Computational Statistics (2020) 35:1637–1658
https://doi.org/10.1007/s00180-020-00967-3

ORIGINAL PAPER

Estimation in partially linear varying-coefficient errors-in-variables models with missing response variables

Yan-Ting Xiao[1] · **Fu-Xiao Li**[1]

Received: 22 February 2018 / Accepted: 5 February 2020 / Published online: 14 February 2020

Abstract
In this paper, a partially linear varying-coefficient model with measurement errors in the nonparametric component as well as missing response variable is studied. Two estimators for the parameter vector and nonparametric function are proposed based on the locally corrected profile least squares method. The first estimator is constructed by using the complete-case data only, and another by using an imputation technique. Both proposed estimators of the parametric component are shown to be asymptotically normal, and the estimators of nonparametric function are proved to achieve the optimal strong convergence rate as the usual nonparametric regression. Some simulation studies are conducted to compare the behavior of these estimators and the results confirm that the estimators based on the imputation technique perform better than the complete-case data estimator in finite samples. Finally, an application to a real data set is illustrated.

Keywords Partially linear varying-coefficient models · Measurement error · Missing response · Locally corrected profile least squares · Imputation technique

1 Introduction

The partially linear varying-coefficient model, as a very important semi-parametric model, takes the form as

$$Y = \mathbf{X}^T \boldsymbol{\beta} + \mathbf{Z}^T \boldsymbol{\alpha}(U) + \varepsilon, \tag{1}$$

where Y is the response variable, $\mathbf{X} \in R^p$, $\mathbf{Z} \in R^q$ and U are the associated covariates, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ is a p-dimensional vector of unknown parameter and $\boldsymbol{\alpha}(.) = (\alpha_1(.), \ldots, \alpha_q(.))^T$ is a q-dimensional vector of unknown coefficient function, ε is

✉ Yan-Ting Xiao
xiaoyanting03@163.com

[1] Department of Applied Mathematics, Xi'an University of Technology, Xi'an 710048, China

the random error that is assumed to be independent of $(U, \mathbf{X}, \mathbf{Z})$ with mean zero and finite variance σ^2. Since model (1) keeps both the interpretation power of parametric model and the flexibility of nonparametric model, it has been extensively studied by researches (Ahmad et al. 2005; Fan and Huang 2005; Kai et al. 2011; Long et al. 2013; You and Zhou 2006; Zhang et al. 2002; among others).

With the development of science and technology, the study of data with missing observations has been attracted more attention in various scientific fields, such as economics, engineering, biology and epidemiology. Dealing with missing data, several problems may arise when traditional statistical inference procedures for complete data sets are applied directly. There has been extensive research on statistical models with missing observations. In the partially linear model with the missing response data, Wang et al. (2004) proposed a class of semiparametric estimators for the regression coefficient and the response mean. Wang and Sun (2007) developed the imputation, semi-parametric surrogate regression and inverse marginal probability weighted methods to estimate unknown parameter. Xue and Xue (2011) proposed the bias-corrected method to calibrate the empirical likelihood ratios so that the estimator has asymptotically chi-squared distribution. Besides, with the missing response data in the partially linear varying-coefficient model (1), Wei (2012a) presented a profile least squares estimator for the parametric part based on the complete-case data.

Besides missing data, error-in-variables(EV) data, as another complex data can always be seen in real problems. It is well known that, if the measurement errors are ignored directly, the resulting estimators will not be unbiased. A great deal of researches on regression models with EV data have been studied. The simple specification of EV data is that the variables are measured with additive errors. Instead of observing certain covarites $\mathbf{X}$, we observe $\mathbf{W} = \mathbf{X} + \boldsymbol{\xi}$, where the measurement error $\boldsymbol{\xi}$ is independent of other variables. Taking model (1) as an example, under the situation of $\mathbf{X}$ is measured with additive error, You and Chen (2006) proposed a locally corrected profile least squares procedure to estimate the parameter and showed that the estimator is consistent and asymptotically normal. Zhang et al. (2011) and Wei (2012b) developed a restricted modified profile least squares estimator of the parameter under some additional linear restrictions. Hu et al. (2009) and Wang et al. (2011) constructed confidence regions for the unknown parameters with the empirical likelihood inference. On the other hand, when the nonparametric part $\mathbf{Z}$ is measured with additive error in model (1), Feng and Xue (2014) conducted a locally bias-corrected restricted profile least squares estimators of both parameter and nonparametric functions. Fan et al. (2016a) used some auxiliary information to construct empirical log-likelihood ratios and Fan et al. (2016b) extended the penalized empirical likelihood to the high-dimensional model. Fan et al. (2018) suggested a bias-correction penalized profile least squares variable selection method in high dimensional models. Moreover, when $\mathbf{X}$ is measured with additive errors as well as the response Y is missing in model (1), Wei and Mei (2012) applied the empirical likelihood method to construct confidence regions for parameters and Yang and Xia (2014) obtained restricted estimators under the linear constraint. However, the simultaneous existence of missing response and measurement error in the nonparametric part of model (1) has been seldom discussed. In addition, it should be noted that the assumption of additive measurement errors may be too simple in some applications. To analyze data from certain biomedical and health-related studies, one cannot directly observe some covariates and the response variable, but may obtain their distorted observations by certain functions of an observed confounding variable. Zhang et al. (2018) considered the nonlinear regression model under the assumption that both the response and predictors are unobservable and distorted by the multiplicative effects of some observable confounding variables. More interesting work for further study with model (1) will consider this situation.

In this paper, we study partially linear varying-coefficient models in which the response variable Y cannot be observed completely and the covariate $\mathbf{Z}$ cannot be observed accurately. Throughout this paper, we introduce an indicator variable δ such that $\delta = 1$ means that Y is observed and $\delta = 0$ indicates that Y is missing. We assume that data missing mechanism follows

$$\Pr(\delta = 1|Y, \mathbf{X},\mathbf{Z},U) = \Pr(\delta = 1|\mathbf{X},\mathbf{Z},U) = \pi(\mathbf{X},\mathbf{Z},U). \tag{2}$$

Meanwhile, the variable $\mathbf{Z}$ is measured with additive errors. That is

$$\mathbf{W} = \mathbf{Z} + \boldsymbol{\xi}, \tag{3}$$

where $\boldsymbol{\xi}$ is the measurement error and independent of $(Y, \mathbf{X}, \mathbf{Z}, U, \boldsymbol{\varepsilon}, \delta)$ and has mean zero and known covariance $\text{Cov}(\boldsymbol{\xi}) = \Sigma_{\boldsymbol{\xi}}$. Even if covariance $\Sigma_{\boldsymbol{\xi}}$ is unknown, a consistent and unbiased estimator can still be obtained by repeatedly observing $\mathbf{W}_i$ (see Liang et al. (2007) for details). If $\mathbf{Z}$ is observed exactly, then the probability of missingness is independent of missing responses and the resulting mechanism is called missing at random (MAR). However, considering model (1) under assumption (3), the covariate $\mathbf{Z}$ is observed with measurement error and therefore Y is not missing at random which has been pointed out by Wei and Mei (2012) and Liang et al. (2007).

The rest of this paper is organized as follows. In Sect. 2, the locally corrected profile linear least squares estimation procedure with complete-case data is proposed, and then the asymptotic properties of the estimators are proved under some assumptions. In Sect. 3, an imputation technique is used to improve the accuracy of the estimator and corresponding asymptotic results are obtained. Some simulation studies are conducted in Sect. 4 to assess the performances of the proposed two estimators. In Sect. 5, the methodologies are illustrated by a real data example. Sect. 6 is conclusion and the proofs of the main Theorems are left in the "Appendix".

2 Estimation method based on complete-case data

Firstly, we assume that measurement errors do not exist thus covariate $\mathbf{Z}$ can be observed exactly. Suppose that the observation data $\{Y_i; \delta_i, \mathbf{X}_i, \mathbf{Z}_i, U_i\}_{i=1}^n$ is generated from model (1) under assumption (2), then we have the following equation

$$\delta_i Y_i = \delta_i \mathbf{X}_i^T \boldsymbol{\beta} + \delta_i \mathbf{Z}_i^T \boldsymbol{\alpha}(U_i) + \delta_i \varepsilon_i, \quad i = 1, \ldots, n. \tag{4}$$

We assume that $\boldsymbol{\beta}$ is known, then the model (4) can be rewritten as the following varying coefficient regression model,

$$\delta_i Y_i - \delta_i \mathbf{X}_i^T \boldsymbol{\beta} = \delta_i \sum_{j=1}^{q} Z_{ij} \alpha_j(U_i) + \delta_i \varepsilon_i, \tag{5}$$

where Z_{ij} is the jth element of $\mathbf{Z}_i$ and $\alpha_j(.)$ is the jth function of $\boldsymbol{\alpha}(.)$, $j = 1, \ldots, q$. We can estimate the coefficient functions $\alpha_j(.)$, $j = 1 \cdots, q$ by the local linear fitting procedure. Specifically, for u in a small neighborhood of u_0, $\alpha_j(u)$ can be locally approximated by a linear function as following:

$$\alpha_j(u) \approx \alpha_j(u_0) + \alpha_j^{(1)}(u_0)(u - u_0) = a_j + b_j(u - u_0), \quad j = 1, \ldots, q,$$

where $\alpha_j^{(1)}(u) = \partial \alpha_j(u)/\partial u$ denotes the first order derivative of $\alpha_j(u)$. Then, the estimators of $\alpha_j(.)$ can be obtained by selecting $\{(a_j, b_j), j = 1, \ldots, q\}$ to minimize:

$$\sum_{i=1}^{n} \left\{ Y_i - \mathbf{X}_i^T \boldsymbol{\beta} - \sum_{j=1}^{q} [a_j + b_j(U_i - u)] Z_{ij} \right\}^2 K_{h_1}(U_i - u)\delta_i, \tag{6}$$

where $K_{h_1}(.) = K(./h_1)/h_1$, $K(.)$ is a kernel function and h_1 is the bandwidth. The solution to problem (6) is obtained by

$$\hat{\boldsymbol{\alpha}}(u; \boldsymbol{\beta}) = (\mathbf{I}_q, \mathbf{0}_q) \left[(\mathbf{D}_u^{\mathbf{Z}})^T \boldsymbol{\omega}_u^{\delta} \mathbf{D}_u^{\mathbf{Z}} \right]^{-1} (\mathbf{D}_u^{\mathbf{Z}})^T \boldsymbol{\omega}_u^{\delta} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{7}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$, $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)^T$, $\boldsymbol{\omega}_u^{\delta} = \mathrm{diag}(K_{h_1}(U_1 - u)\delta_1, \ldots, K_{h_1}(U_n - u)\delta_n)$, and

$$\mathbf{D}_u^{\mathbf{Z}} = \begin{pmatrix} \mathbf{Z}_1^T & h_1^{-1}(U_1 - u)\mathbf{Z}_1^T \\ \vdots & \vdots \\ \mathbf{Z}_n^T & h_1^{-1}(U_n - u)\mathbf{Z}_n^T \end{pmatrix}.$$

Now consider $\mathbf{Z}_i$'s are not observed due to measurement error and $\mathbf{W}_i$'s are the observable surrogate data. Thus, $\hat{\boldsymbol{\alpha}}(u; \boldsymbol{\beta})$ is not consistent and unbiased if $\mathbf{W}_i$ is replaced by $\mathbf{Z}_i$ directly in (7). Based on the idea of Feng and Xue (2014), a modified locally corrected linear estimators of $\boldsymbol{\alpha}(.)$ can be given by

$$\hat{\boldsymbol{\alpha}}(u; \boldsymbol{\beta}) = (\mathbf{I}_q, \mathbf{0}_q) \left[(\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u^{\delta} \mathbf{D}_u^{\mathbf{W}} - \boldsymbol{\Omega}_u^{\delta} \right]^{-1} (\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u^{\delta} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{8}$$

where $\mathbf{D}_u^{\mathbf{W}}$ has the same form as $\mathbf{D}_u^{\mathbf{Z}}$ except that $\mathbf{Z}_i$ is replaced by $\mathbf{W}_i$ and

$$\boldsymbol{\Omega}_u^{\delta} = \sum_{i=1}^{n} \boldsymbol{\Sigma}_{\xi} \otimes \begin{pmatrix} 1 & \frac{U_i - u}{h_1} \\ \frac{U_i - u}{h_1} & \left(\frac{U_i - u}{h_1}\right)^2 \end{pmatrix} K_{h_1}(U_i - u)\delta_i,$$

with $\otimes$ is the Kronecker product.

Taking u to be $U_1, \ldots, U_n$ in (8), we can get that $\hat{\boldsymbol{\alpha}}(U_i; \boldsymbol{\beta}) = \mathbf{Q}_i(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$, where $\mathbf{Q}_i = (\mathbf{I}_q, \mathbf{0}_q)[(\mathbf{D}_{U_i}^{\mathbf{W}})^T \boldsymbol{\omega}_{U_i}^{\delta} \mathbf{D}_{U_i}^{\mathbf{W}} - \boldsymbol{\Omega}_{U_i}^{\delta}]^{-1} (\mathbf{D}_{U_i}^{\mathbf{W}})^T \boldsymbol{\omega}_{U_i}^{\delta}$. For the convenience of expression, let

$$\mathbf{S}_c = \begin{pmatrix} (\mathbf{W}_1^T, 0_{1\times q})[(\mathbf{D}_{U_1}^{\mathbf{W}})^T \boldsymbol{\omega}_{U_1}^{\delta} \mathbf{D}_{U_1}^{\mathbf{W}} - \boldsymbol{\Omega}_{U_1}^{\delta}]^{-1} (\mathbf{D}_{U_1}^{\mathbf{W}})^T \boldsymbol{\omega}_{U_1}^{\delta} \\ \vdots \\ (\mathbf{W}_n^T, 0_{1\times q})[(\mathbf{D}_{U_n}^{\mathbf{W}})^T \boldsymbol{\omega}_{U_n}^{\delta} \mathbf{D}_{U_n}^{\mathbf{W}} - \boldsymbol{\Omega}_{U_n}^{\delta}]^{-1} (\mathbf{D}_{U_n}^{\mathbf{W}})^T \boldsymbol{\omega}_{U_n}^{\delta} \end{pmatrix},$$

and denote $\tilde{Y}_i = Y_i - \sum_{k=1}^n \mathbf{S}_{ik}^c Y_k$ and $\tilde{\mathbf{X}}_i = \mathbf{X}_i - \sum_{k=1}^n \mathbf{S}_{ik}^c \mathbf{X}_k$, where $\mathbf{S}_{ik}^c$ is the (i, k)th component of matrix $\mathbf{S}_c$.

Then, the locally corrected profile least squares estimator $\hat{\boldsymbol{\beta}}_c$ of $\boldsymbol{\beta}$ based on complete-case data is obtained by minimizing

$$\sum_{i=1}^n \delta_i \left[Y_i - \mathbf{X}_i^T \boldsymbol{\beta} - \mathbf{W}_i^T \hat{\boldsymbol{\alpha}}(U_i; \boldsymbol{\beta}) \right]^2 - \sum_{i=1}^n \delta_i \hat{\boldsymbol{\alpha}}^T (U_i; \boldsymbol{\beta}) \boldsymbol{\Sigma}_{\xi} \hat{\boldsymbol{\alpha}}(U_i; \boldsymbol{\beta}). \tag{9}$$

It is noted that the second term on the right hand side of (9) is included to avoid underestimating for $\boldsymbol{\beta}$ which is caused by measurement errors. By simple calculation, estimator $\hat{\boldsymbol{\beta}}_c$ can be obtained by

$$\hat{\boldsymbol{\beta}}_c = \left[\sum_{i=1}^n \delta_i \left(\tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i^T - \mathbf{X}^T \mathbf{Q}_i^T \boldsymbol{\Sigma}_{\xi} \mathbf{Q}_i \mathbf{X} \right) \right]^{-1} \left[\sum_{i=1}^n \delta_i \left(\tilde{\mathbf{X}}_i \tilde{Y}_i - \mathbf{X}^T \mathbf{Q}_i^T \boldsymbol{\Sigma}_{\xi} \mathbf{Q}_i \mathbf{Y} \right) \right]. \tag{10}$$

Then, substituting $\hat{\boldsymbol{\beta}}_c$ into $\hat{\boldsymbol{\alpha}}(u; \boldsymbol{\beta})$ of (8) gives the estimator $\hat{\boldsymbol{\alpha}}(u; \hat{\boldsymbol{\beta}}_c)$ of $\boldsymbol{\alpha}(u)$,that is

$$\hat{\boldsymbol{\alpha}}_c(u) = \hat{\boldsymbol{\alpha}}(u; \hat{\boldsymbol{\beta}}_c) = (\mathbf{I}_q, \mathbf{0}_q) \left[(\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u^{\delta} \mathbf{D}_u^{\mathbf{W}} - \boldsymbol{\Omega}_u^{\delta} \right]^{-1} (\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u^{\delta} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_c). \tag{11}$$

The asymptotic properties of $\hat{\boldsymbol{\beta}}_c$ and $\hat{\boldsymbol{\alpha}}_c(u)$ are given in the following Theorems.

Theorem 1 *Suppose that the Conditions in the Appendix C1–C5 hold. Then we have*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Omega}_1 \boldsymbol{\Sigma}_1^{-1}),$$

where "$\xrightarrow{d}$" denotes convergence in distribution,

$$\begin{aligned} \boldsymbol{\Sigma}_1 &= \mathrm{E}\{\delta_1 [\boldsymbol{X}_1 - \boldsymbol{\Phi}_c^T(U_1) \boldsymbol{\Gamma}_c^{-1}(U_1) \mathbf{Z}_1]^{\otimes 2}\}, \\ \boldsymbol{\Omega}_1 &= \mathrm{E}\{\delta_1 (\varepsilon_1 - \boldsymbol{\xi}_1^T \boldsymbol{\alpha}(U_1))^2 \boldsymbol{\Sigma}_1\} + \sigma^2 \mathrm{E}\{\delta_1 [\boldsymbol{\Phi}_c^T(U_1) \boldsymbol{\Gamma}_c^{-1}(U_1) \boldsymbol{\Sigma}_{\xi} \boldsymbol{\Gamma}_c^{-1}(U_1) \boldsymbol{\Phi}_c(U_1)]\} \\ &\quad + \mathrm{E}\{\delta_1 [\boldsymbol{\Phi}_c^T(U_1) \boldsymbol{\Gamma}_c^{-1}(U_1) (\boldsymbol{\xi}_1 \boldsymbol{\xi}_1^T - \boldsymbol{\Sigma}_{\xi}) \boldsymbol{\alpha}(U_1)]^{\otimes 2}\}, \\ &\qquad with \ \ \boldsymbol{\Gamma}_c(u) = \mathrm{E}(\delta_1 \mathbf{Z}_1 \mathbf{Z}_1^T | U = u) \ \ and \ \ \boldsymbol{\Phi}_c(u) = \mathrm{E}(\delta_1 \mathbf{Z}_1 \boldsymbol{X}_1^T | U = u). \end{aligned}$$

When we make statistical inference for $\boldsymbol{\beta}$ by Theorem 1, asymptotic variance of $\boldsymbol{\beta}$ is required to be estimated firstly. $\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Omega}_1 \boldsymbol{\Sigma}_1^{-1}$ is estimated by $\hat{\boldsymbol{\Sigma}}_1^{-1} \hat{\boldsymbol{\Omega}}_1 \hat{\boldsymbol{\Sigma}}_1^{-1}$ with plug-in method, where $\hat{\boldsymbol{\Sigma}}_1 = \frac{1}{n} \sum_{i=1}^n \delta_i \{\tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i^T - \mathbf{X}^T \mathbf{Q}_i^T \boldsymbol{\Sigma}_{\xi} \mathbf{Q}_i \mathbf{X}\}$, and $\hat{\boldsymbol{\Omega}}_1 =$

$\frac{1}{n}\sum_{i=1}^{n}\delta_i\left\{\tilde{\mathbf{X}}_i(\tilde{\mathbf{Y}}_i-\tilde{\mathbf{X}}_i^T\hat{\boldsymbol{\beta}}_c)-\mathbf{X}^T\mathbf{Q}_i^T\boldsymbol{\Sigma}_\xi\mathbf{Q}_i[\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}_c]\right\}^{\otimes 2}$.

Theorem 2 *Suppose that the Conditions C1–C5 in the Appendix hold and* $h_1 = cn^{-1/5}$, *where c is a constant. Then we have*

$$\max_{1\le j\le p}\sup_{u\in\Pi}|\hat{\alpha}_{cj}(u)-\alpha_j(u)| = O(n^{-2/5}+(\log n)^{1/2}), \quad a.s.$$

3 Estimation method based on imputation technique

It is noted that the estimator $\hat{\boldsymbol{\beta}}_c$ defined by (10) use complete-case data only and discard sample when Y_i is missing. This procedure may reduce the efficiency of the estimators of $\boldsymbol{\beta}$ which is caused without making full use of sample information.

When we are dealing with missing data, an imputation technique is prevalent which has been applied to various semi-parametric models, such examples can be found in Yang et al. (2011) and Xue and Xue (2011). The main idea of this method is to firstly impute a reasonable value for each missing data and then make statistical inference as if the data set is complete. Specifically, if covariate $\mathbf{Z}$ can be observed directly, based on the estimator $\hat{\boldsymbol{\beta}}_c$ and $\hat{\boldsymbol{\alpha}}_c(u;\hat{\boldsymbol{\beta}}_c)$, we have $(\hat{H}_i^0;\mathbf{X}_i,\mathbf{Z}_i,U_i)_{i=1}^n$, where

$$\hat{H}_i^0 = \delta_i Y_i + (1-\delta_i)[\mathbf{X}_i^T\hat{\boldsymbol{\beta}}_c + \mathbf{Z}_i^T\hat{\boldsymbol{\alpha}}_c(U_i;\hat{\boldsymbol{\beta}}_c)].$$

However, $\hat{H}_i^0$ can not be obtained since $\mathbf{Z}_i$ can not be observed in practice. Instead, $\hat{H}_i = \delta_i Y_i + (1-\delta_i)[\mathbf{X}_i^T\hat{\boldsymbol{\beta}}_c + \mathbf{W}_i^T\hat{\boldsymbol{\alpha}}_c(U_i;\hat{\boldsymbol{\beta}}_c)]$ is available. Based on data $(\hat{H}_i;\mathbf{X}_i,\mathbf{W}_i,U_i)_{i=1}^n$, the following partially linear varying-coefficient model with measurement errors in both covariate and response can be written as

$$\begin{cases}\hat{H}_i^0 = \mathbf{X}_i^T\boldsymbol{\beta} + \mathbf{Z}_i^T\boldsymbol{\alpha}(U_i) + e_i\\ \mathbf{W}_i = \mathbf{Z}_i + \boldsymbol{\xi}_i\\ \hat{H}_i = \hat{H}_i^0 + (1-\delta_i)\boldsymbol{\xi}_i^T\hat{\boldsymbol{\alpha}}_c(U_i;\hat{\boldsymbol{\beta}}_c)\end{cases} \tag{12}$$

where $e_i = \hat{H}_i^0 - Y_i + \varepsilon_i$ is the model error.

Then, the estimator $\hat{\boldsymbol{\beta}}_I$ of parameter $\boldsymbol{\beta}$ based on model (12) can be obtained by minimizing

$$\sum_{i=1}^n\left[\hat{H}_i - \mathbf{X}_i^T\boldsymbol{\beta} - \mathbf{W}_i^T\check{\boldsymbol{\alpha}}(U_i;\boldsymbol{\beta})\right]^2 - \sum_{i=1}^n\check{\boldsymbol{\alpha}}^T(U_i;\boldsymbol{\beta})\boldsymbol{\Sigma}_\xi\check{\boldsymbol{\alpha}}(U_i;\boldsymbol{\beta})$$
$$+\sum_{i=1}^n(1-\delta_i)\check{\boldsymbol{\alpha}}^T(U_i;\boldsymbol{\beta})\boldsymbol{\Sigma}_\xi\hat{\boldsymbol{\alpha}}(U_i;\hat{\boldsymbol{\beta}}_c), \tag{13}$$

where $\check{\boldsymbol{\alpha}}(u;\boldsymbol{\beta})$ has the same form as $\hat{\boldsymbol{\alpha}}(u;\boldsymbol{\beta})$ defined in (8), except that $\boldsymbol{\omega}_u^\delta$ and $\boldsymbol{\Omega}_u^\delta$ are replaced by $\boldsymbol{\omega}_u$ and $\boldsymbol{\Omega}_u$, respectively. That is

$$\check{\boldsymbol{\alpha}}(u;\boldsymbol{\beta}) = (\mathbf{I}_q, \mathbf{0}_q)\left[(\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{W}} - \boldsymbol{\Omega}_u\right]^{-1} (\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u (\hat{\mathbf{H}} - \mathbf{X}\boldsymbol{\beta}), \tag{14}$$

with $\boldsymbol{\omega}_u = \mathrm{diag}(K_{h_2}(U_1 - u), \ldots, K_{h_2}(U_n - u))$, $\hat{\mathbf{H}} = (\hat{H}_1, \hat{H}_2, \ldots, \hat{H}_n)^T$, and

$$\boldsymbol{\Omega}_u = \sum_{i=1}^{n} \boldsymbol{\Sigma}_\xi \otimes \begin{pmatrix} 1 & \frac{U_i-u}{h_2} \\ \frac{U_i-u}{h_2} & \left(\frac{U_i-u}{h_2}\right)^2 \end{pmatrix} K_{h_2}(U_i - u),$$

and $K_{h_2}(.) = K(./h_2)/h_2$ with a kernel function $K(.)$ and a bandwidth h_2.

Besides the second term in (13), the third term is added to correct the bias which is induced by $\mathbf{W}_i$ contained in $\hat{H}_i$. Similarity, denote $\mathbf{R}_i = (\mathbf{I}_q, \mathbf{0}_q)[(\mathbf{D}_{U_i}^{\mathbf{W}})^T \boldsymbol{\omega}_{U_i} \mathbf{D}_{U_i}^{\mathbf{W}} - \boldsymbol{\Omega}_{U_i}]^{-1} (\mathbf{D}_{U_i}^{\mathbf{W}})^T \boldsymbol{\omega}_{U_i}$, then $\check{\boldsymbol{\alpha}}(U_i;\boldsymbol{\beta}) = \mathbf{R}_i(\hat{\mathbf{H}} - \mathbf{X}\boldsymbol{\beta})$. Let

$$\mathbf{S}_I = \begin{pmatrix} (\mathbf{W}_1^T, 0_{1\times q})\left[(\mathbf{D}_{U_1}^{\mathbf{W}})^T \boldsymbol{\omega}_{U_1} \mathbf{D}_{U_1}^{\mathbf{W}} - \boldsymbol{\Omega}_{U_1}\right]^{-1} (\mathbf{D}_{U_1}^{\mathbf{W}})^T \boldsymbol{\omega}_{U_1} \\ \vdots \\ (\mathbf{W}_n^T, 0_{1\times q})\left[(\mathbf{D}_{U_n}^{\mathbf{W}})^T \boldsymbol{\omega}_{U_n} \mathbf{D}_{U_n}^{\mathbf{W}} - \boldsymbol{\Omega}_{U_n}\right]^{-1} (\mathbf{D}_{U_n}^{\mathbf{W}})^T \boldsymbol{\omega}_{U_n} \end{pmatrix}.$$

Denote $\bar{H}_i = \hat{H}_i - \sum_{k=1}^n \mathbf{S}_{ik}^I \hat{H}_k$ and $\bar{\mathbf{X}}_i = \mathbf{X}_i - \sum_{k=1}^n \mathbf{S}_{ik}^I \mathbf{X}_k$, where $\mathbf{S}_{ik}^I$ is the (i,k)th component of matrix $\mathbf{S}_I$.

By simple calculation, the estimator $\hat{\boldsymbol{\beta}}_I$ based on the imputation method is obtained by

$$\begin{aligned}\hat{\boldsymbol{\beta}}_I = &\left[\sum_{i=1}^{n} (\bar{\mathbf{X}}_i \bar{\mathbf{X}}_i^T - \mathbf{X}^T \mathbf{R}_i^T \boldsymbol{\Sigma}_\xi \mathbf{R}_i \mathbf{X})\right]^{-1} \\ &\left[\sum_{i=1}^{n} (\bar{\mathbf{X}}_i \bar{H}_i - \mathbf{X}^T \mathbf{R}_i^T \boldsymbol{\Sigma}_\xi \mathbf{R}_i \hat{\mathbf{H}}) + (1-\delta_i)\mathbf{X}^T \mathbf{R}_i^T \boldsymbol{\Sigma}_\xi \mathbf{Q}_i (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_c)\right].\end{aligned} \tag{15}$$

Then, the corresponding imputation estimator $\hat{\boldsymbol{\alpha}}_I(u)$ of $\boldsymbol{\alpha}(u)$ is defined as

$$\hat{\boldsymbol{\alpha}}_I(u) = \check{\boldsymbol{\alpha}}(u;\hat{\boldsymbol{\beta}}_I) = (\mathbf{I}_q, \mathbf{0}_q)\left[(\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u \mathbf{D}_u^{\mathbf{W}} - \boldsymbol{\Omega}_u\right]^{-1} (\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u (\hat{\mathbf{H}} - \mathbf{X}\hat{\boldsymbol{\beta}}_I). \tag{16}$$

The asymptotic normality of $\hat{\boldsymbol{\beta}}_I$ and the convergence of $\hat{\boldsymbol{\alpha}}_I(u)$ are given in the following Theorems.

Theorem 3 *Suppose that the Conditions C1–C5 in the Appendix hold. Then we have*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}^{-1}\boldsymbol{\Omega}_2\boldsymbol{\Sigma}^{-1}),$$

where

$$\begin{aligned}
\boldsymbol{\Sigma} &= \mathrm{E}\{[\boldsymbol{X}_1 - \boldsymbol{\Phi}^T(U_1)\boldsymbol{\Gamma}^{-1}(U_1)\mathbf{Z}_1]^{\otimes 2}\},\\
\boldsymbol{\Omega}_2 &= (\boldsymbol{\Sigma}_2 + \boldsymbol{\Sigma}_1)\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Omega}_1\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\Sigma}_2 + \boldsymbol{\Sigma}_1),\\
\boldsymbol{\Sigma}_2 &= \mathrm{E}\{(1-\delta_1)[\boldsymbol{X}_1 - \boldsymbol{\Phi}^T(U_1)\boldsymbol{\Gamma}^{-1}(U_1)\mathbf{Z}_1][\boldsymbol{X}_1 - \boldsymbol{\Phi}_c^T(U_1)\boldsymbol{\Gamma}_c^{-1}(U_1)\mathbf{Z}_1]^T\}\\
&\quad \text{with } \boldsymbol{\Gamma}(u) = \mathrm{E}(\mathbf{Z}_1\mathbf{Z}_1^T|U=u) \text{ and } \boldsymbol{\Phi}(u) = \mathrm{E}(\mathbf{Z}_1\boldsymbol{X}_1^T|U=u).
\end{aligned}$$

where $\boldsymbol{\Sigma}_1$ *and* $\boldsymbol{\Omega}_1$ *are defined in Theorem 1.*

Theorem 4 *Suppose that the Conditions C1–C5 in the Appendix hold and* $h_2 = cn^{-1/5}$*, where c is a constant. Then we have*

$$\max_{1\le j\le p}\sup_{u\in\Pi}|\hat{\alpha}_{Ij}(u) - \alpha_j(u)| = O(n^{-2/5} + (\mathrm{logn})^{1/2}), \quad \text{a.s.}$$

4 Simulation study

In this section, we conduct some simulations to assess the performances of the proposed estimators in finite samples. The data are generated from the following partially linear varying-coefficient measurement error model with missing responses

$$\begin{cases} Y_i = \mathbf{X}_i^T\boldsymbol{\beta} + Z_{1i}\alpha_1(U_i) + Z_{2i}\alpha_2(U_i) + \varepsilon_i,\\ W_{ji} = Z_{ji} + \xi_{ji}, \quad j=1,2, \quad i=1,\ldots,n, \end{cases}$$

where parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^T = (1, 1.5, 2, 1.5, 1)^T$, coefficient functions $\alpha_1(u) = \cos(2\pi u)$ and $\alpha_2(u) = \sin(2\pi u)$. The covariate variables X_1, X_2, X_3, X_4, X_5 are independently generated from $N(1, 1)$, Z_1, Z_2 are independently generated from $N(-1, 1)$ and U is independently drawn from a uniform distribution on [0,1]. In addition, the model error $\varepsilon \sim N(0, 1)$ and the measurement error $\boldsymbol{\xi} = (\xi_1, \xi_2)^T \sim N(\mathbf{0}, \Sigma_\xi)$ with $\Sigma_\xi = 0.2I_2$ and $\Sigma_\xi = 0.4I_2$, respectively. I_2 is identity matrix with order 2. We consider the following two missing schemes as

Case (i). $\Pr(\delta = 1|X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4, X_5 = x_5, Z_1 = z_1, Z_2 = z_2, U = u) = 0.8$ for all $x_1, x_2, x_3, x_4, x_5, z_1, z_2, u$.

Case (ii). $\Pr(\delta = 1|X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4, X_5 = x_5, Z_1 = z_1, Z_2 = z_2, U = u) = 0.8 + 0.6(|z_1| + |z_2| + |u - 0.5|)$ if $|z_1| + |z_2| + |u - 0.5| < 1$, and otherwise 0.8. In this case, the mean response rates is approximately 0.87.

Kernel function $K(t)$ is chosen as Epanechnikov kernel $K(t) = (3/4)(1 - t^2)$ if $|t| \le 1$, and 0 otherwise. In our simulation, we set the sample size n to be 100, 200 and 300. For each sample size, we generate 1000 random samples.

Some simulation results are reported in Tables 1, 2, 3, 4, 5, 6 and 7 to evaluate the performances of the proposed estimators $\hat{\boldsymbol{\beta}}_c$ and $\hat{\boldsymbol{\beta}}_I$. Firstly, to determine whether the choice of bandwidth has influences on the performance of the estimators, we choose three bandwidths with $h_1 = h_2 = h_{opt} = 2.34 * \mathrm{sd}(U) * n^{-1/5}$, where $\mathrm{sd}(U)$ is the standard deviation of the observations of $U_1, U_2, \ldots, U_n$. The average

Table 1 The average estimation errors of estimators for the parametric components with $\Sigma_\xi = 0.2I_2$ under missing case (i)

n	$\|\|\hat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}\|\|$			$\|\|\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}\|\|$		
	$0.5h_{opt}$	h_{opt}	$1.5h_{opt}$	$0.5h_{opt}$	h_{opt}	$1.5h_{opt}$
100	0.3441	0.2806	0.3130	0.3221	0.2789	0.2938
200	0.2071	0.1809	0.1881	0.1889	0.1806	0.1865
300	0.1557	0.1438	0.1494	0.1541	0.1432	0.1488

Table 2 Finite sample performance of estimator $\hat{\beta}_T$, $\hat{\beta}_c$, $\hat{\beta}_I$ and $\hat{\beta}_N$ for β_1

Σ_ξ	n	$\hat{\beta}_T$		$\hat{\beta}_c$		$\hat{\beta}_I$		$\hat{\beta}_N$	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD
case(i)									
$0.2I_2$	100	0.0027	0.1197	0.0035	0.1355	0.0032	0.1354	0.1109	0.2133
	200	0.0013	0.0797	0.0034	0.0878	0.0032	0.0876	0.1101	0.1363
	300	0.0005	0.0645	0.0019	0.0701	0.0014	0.0700	0.1082	0.1086
$0.4I_2$	100	0.0042	0.1145	0.0035	0.1393	0.0038	0.1376	0.2016	0.2575
	200	0.0027	0.0774	0.0023	0.0912	0.0021	0.0904	0.1998	0.1621
	300	0.0008	0.0648	0.0018	0.0725	0.0013	0.0721	0.1945	0.1310
case(ii)									
$0.2I_2$	100	0.0021	0.1129	0.0037	0.1274	0.0031	0.1264	0.1075	0.2049
	200	0.0009	0.0775	0.0018	0.0866	0.0018	0.0860	0.1013	0.1357
	300	0.0001	0.0613	0.0013	0.0675	0.0011	0.0673	0.1010	0.1065
$0.4I_2$	100	0.0026	0.1120	0.0057	0.1362	0.0053	0.1359	0.1910	0.2443
	200	0.0013	0.0749	0.0049	0.0907	0.0045	0.0899	0.1779	0.1680
	300	0.0006	0.0606	0.0013	0.0717	0.0011	0.0710	0.1659	0.1338

estimation errors $||\hat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}||$ and $||\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}||$ in L_2-norm are computed with three different bandwidths in Table 1. We can see that the choice of bandwidth shows a very slight impact on the estimators $\hat{\boldsymbol{\beta}}_c$ and $\hat{\boldsymbol{\beta}}_I$, especially when the sample size is large. Hence, we choose h_{opt} as the selected bandwidth for the later examples.

Secondly, in Tables 2, 3, 4, 5 and 6, "Bias" and "SD" denote the bias and the standard deviation of the 1000 estimators, respectively. For comparison, we not only report the proposed estimates $\hat{\beta}_c$ and $\hat{\beta}_I$, but also $\hat{\beta}_T$ and $\hat{\beta}_N$, which stand respectively for the true and naive estimates. The true estimate $\hat{\beta}_T$ is obtained via the standard profile least squares approach by using the complete data $(Y_i; X_i, Z_i, U_i)$, $i = 1, \ldots, n$. However, $\hat{\beta}_T$ is not applicable in practice since some observations of Y_i are not available due to missing and Z_i can not be obtained as a result of measurement errors. The naive estimate $\hat{\beta}_N$ is calculated by ignoring the measurement errors, not performing the bias correction as in equations (8) and (9), and applying the complete-case data only.

From Tables 2, 3, 4, 5 and 6, it is observed that the bias and SD of both estimators $\hat{\beta}_c$ and $\hat{\beta}_I$ are relatively small, which show that the proposed estimation procedures in

Table 3 Finite sample performance of estimator $\hat{\beta}_T$, $\hat{\beta}_c$, $\hat{\beta}_I$ and $\hat{\beta}_N$ for β_2

Σ_ξ	n	$\hat{\beta}_T$		$\hat{\beta}_c$		$\hat{\beta}_I$		$\hat{\beta}_N$	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD
case(i)									
$0.2I_2$	100	0.0039	0.1156	0.0083	0.1291	0.0078	0.1288	0.1297	0.2133
	200	0.0013	0.0813	0.0061	0.0888	0.0054	0.0881	0.1106	0.1420
	300	0.0005	0.0609	0.0006	0.0701	0.0003	0.0667	0.1071	0.1067
$0.4I_2$	100	0.0055	0.1131	0.0100	0.1437	0.0113	0.1383	0.2016	0.2575
	200	0.0026	0.0758	0.0062	0.0920	0.0069	0.0913	0.1924	0.1702
	300	0.0014	0.0649	0.0054	0.0765	0.0052	0.0751	0.1842	0.1347
case (ii)									
$0.2I_2$	100	0.0030	0.1180	0.0052	0.1306	0.0049	0.1298	0.1217	0.2070
	200	0.0012	0.0747	0.0017	0.0821	0.0015	0.0819	0.1091	0.1317
	300	0.0003	0.0644	0.0029	0.0696	0.0014	0.0693	0.1061	0.1073
$0.4I_2$	100	0.0087	0.1146	0.0135	0.1400	0.0057	0.1374	0.2154	0.2470
	200	0.0027	0.0744	0.0018	0.0888	0.0017	0.0881	0.1920	0.1593
	300	0.0020	0.0602	0.0021	0.0720	0.0013	0.0719	0.1900	0.1267

Table 4 Finite sample performance of estimator $\hat{\beta}_T$, $\hat{\beta}_c$, $\hat{\beta}_I$ and $\hat{\beta}_N$ for β_3

Σ_ξ	n	$\hat{\beta}_T$		$\hat{\beta}_c$		$\hat{\beta}_I$		$\hat{\beta}_N$	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD
case(i)									
$0.2I_2$	100	0.0024	0.1096	0.0073	0.1206	0.0070	0.1201	0.1234	0.2177
	200	0.0011	0.0806	0.0045	0.0900	0.0042	0.0898	0.1130	0.1417
	300	0.0008	0.0628	0.0025	0.0703	0.0021	0.0700	0.1091	0.1058
$0.4I_2$	100	0.0031	0.1188	0.0081	0.1487	0.0080	0.1440	0.2243	0.2527
	200	0.0028	0.0890	0.0064	0.0960	0.0060	0.0956	0.1982	0.1638
	300	0.0023	0.0637	0.0027	0.0741	0.0023	0.0741	0.1008	0.1373
case(ii)									
$0.2I_2$	100	0.0093	0.1202	0.0075	0.1325	0.0068	0.1313	0.1307	0.2095
	200	0.0025	0.0779	0.0044	0.0858	0.0041	0.0856	0.1133	0.1353
	300	0.0013	0.0595	0.0031	0.0648	0.0032	0.0648	0.1041	0.1036
$0.4I_2$	100	0.0029	0.1139	0.0167	0.1399	0.0164	0.1378	0.2087	0.2549
	200	0.0022	0.0868	0.0046	0.0894	0.0041	0.0884	0.1886	0.1660
	300	0.0016	0.0618	0.0028	0.0728	0.0027	0.0725	0.1865	0.1326

this paper can work well in finite samples. The estimators $\hat{\beta}_c$ and $\hat{\beta}_I$ are comparable to $\hat{\beta}_T$, though it is impossible to obtain in practice. The bias and the SD of $\hat{\beta}_N$ is much larger than other three estimators, which indicate that the measurement error should not be ignored directly. It is noted that the estimator $\hat{\beta}_I$ based on the imputation technique outperform the complete-case estimator $\hat{\beta}_c$ in terms that it gives smaller SD

Table 5 Finite sample performance of estimator $\hat{\beta}_T$, $\hat{\beta}_c$, $\hat{\beta}_I$ and $\hat{\beta}_N$ for β_4

Σ_ξ	n	$\hat{\beta}_T$		$\hat{\beta}_c$		$\hat{\beta}_I$		$\hat{\beta}_N$	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD
case(i)									
$0.2I_2$	100	0.0057	0.1128	0.0093	0.1279	0.0089	0.1270	0.1217	0.2063
	200	0.0039	0.0791	0.0073	0.0863	0.0071	0.0858	0.1070	0.1366
	300	0.0033	0.0611	0.0057	0.0670	0.0055	0.0698	0.1008	0.1065
$0.4I_2$	100	0.0077	0.1178	0.0100	0.1488	0.0096	0.1461	0.2052	0.2572
	200	0.0042	0.0808	0.0083	0.0936	0.0080	0.0932	0.1964	0.1692
	300	0.0023	0.0638	0.0029	0.0746	0.0027	0.0742	0.1843	0.1324
case(ii)									
$0.2I_2$	100	0.0042	0.1110	0.0093	0.1243	0.0088	0.1232	0.1193	0.2036
	200	0.0031	0.0758	0.0059	0.0833	0.0055	0.0830	0.1024	0.1322
	300	0.0021	0.0617	0.0025	0.0689	0.0023	0.0687	0.1007	0.1036
$0.4I_2$	100	0.0098	0.1112	0.0167	0.1376	0.0164	0.1345	0.2043	0.2441
	200	0.0052	0.0763	0.0046	0.0896	0.0037	0.0894	0.1923	0.1642
	300	0.0007	0.0618	0.0064	0.0749	0.0062	0.0746	0.1826	0.1374

Table 6 Finite sample performance of estimator $\hat{\beta}_T$, $\hat{\beta}_c$, $\hat{\beta}_I$ and $\hat{\beta}_N$ for β_5

Σ_ξ	n	$\hat{\beta}_T$		$\hat{\beta}_c$		$\hat{\beta}_I$		$\hat{\beta}_N$	
		Bias	SD	Bias	SD	Bias	SD	Bias	SD
case(i)									
$0.2I_2$	100	0.0051	0.1216	0.0052	0.1356	0.0049	0.1349	0.1124	0.2135
	200	0.0020	0.0797	0.0030	0.0877	0.0032	0.0876	0.1070	0.1336
	300	0.0016	0.0656	0.0031	0.0731	0.0031	0.0727	0.1008	0.1107
$0.4I_2$	100	0.0059	0.1241	0.0165	0.1400	0.0156	0.1380	0.1982	0.2587
	200	0.0010	0.0795	0.0037	0.0906	0.0031	0.0899	0.1908	0.1713
	300	0.0017	0.0653	0.0034	0.0758	0.0031	0.0756	0.1856	0.1334
case(ii)									
$0.2I_2$	100	0.0042	0.1196	0.0082	0.1331	0.0080	0.1326	0.1129	0.2162
	200	0.0023	0.0765	0.0051	0.0842	0.0047	0.0841	0.1073	0.1302
	300	0.0002	0.0594	0.0017	0.0670	0.0017	0.0670	0.1056	0.1049
$0.4I_2$	100	0.0020	0.1096	0.0067	0.1373	0.0065	0.1352	0.2003	0.2347
	200	0.0023	0.0760	0.0031	0.0902	0.0035	0.0894	0.1886	0.1620
	300	0.0007	0.0636	0.0009	0.0724	0.0008	0.0722	0.1814	0.1322

in most cases. This is due to the fact that $\hat{\beta}_I$ can make full use of the sample information. Furthermore, the values of SD in the missing data case (i) are usually greater than those in case (ii). The reason for this is that the number of observed responses generated from the missing data case (i) is less than that from case (ii). In addition, the larger

Table 7 The average MSE of estimator $\hat{\beta}_T$, $\hat{\beta}_c$, $\hat{\beta}_I$ and $\hat{\beta}_N$

Σ_ξ	n	$\varepsilon \sim N(0, 0.5)$				$\varepsilon \sim N(0, 1)$			
		$\hat{\beta}_T$	$\hat{\beta}_c$	$\hat{\beta}_I$	$\hat{\beta}_N$	$\hat{\beta}_T$	$\hat{\beta}_c$	$\hat{\beta}_I$	$\hat{\beta}_N$
case(i)									
$0.2I_2$	100	0.0072	0.0113	0.0103	0.0505	0.0133	0.0168	0.0167	0.0579
	200	0.0030	0.0043	0.0043	0.0272	0.0059	0.0072	0.0072	0.0299
	300	0.0019	0.0027	0.0027	0.0214	0.0039	0.0047	0.0046	0.0232
$0.4I_2$	100	0.0070	0.0132	0.0128	0.1015	0.0134	0.0215	0.0205	0.1103
	200	0.0032	0.0056	0.0055	0.0623	0.0062	0.0088	0.0088	0.0659
	300	0.0020	0.0035	0.0034	0.0510	0.0038	0.0054	0.0054	0.0538
case(ii)									
$0.2I_2$	100	0.0064	0.0094	0.0093	0.0502	0.0135	0.0167	0.0166	0.0565
	200	0.0031	0.0043	0.0043	0.0268	0.0058	0.0072	0.0071	0.0298
	300	0.0018	0.0026	0.0025	0.0205	0.0039	0.0047	0.0046	0.0222
$0.4I_2$	100	0.0065	0.0134	0.0123	0.0967	0.0136	0.0205	0.0200	0.1035
	200	0.0029	0.0051	0.0051	0.0604	0.0059	0.0085	0.0085	0.0621
	300	0.0019	0.0034	0.0033	0.0496	0.0039	0.0053	0.0052	0.0526

variance of measurement error Σ_ξ yields larger SD. It can also be observed that all methods perform better with smaller bias and SD as the sample size increases.

To illustrate the effect of the variance of model error ε on the proposed estimation methods, we compare the average mean square error(MSE) of vector β in Table 7. The smaller variance of model error ε, the smaller MSE. It can be seen that all the proposed estimation procedures perform better with the small variance of model error.

In addition, we report the performances of the proposed estimation procedures for the nonparametric function. We plot the estimated curve of the nonparametric function when the measurement error covariance is $\Sigma_\xi = 0.2I_2$ and the missing scheme is case (i) with sample size 200 in Fig. 1. We also evaluate the performance of the estimator $\boldsymbol{\alpha}(.)$ by using the square root of mean-squared errors (RMSE) which is defined as

$$\text{RMSE} = \left\{ \frac{1}{N} \sum_{k=1}^{N} ||\hat{\boldsymbol{\alpha}}(U_k) - \boldsymbol{\alpha}(U_k)||^2 \right\}^{1/2}$$

where $U_k, k = 1, \ldots, N$ are the grid points at which the function is evaluated. In our simulation, we set $N = 200$ and U_k is equally taken on interval (0,1). Figure 2 shows the box-plots for 1000 RMSE values for the nonparametric functions $\boldsymbol{\alpha}(.)$ with different methods.

From Fig. 1, we can see that the estimators $\hat{\alpha}_c(.)$ and $\hat{\alpha}_I(.)$ are almost same, and they all approximate the real curve. It shows that both the proposed methods can perform well in terms of nonparametric functions. From Fig. 2, it is observed that the RMSE

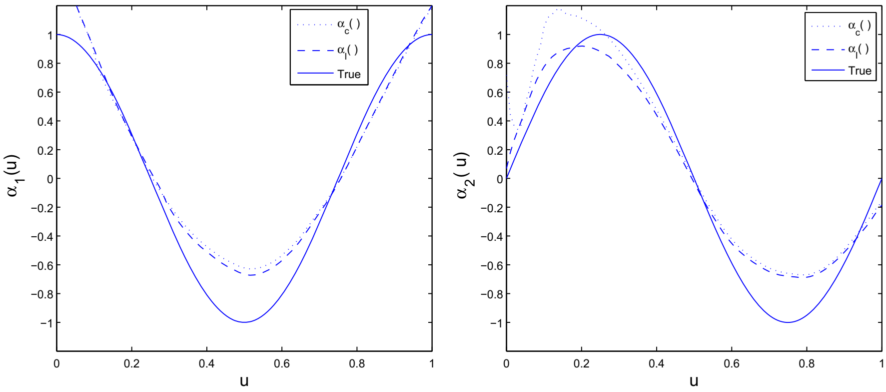

Fig. 1 The plot of the nonparametric estimator. The dotted, the dashed and the solid lines respectively denote $\hat{\alpha}_C(.)$, $\hat{\alpha}_I(.)$ and the true curve $\alpha(.)$

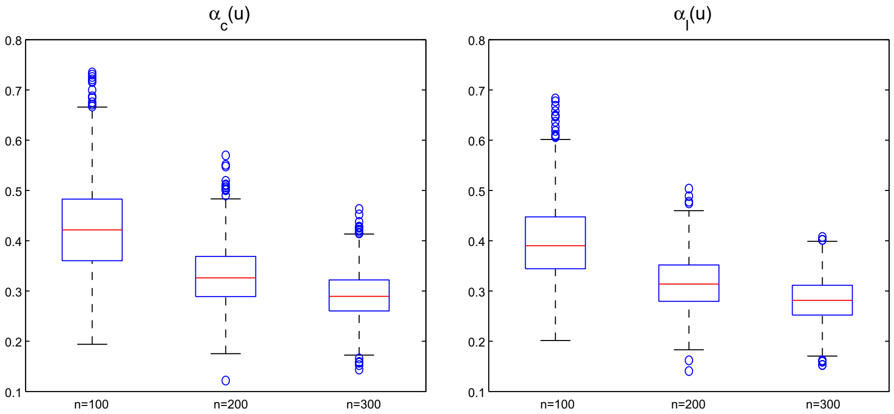

Fig. 2 The boxplots of the 1000 RMSE values for the nonparametric functions based on the complete-case data (left panel) and imputation technique (right panel)

values, obtained by the complete-case data and imputation technique, both decrease as the sample size increases. In addition, $\hat{\alpha}_I(.)$ performs better than $\hat{\alpha}_c(.)$ since it has smaller RMSE values.

In this simulation, we assume that the dimension p of the parameter β is fixed. In a general setup, the dimension p can grow with the sample size n, and thus model (1) extends to a high-dimensional partially linear varyingcoefficient model. As there would be some spurious covariates in the parametric component, some penalized profile least squares estimation procedures should be developed. The assumption that there are simultaneous missing response observations and additive errors in the nonparametric component in high-dimensional partially linear varying-coefficient model (1), would be more practical, but more challenging, which is left for the future research.

Table 8 The estimates of β_1 and β_2

Missing rate (%)	$\hat{\beta}_1$		$\hat{\beta}_2$	
	$\hat{\beta}_c$	$\hat{\beta}_I$	$\hat{\beta}_c$	$\hat{\beta}_I$
10	0.0648	0.0634	− 0.1423	− 0.1443
15	0.0750	0.0728	− 0.1405	− 0.1412
20	0.0989	0.0971	− 0.1319	− 0.1337

5 A real example

In this section, we apply our proposed estimation procedures to the Boston housing data set, which has been analyzed by several researches, such as Fan and Huang (2005), Wang and Xue (2011) and Li and Mei (2013) via different regression models. The median value of houses and several associated variables which might explain the variation of housing values are our main interest. In this study, we take the median value of owner-occupied homes in \$1000s(MEDV) as the response variable Y, per capita crime rate by town(CRIM), nitric oxide concentration parts per 10 million(NOX), average number of rooms per dwelling(RM), full-value property tax per \$10,000(TAX), proportion of owner-occupied units built prior to 1940(AGE) and pupil-teacher ratio by town school district(PTRATIO) as covariates, denoted by Z_2, Z_3, Z_4, Z_5, X_1 and X_2,respectively. We take $Z_1 = 1$ as the intercept term and $U = \sqrt{\text{LSTAT}}$ as the index variable, where LSTAT means lower status of the population. We employ the following partially linear varying-coefficient model

$$Y = X_1\beta_1 + X_2\beta_2 + \sum_{i=1}^{5} Z_i\alpha_i(U) + \varepsilon. \tag{17}$$

to fit the given data.

Before building the model, the response and covariate variables should be standardized for mean zero and unit sample standard deviation. In addition, the index variable U is transformed so that its marginal distribution is U[0,1]. To illustrate our method to this data set, as mentioned in Feng and Xue (2014), we consider the situation that covariate Z_5 has measurement error and can not be observed directly. Instead of Z_5, W_5 can be observed and has the following form

$$W_5 = Z_5 + U_5, \tag{18}$$

where $U_5 \sim N(0, 0.3^2)$. Firstly, we fit the data set by models (17) and (18) without response missing. The estimator of $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}_0 = (0.0435, -0.1446)^T$ is obtained with all observation data. Secondly, we remove 10%,15% and 20% of the response Y values at random. Since δ is randomly generated, we estimate $\boldsymbol{\beta}$ from 100 simulation runs and the average results can be found in Table 8.

From Table 8, we can obtain that two estimates of the parameter with complete-case data and imputation technique are almost same, and close to the case with no response

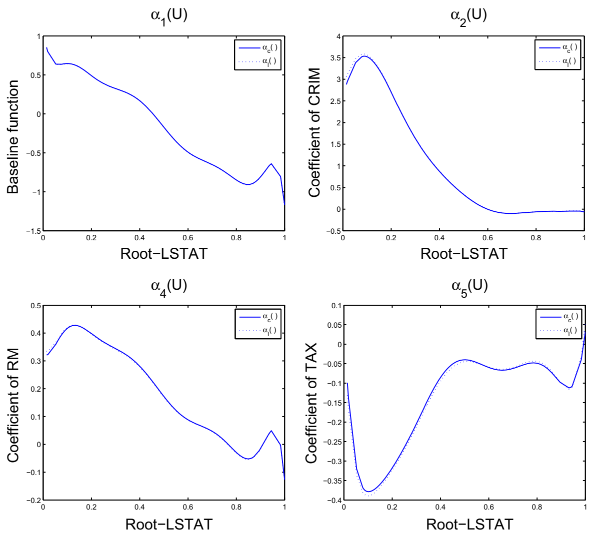

Fig. 3 The estimated coefficient functions, where the solid line and the dotted line represent the estimated coefficient functions $\hat{\alpha}_c(.)$ and $\hat{\alpha}_I(.)$, respectively

missing. In addition, the missing rate is smaller, the estimator value is closer to the case of no missing.

The estimated coefficient functions when the missing rata is 20% are depicted in Fig. 3. From Fig. 3, we can observe that the shapes of the $\hat{\alpha}_c(.)$ and $\hat{\alpha}_I(.)$ are very similar in five different coefficient functions.

6 Conclusions

In this paper, we study the partially linear varying-coefficient model when the nonparametric component is measured with additive error and the response variable is missing simultaneously. Firstly, we propose a locally corrected profile linear least squares estimation procedure based on the complete-case data only. Furthermore, a semiparametric imputation technique is applied to construct another estimator for improving the accuracy of the estimator. We establish the asymptotic normality property of the proposed two estimators of the parameters. As well, we show that the estimator of nonparametric component converge at an optimal rate. Theoretically, the estimator based on the imputation technique has advantages compared to the complete-case data method because it makes full use of information of the observation data. This conclusion is confirmed by the simulation studies and a real example.

However, we only consider the case in which there are a fixed number of predictors in this study. Currently highdimensional data analysis has attracted extensive attention. One important aspect of a regression model for highdimensional data is that the number of covariates is diverging. There has been some remarkable results on variable selection and parameter estimation in partially linear varying coefficient errors-invariables model with no missing data (Fan et al. 2016b, 2018). The simultaneous existence of missing response observations as well as measurement errors in the covariates would be extremely challenge in the highdimensional data modeling. We may apply some penalization methods for variable selection. Specifically, a penalty function could be added to Eqs. (9) or (13). The penalized estimator of β based on complete-case data can be obtained by minimizing the following bias-corrected penalized least square function

$$\frac{1}{2}\sum_{i=1}^{n}\delta_i\left[Y_i-\mathbf{X}_i^T\boldsymbol{\beta}-\mathbf{W}_i^T\hat{\boldsymbol{\alpha}}(U_i;\boldsymbol{\beta})\right]^2-\frac{1}{2}\sum_{i=1}^{n}\delta_i\hat{\boldsymbol{\alpha}}^T(U_i;\boldsymbol{\beta})\boldsymbol{\Sigma}_\xi\hat{\boldsymbol{\alpha}}(U_i;\boldsymbol{\beta})+n\sum_{j=1}^{p}p_\lambda(|\beta_j|). \tag{19}$$

where $p_\lambda(.)$ is a pre-specified penalty function, such as the SCAD penalty. The tuning parameter λ can be selected by some data-driven criteria, such as BIC, AIC, CV. Since the SCAD penalty function is irregular at the origin, the commonly used gradient method is not applicable. To solve this difficulty, an iterative algorithm is proposed by Fan and Li (2001). The penalty function is locally approximated by a quadratic function and then Newton–Raphson algorithm can be used to minimize problem (19). This method can significant reduce the computational burden and should be studied in the future work.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Nos. 11601419, 11801438).

Appendix: Proofs of the main results

We begin with the following assumption conditions required to derive the main results. These conditions are quite mild and can be easily satisfied.

C1: The random variable u has a bounded support Π. Its probability density function $f(.)$ is Lipschitz continuous and bounded away from 0 on its support.

C2: The $q\times q$ matrix $\mathrm{E}(\mathbf{ZZ}^T|U)$ and $\mathrm{E}(\delta\mathbf{ZZ}^T|U)$ are nonsingular for each $U\in\Pi$. The matrix $\mathrm{E}(\mathbf{ZZ}^T|U)$, $\mathrm{E}(\mathbf{ZZ}^T|U)^{-1}$, $\mathrm{E}(\delta\mathbf{ZZ}^T|U)$, $\mathrm{E}(\delta\mathbf{ZZ}^T|U)^{-1}$, $\mathrm{E}(\mathbf{ZX}^T|U)$ and $\mathrm{E}(\delta\mathbf{ZX}^T|U)$ are all Lipschitz continuous.

C3: There exists an $s>0$ such that $\mathrm{E}||\mathbf{X}||^{2s}<\infty$,$\mathrm{E}||\mathbf{Z}||^{2s}<\infty$ and for some $k<2-s^{-1}$ such that $n^{2k-1}h\longrightarrow\infty$.

C4: $\alpha_j(u)$, $j=1,\ldots,q$ have continuous second derivative for $u\in\Pi$.

C5: The Kernel $K(.)$ is a symmetric probability density function with compact support and the bandwidth h satisfies $nh^8\longrightarrow 0$ and $nh^2/(\log n)^2\longrightarrow\infty$ when $n\longrightarrow\infty$.

In order to prove the main results, we first give several Lemmas. The following notations will be used in the proof of the Lemmas and Theorems. Let $c_n=(\log n/nh)^{1/2}$, $\mu_i=\int_0^\infty t^iK(t)\mathrm{d}t$, $\mathbf{M}=[\mathbf{Z}_1^T\boldsymbol{\alpha}(U_1),\ldots,\mathbf{Z}_n^T\boldsymbol{\alpha}(U_n)]^T$, $\mathbf{M}^{\mathbf{W}}=[\mathbf{W}_1^T\boldsymbol{\alpha}(U_1),\ldots,\mathbf{W}_n^T\boldsymbol{\alpha}(U_n)]^T$, $\tilde{\varepsilon}_i=\varepsilon_i-\sum_{k=1}^n\mathbf{S}_{ik}^c\varepsilon_k$ and $\tilde{\mathbf{Z}}_i=\mathbf{Z}_i-\sum_{k=1}^n\mathbf{S}_{ik}^c\mathbf{Z}_k$.

Lemma 1 *Suppose that conditions C1–C5 hold. Then the followings hold uniformly*

$$(\boldsymbol{D}_u^{\mathbf{Z}})^T \boldsymbol{\omega}_u \boldsymbol{D}_u^{\mathbf{Z}} - \boldsymbol{\Omega}_u = nf(u)\boldsymbol{\Gamma}(u) \otimes \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} [1 + O_p(c_n)]. \tag{20}$$

$$(\boldsymbol{D}_u^{\mathbf{Z}})^T \boldsymbol{\omega}_u \boldsymbol{X} = nf(u)\boldsymbol{\Phi}(u) \otimes (1, \mu_1)^T [1 + O_p(c_n)]. \tag{21}$$

$$(\boldsymbol{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u^{\delta} \boldsymbol{D}_u^{\mathbf{W}} - \boldsymbol{\Omega}_u^{\delta} = nf(u)\boldsymbol{\Gamma}_c(u) \otimes \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} [1 + O_p(c_n)]. \tag{22}$$

$$(\boldsymbol{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u^{\delta} \boldsymbol{X} = nf(u)\boldsymbol{\Phi}_c(u) \otimes (1, \mu_1)^T [1 + O_p(c_n)]. \tag{23}$$

Proof Equations (20) and (21) are given in Lemma 2 in Feng and Xue (2014). Similarly, Eqs. (22) and (23) can also be obtained.

Lemma 2 *Suppose that conditions C1–C5 hold. Then*

$$\frac{1}{n}\sum_{i=1}^{n} \delta_i (\tilde{\boldsymbol{X}}_i \tilde{\boldsymbol{X}}_i^T - \boldsymbol{X}^T \boldsymbol{Q}_i^T \boldsymbol{\Sigma}_\xi \boldsymbol{Q}_i \boldsymbol{X}) \longrightarrow \boldsymbol{\Sigma}_{\mathbf{1}}, \;\; a.s,$$

$$\frac{1}{n}\sum_{i=1}^{n} (\bar{\boldsymbol{X}}_i \bar{\boldsymbol{X}}_i^T - \boldsymbol{X}^T \boldsymbol{R}_i^T \boldsymbol{\Sigma}_\xi \boldsymbol{R}_i \boldsymbol{X}) \longrightarrow \boldsymbol{\Sigma}, \;\; a.s,$$

$$\frac{1}{n}\sum_{i=1}^{n} (1 - \delta_i)(\bar{\boldsymbol{X}}_i \tilde{\boldsymbol{X}}_i^T - \boldsymbol{X}^T \boldsymbol{R}_i^T \boldsymbol{\Sigma}_\xi \boldsymbol{Q}_i \boldsymbol{X}) \longrightarrow \boldsymbol{\Sigma}_2, \;\; a.s,$$

where $\boldsymbol{\Sigma}_{\mathbf{1}}$ *is defined in Theorem 1,* $\boldsymbol{\Sigma}$ *and* $\boldsymbol{\Sigma}_2$ *are defined in Theorem 3.*

Proof The proof of this Lemma is similar to that of Lemma 7.2 in Fan and Huang (2005). Hence, the details are omitted.

Proof of Theorem 1 Let

$$B_n = \frac{1}{n}\sum_{i=1}^{n} \delta_i (\tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_i^T - \mathbf{X}^T \mathbf{Q}_i^T \boldsymbol{\Sigma}_\xi \mathbf{Q}_i \mathbf{X}),$$

and

$$A_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \delta_i \left[\tilde{\mathbf{X}}_i (\tilde{Y}_i - \tilde{\mathbf{X}}_i^T \boldsymbol{\beta}) - \mathbf{X}^T \mathbf{Q}_i^T \boldsymbol{\Sigma}_\xi \mathbf{Q}_i (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right].$$

Then,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}) = B_n^{-1} A_n. \tag{24}$$

For A_n, by simple calculation and similar proof of Lemma 4 in Feng and Xue (2014), we have

$$\begin{aligned}
A_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \left[\tilde{\mathbf{X}}_i (\tilde{\mathbf{Z}}_i^T \boldsymbol{\alpha}(U_i) + \tilde{\varepsilon}_i) - \mathbf{X}^T \mathbf{Q}_i^T \boldsymbol{\Sigma}_\xi \mathbf{Q}_i (\mathbf{M} + \boldsymbol{\varepsilon}) \right] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \Big\{ [\mathbf{X}_i - \boldsymbol{\Phi}_c^T (U_i) \boldsymbol{\Gamma}_c^{-1} (U_i) \mathbf{Z}_i][\varepsilon_i - \boldsymbol{\xi}_i^T \boldsymbol{\alpha}(U_i)] \\
&\quad - \boldsymbol{\Phi}_c^T (U_i) \boldsymbol{\Gamma}_c^{-1} (U_i) \boldsymbol{\xi}_i \varepsilon_i + \boldsymbol{\Phi}_c^T (U_i) \boldsymbol{\Gamma}_c^{-1} (U_i) (\boldsymbol{\xi}_i \boldsymbol{\xi}_i^T - \boldsymbol{\Sigma}_\xi) \boldsymbol{\alpha}(U_i) \Big\} + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{G}_i + o_p(1). \qquad (25)
\end{aligned}$$

It is easy to see that $\mathbf{G}_i$ is independent and identical distributed with mean zero and $\mathrm{Cov}(\mathbf{G}_i) = \boldsymbol{\Omega}_1$.

Thus, by the Slutsky theorem, Lemma 2 and the central limit theorem, we complete the Theorem.

Proof of Theorem 2 By the definition of $\hat{\boldsymbol{\alpha}}_c(u)$, we can obtain that

$$\begin{aligned}
\hat{\boldsymbol{\alpha}}_c(u) = &(\mathbf{I}_q, \mathbf{0}_q)[(\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u^\delta \mathbf{D}_u^{\mathbf{W}} - \boldsymbol{\Omega}_u^\delta]^{-1} (\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u^\delta \mathbf{M} \\
&+ (\mathbf{I}_q, \mathbf{0}_q)[(\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u^\delta \mathbf{D}_u^{\mathbf{W}} - \boldsymbol{\Omega}_u^\delta]^{-1} (\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u^\delta \boldsymbol{\varepsilon} \\
&+ (\mathbf{I}_q, \mathbf{0}_q)[(\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u^\delta \mathbf{D}_u^{\mathbf{W}} - \boldsymbol{\Omega}_u^\delta]^{-1} (\mathbf{D}_u^{\mathbf{W}})^T \boldsymbol{\omega}_u^\delta \mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_c).
\end{aligned}$$

By Theorem 1, similar to the proof of Theorem 3.1 in Xia and Li (1999), it is easy to show that

$$\max_{1 \le j \le p} \sup_{u \in \Pi} |\hat{\alpha}_{cj}(u) - \alpha_j(u)| = O\{h_1^2 + (\log n / n h_1)^{1/2}\}, \quad a.s.$$

Let $h_1 = cn^{-1/5}$, where c is a constant. Then it yields that

$$\max_{1 \le j \le p} \sup_{u \in \Pi} |\hat{\alpha}_{cj}(u) - \alpha_j(u)| = O(n^{-2/5} + (\log n)^{1/2}), \quad a.s.$$

Proof of Theorem 3 Similar to Theorem 1, it can be shown that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}) = D_n^{-1} E_n, \qquad (26)$$

where

$$D_n = \frac{1}{n} \sum_{i=1}^{n} (\bar{\mathbf{X}}_i \bar{\mathbf{X}}_i^T - \mathbf{X}^T \mathbf{R}_i^T \boldsymbol{\Sigma}_\xi \mathbf{R}_i \mathbf{X}),$$

and

$$E_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \bar{\mathbf{X}}_i (\bar{H}_i - \bar{\mathbf{X}}_i^T \boldsymbol{\beta}) - \mathbf{X}^T \mathbf{Q}_i^T \boldsymbol{\Sigma}_\xi \mathbf{Q}_i \big(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ + (1-\delta_i)\mathbf{X}^T \mathbf{R}_i^T \boldsymbol{\Sigma}_\xi \mathbf{Q}_i \big(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_c) + o_p(1).$$

For convenience, we denote $[\mathbf{S}_c(\mathbf{A})]_i$ and $[\mathbf{S}_I(\mathbf{A})]_i$ to respectively be the ith row of product of $\mathbf{S}_c\mathbf{A}$ and $\mathbf{S}_I\mathbf{A}$ for a given matrix $\mathbf{A}$.

By simple calculation, it is obtained that

$$\begin{aligned} E_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \bar{\mathbf{X}}_i \delta_i (\varepsilon_i - \boldsymbol{\xi}_i^T \boldsymbol{\alpha}(U_i)) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1-\delta_i) \bar{\mathbf{X}}_i \tilde{\mathbf{X}}_i^T (\hat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1-\delta_i) \bar{\mathbf{X}}_i [\mathbf{S}_c(\boldsymbol{\varepsilon} - \boldsymbol{\xi}^T \boldsymbol{\alpha}(u))]_i \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \bar{\mathbf{X}}_i \{\mathbf{W}_i^T \boldsymbol{\alpha}(U_i) + [\mathbf{S}_I(\mathbf{X})]_i^T \boldsymbol{\beta} - [\mathbf{S}_I(\hat{\mathbf{H}})]_i\} \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1-\delta_i) \bar{\mathbf{X}}_i \{[\mathbf{S}_c(\mathbf{M}^{\mathbf{W}})]_i - \mathbf{W}_i^T \boldsymbol{\alpha}(U_i)\} \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \mathbf{X}^T \mathbf{Q}_i^T \boldsymbol{\Sigma}_\xi \mathbf{Q}_i (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1-\delta_i) \mathbf{X}^T \mathbf{R}_i^T \boldsymbol{\Sigma}_\xi \mathbf{Q}_i \mathbf{X} (\hat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}) \\ &= \sum_{i=1}^{7} I_i. \end{aligned}$$

By Lemma 1, we have

$$I_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbf{X}_i - \boldsymbol{\Phi}(U_i)\boldsymbol{\Gamma}^{-1}(U_i)\mathbf{W}_i)\delta_i(\varepsilon_i - \boldsymbol{\xi}_i^T \boldsymbol{\alpha}(U_i)) + o_p(1). \tag{27}$$

In view of Theorem 1 and the law of large numbers, it follows that

$$I_2 = \frac{1}{n} \sum_{i=1}^{n} (1-\delta_i) \bar{\mathbf{X}}_i \tilde{\mathbf{X}}_i^T \sqrt{n} (\hat{\boldsymbol{\beta}}_c - \boldsymbol{\beta})$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)(\bar{\mathbf{X}}_i\tilde{\mathbf{X}}_i^T - \mathbf{X}^T\mathbf{R}_i^T\boldsymbol{\Sigma}_\xi\mathbf{Q}_i\mathbf{X})\sqrt{n}(\hat{\boldsymbol{\beta}}_c-\boldsymbol{\beta}) \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathbf{X}^T\mathbf{R}_i^T\boldsymbol{\Sigma}_\xi\mathbf{Q}_i\mathbf{X}(\hat{\boldsymbol{\beta}}_c-\boldsymbol{\beta}) + o_p(1) \\
&= \boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{G}_i + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathbf{X}^T\mathbf{R}_i^T\boldsymbol{\Sigma}_\xi\mathbf{Q}_i\mathbf{X}(\hat{\boldsymbol{\beta}}_c-\boldsymbol{\beta}) + o_p(1),
\end{aligned} \tag{28}
$$

where $\mathbf{G}_i$ is defined in Theorem 1.

I_3 can be written as

$$
\begin{aligned}
I_3 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathbf{X}_i[\mathbf{S}_c(\boldsymbol{\varepsilon}-\boldsymbol{\xi}^T\boldsymbol{\alpha}(u))]_i - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)[\mathbf{S}_I(\mathbf{X})]_i^T[\mathbf{S}_c(\boldsymbol{\varepsilon}-\boldsymbol{\xi}^T\boldsymbol{\alpha}(u))]_i \\
&= I_{31} - I_{32}.
\end{aligned}
$$

By Lemma 1, it can be shown that

$$
\begin{aligned}
I_{31} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathbf{X}_i\mathbf{W}_i^T(nf(U_i)\boldsymbol{\Gamma}_c(U_i))^{-1}\sum_{j=1}^{n}K_{h_1}(U_j-U_i)\mathbf{W}_j(\varepsilon_j-\boldsymbol{\xi}_j^T\boldsymbol{\alpha}(U_j))\delta_j \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\Phi}(U_i)\boldsymbol{\Gamma}_c^{-1}(U_i)\mathbf{W}_i\delta_i(\varepsilon_i-\boldsymbol{\xi}_i^T\boldsymbol{\alpha}(U_i)) \\
&\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\Phi}_c(U_i)\boldsymbol{\Gamma}_c^{-1}(U_i)\mathbf{W}_i\delta_i(\varepsilon_i-\boldsymbol{\xi}_i^T\boldsymbol{\alpha}(U_i)) + o_p(1).
\end{aligned}
$$

In a similar way, we obtain that,

$$
\begin{aligned}
I_{32} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\Phi}(U_i)\boldsymbol{\Gamma}_c^{-1}(U_i)\mathbf{W}_i\delta_i(\varepsilon_i-\boldsymbol{\xi}_i^T\boldsymbol{\alpha}(U_i)) \\
&\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\Phi}(U_i)\boldsymbol{\Gamma}^{-1}(U_i)\mathbf{W}_i\delta_i(\varepsilon_i-\boldsymbol{\xi}_i^T\boldsymbol{\alpha}(U_i)) + o_p(1)
\end{aligned}
$$

Therefor,

$$
I_3 = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\boldsymbol{\Phi}(U_i)\boldsymbol{\Gamma}^{-1}(U_i)\mathbf{W}_i - \boldsymbol{\Phi}_c(U_i)\boldsymbol{\Gamma}_c^{-1}(U_i)\mathbf{W}_i]\delta_i(\varepsilon_i-\boldsymbol{\xi}_i^T\boldsymbol{\alpha}(U_i)) + o_p(1). \tag{29}
$$

I_4 can be expressed as

$$\begin{aligned} I_4 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{\mathbf{X}}_i(\mathbf{W}_i^T\boldsymbol{\alpha}(U_i) - [\mathbf{S}_I(\mathbf{M}^{\mathbf{W}})]_i) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{\mathbf{X}}_i\delta_i[\mathbf{S}_I(\varepsilon - \boldsymbol{\xi}^T\boldsymbol{\alpha}(u))]_i \\ &\quad - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{\mathbf{X}}_i(1-\delta_i)[\mathbf{S}_I(\mathbf{X})]_i^T(\hat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{\mathbf{X}}_i(1-\delta_i)[\mathbf{S}_I(\hat{\mathbf{M}}_c^{\mathbf{W}} - \mathbf{M}^{\mathbf{W}})]_i \\ &= I_{41} + I_{42} + I_{43} + I_{44} \end{aligned}$$

where $\hat{\mathbf{M}}_c^{\mathbf{W}} = [\mathbf{W}_1^T\hat{\boldsymbol{\alpha}}_c(U_1), \ldots, \mathbf{W}_n^T\hat{\boldsymbol{\alpha}}_c(U_n)]^T$. By Lemma 1, it can be shown that $I_{41} = o_p(1)$ and $I_{42} = o_p(1)$. By the fact that $\hat{\boldsymbol{\beta}}_c - \boldsymbol{\beta} = O_p(n^{-1/2})$ from Theorem 1 and $\frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{X}}_i[\mathbf{S}_I(\mathbf{X})]_i = o_p(1)$, $I_{43} = o_p(1)$ is obtained. $I_{44} = o_p(1)$ can also be proved similarly. Thus, we have

$$I_4 = o_p(1). \tag{30}$$

Similar to the calculation of I_4, we can show that

$$\begin{aligned} I_5 &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\bar{\mathbf{X}}_i([\mathbf{S}_I(\mathbf{M}^{\mathbf{W}})]_i - \mathbf{W}_i^T\boldsymbol{\alpha}(U_i)) \\ &= o_p(1). \end{aligned} \tag{31}$$

Invoking (26)–(31), it can be obtained that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}) = \Sigma^{-1}(\Sigma_1 + \Sigma_2)\Sigma_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{G}_i + o_p(1).$$

Thus, by the Slutsky theorem, Lemma 2 and the central limit theorem, we concludes the theorem.

Proof of Theorem 4 The proof of Theorem 4 is similar to Theorem 2, then, we omit it.

References

Ahmad I, Leelahanon S, Li Q (2005) Efficient estimation of a semiparametric partially linear varying coefficient model. Ann Stat 33:258–283

Fan JQ, Huang T (2005) Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli 11:1031–1057

Fan JQ, Li R (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc 9:1348–1360

Fan GL, Xu HX, Huang ZS (2016a) Empirical likelihood for semivarying coefficient model with measurement error in the nonparametric part. Adv Stat Anal 100:21–41

Fan GL, Liang HY, Shen Y (2016b) Penalized empirical likelihood for high-dimensional partially linear varying coefficient model with measurement errors. J Multivar Anal 147:183–201

Fan GL, Liang HY, Zhu LX (2018) Penalized profile least squares-based statistical inference for varying coefficient partially linear errors-in-variables models. Sci China Math 61:1677–1694

Feng SY, Xue LG (2014) Bias-corrected statistical inference for partially linear varying coefficient errors-in-variables models with restricted condition. Ann Inst Stat Math 66:121–140
Hu XM, Wang ZZ, Zhao ZW (2009) Empirical likelihood for semiparametric varying-coefficient partially linear errors-in-variables models. Stat Probab Lett 79:1044–1052
Kai B, Li R, Zou H (2011) New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. Ann Stat 39:305–332
Li TZ, Mei CL (2013) Estimation and inference for varying coefficient partially nonlinear models. J Stat Plann Infer 143:2023–2037
Liang H, Wang SJ, Carroll RJ (2007) Partially linear models with missing response variables and error-prone covariates. Biometrika 94:185–198
Long W, Ouyang M, Shang Y (2013) Efficient estimation of partially linear varying coefficient models. Econ Lett 121:79–81
Wang QH, Sun ZH (2007) Estimation in partially linear models with missing responses at random. J Multivar Anal 98:1470–1493
Wang QH, Xue LG (2011) Statistical inference in partially-varying-coefficient single-index model. J Multivar Anal 201:1–19
Wang QH, Linton O, Härdle W (2004) Semiparametric regression analysis with missing response at random. J Am Stat Assoc 99:334–345
Wang XL, Li GR, Lin L (2011) Empirical likelihood inference for semi-parametric varying-coefficient partially linear EV models. Metrika 73:171–185
Wei CH (2012a) Statistical inference in partially linear varying-coefficient models with missing responses at random. Commun Stat Theory Methods 41:1284–1298
Wei CH (2012b) Statistical inference for restricted partially linear varying coefficient errors-in-variables models. J Stat Plan Inference 142:2464–2472
Wei CH, Mei CL (2012) Empirical likelihood for partially linear varying-coefficient models with missing response variables and error-prone covariates. J Korean Stat Soc 41:97–103
Xia YC, Li WK (1999) On the estimation and testing of functional-coefficient linear models. Stat Sin 9:737–757
Xue LG, Xue D (2011) Empirical likelihood for semiparametric regression model with missing response data. J Multivar Anal 102:723–740
Yang H, Xia XC (2014) Equivalence of two tests in varying coefficient partially linear errors in variable model with missing responses. J Korean Stat Soc 43:79–90
Yang YP, Xue LG, Cheng WH (2011) Two-step estimators in partial linear models with missing response variables and error-prone covatiates. J Syst Sci Complex 24:1165–1182
You JH, Chen GM (2006) Estimation of a semiparametric varying-coefficient partially linear errors-in-variables model. J Multivar Anal 97:324–341
You JH, Zhou Y (2006) Empirical likelihood for semi-parametric varying coefficient partially linear model. Stat Probab Lett 76:412–422
Zhang WY, Lee SY, Song XY (2002) Local polynomial fitting in semivarying coefficient model. J Multivar Anal 82:166–188
Zhang WW, Li GR, Xue LG (2011) Profile inference on partially linear varying-coefficient errors-in-variables models under restricted condition. Comput Stat Data Anal 55:3027–3040
Zhang J, Zhou NG, Chen Q, Chu TY (2018) Nonlinear measurement errors models subject to partial linear additive distortion. Braz J Probab Stat 32:86–116