-prone $\left(\sigma_\epsilon^2 = \sqrt{3}/4\right)$ predictors when W calibrated and primary size n', validation size m and missing proportion p (the title for the x axis: Single-index $\hat{\alpha}_0^{(R)T}U^*$, solid circle: the response is observed, circle: the response is missing, solid line: the true curve, dashed line: the fitted curve).

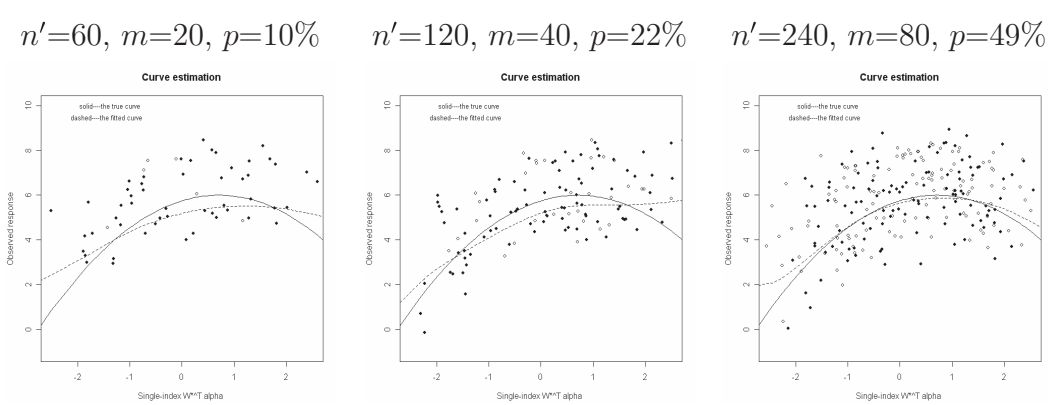

Figure A.6 | Simulated curves of $\hat{\lambda}_0^{(R)}$ with missing response and error-prone $\left(\sigma_\epsilon^2 = \sqrt{3}/4\right)$ predictors when W calibrated and primary size n', validation size m and missing proportion p (the title for the x axis: Single-index $\hat{\alpha}_0^{(R)T}U^*$, solid circle: the response is observed, circle: the response is missing, solid line: the true curve, dashed line: the fitted curve).