



Research article

Robust estimation for varying-coefficient partially nonlinear model with nonignorable missing response

Yanting Xiao* and Yifan Shi

Department of Applied Mathematics, Xi'an University of Technology, Xi'an, Shaanxi 710048, China

* **Correspondence:** Email: xiaoyanting03@163.com.

Abstract: In this paper, we studied the robust estimation for the varying-coefficient partially nonlinear model based on modal regression with nonignorable missing response. First, an instrumental variable was used to handle the identifiability issue of parameters in the propensity score function, and a generalized method of moment was combined to obtain the consistent estimators. Second, inverse probability weighting and modal regression were adopted to construct the estimators of parameters and coefficient function in the model. Under some mild conditions, the asymptotic properties of the resulting estimators were established. Furthermore, simulation studies and a real example were carried out to illustrate the effectiveness of our proposed estimation procedures.

Keywords: varying-coefficient partially nonlinear model; nonignorable missing data; instrumental variable; modal regression; inverse probability weighting

Mathematics Subject Classification: 62G10, 62G05

1. Introduction

Varying-coefficient partially linear model, one of the important semiparametric models, has attracted much attention due to the interpretability of parametric models and the flexibility of nonparametric models. In this model, we assume that the response and covariates with a parametric part have a linear relationship. However, this assumption may be inappropriate in some practical applications. For example, based on the empirical studies in an analysis of a real data set in ecology [1], it is more reasonable to assume that the relationship between net ecosystem CO_2 exchange (NEE) and the amount of photosynthetically active radiation (PAR) is nonlinear. The scatterplot between viral load and CD4 cell count shows that there is no rigorous linearity between them [2].

To describe the complicated potential nonlinear relationship between the response and certain covariates, Li and Mei [3] proposed the varying-coefficient partially nonlinear model, which takes the

form as

$$Y = \mathbf{X}^T \boldsymbol{\alpha}(U) + g(\mathbf{Z}, \boldsymbol{\beta}) + \varepsilon, \quad (1.1)$$

where Y is the response and $\mathbf{X} \in R^p$, $\mathbf{Z} \in R^r$ and U are the associated covariates. $g(\cdot, \cdot)$ is a pre-specified nonlinear function with parameter vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_q)^T$. $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \dots, \alpha_p(\cdot))^T$ is a coefficient function and ε is a model error with mean of zero and variance σ^2 . It is not necessary to demand that the dimension of \mathbf{Z} and $\boldsymbol{\beta}$ is equal.

It is obvious that model (1.1) is more flexible, which covers a varying-coefficient partially linear model as a special case, thus, it has been explored by extensive research since it was proposed. Li and Mei [3] proposed a profile nonlinear least squares estimation procedure and established the asymptotic properties of the resulting estimators. Zhou et al. [4] investigated the empirical likelihood inference and showed the corresponding Wilks phenomena. Furthermore, Yang and Yang [5] developed a two-stage estimation procedure based on the orthogonality-projection method and smooth-threshold estimating equations. In addition, Xiao and Chen [6] developed a local bias-corrected empirical likelihood procedure for model (1.1) with measurement errors in the nonparametric part. Tang et al. [7] proposed an adjusted empirical likelihood statistical inference procedure for model (1.1) with endogenous covariates.

The common above research is based on either the least squares method or the empirical likelihood method which are related to mean regression. As we all know, although they have satisfactory performance in the case of normal distributed data set, it is sensitive and may produce relative large bias when the data is subject to heavy-tailed distribution or contains outliers. Therefore, some robust statistical inferences are developed. For model (1.1), Jiang et al. [8] proposed a robust estimation procedure based on the exponential squared loss function, and Yang et al. [9] developed quantile regression for robust inference. Recently, Xiao and Liang [10] proposed a robust two-stage estimation procedure for model (1.1) based on modal regression.

Missing data is frequently encountered in research fields such as biomedicine, economics and clinical medicine. Common missing mechanism includes missing completely at random (MCAR), missing at random (MAR) and missing nonignorable at random (MNAR) (Little and Rubin [11]). The missing probability, which is called the propensity score function, depends only on the exactly observed variables in the assumption of MAR. At present, some research has explored statistical inference for model (1.1) under the MAR assumption. For example, with the missing covariates at random, Wang et al. [12] proposed the profile nonlinear least squares estimation procedure based on the inverse probability weighted method. Xia et al. [13] developed an empirical likelihood inference based on the weighted imputation method when the response was missing at random.

Nevertheless, it is possible that the propensity score is related to the value of the unobserved variable itself. For example, in the surveys about income, individuals with high income are usually less willing to provide their specific income, and the nonresponse rates tend to be related to the values of response. In this situation, the MNAR assumption would become reasonable. Compared to the MAR assumption, it is a great challenge to handle with MNAR since some parameters are not identifiable without any further restrictions on the propensity score. Over the past few years, MNAR data analysis has received a lot of attention in the literature. Kim and Yu [14] first defined the exponential tilting model for the nonignorably missing mechanism and proposed a semiparametric estimation procedure for the mean functions. Tang et al. [15] developed modified estimating equations by imputing missing data through a kernel regression method in the exponential tilting model. Wang et al. [16] proposed an instrument

variable, which is related to the study variable but unrelated to the missing data propensity, to construct estimating equations. To our knowledge, there is no literature to study the inferences for the varying-coefficient partially nonlinear model with nonignorable missing response. This motivates us to study this interesting topic. First, as an extension of the varying-coefficient partially linear model, the model has the ability of the interpretability of the parametric model and the flexibility of the nonparametric model. In addition, it can overcome the limitation of linear function in traditional semiparametric models. Second, compared to MAR, MNAR is much more common in practical applications and more challenging to study. Furthermore, as one of the robust estimation procedures, modal regression not only has good performances in the presence of outliers or heavy-tailed distributions, but it also has full asymptotic efficiency under the normal error distributions.

In this paper, we propose a robust estimation procedure for model (1.1) with nonignorable missing response. For the nonignorable nonresponse propensity, we assume a semiparametric model because the fully parametric approach is very sensitive to failure of the assumed parametric models. The instrument variable not involved in the propensity has been studied by Shao and Wang [17] to avoid the identifiability issue. The nonparametric component is replaced by a kernel-type estimator, and then the parameter is estimated by the profiled estimating equations and the generalized method of moments. We then propose robust estimators in the varying-coefficient partially nonlinear model based on the inverse probability weighted technique and modal regression. On the one hand, estimation procedures of using complete-case data may result in serious bias, especially when the missing probability is large. In consequence, the inverse probability weighted technique provides a modified way to reduce the estimation bias by using the inverse of the propensity as weights. On the other hand, modal regression, introduced by Yao et al. [18], has been extensively applied to other semiparametric models, such as [19, 20]. Modal regression has many advantages over the traditional mean regression. First, it is easy to implement by involving a tuning parameter that can be automatically selected. Second, this method has good robustness in the presence of outliers or heavy-tailed errors, which has been successfully verified by extensive literature. Meanwhile, the obtained estimator can achieve fully asymptotic efficiency under the normal error distribution in terms of theoretical properties. This encourages us to adopt inverse probability weighting and modal regression to the varying-coefficient partially nonlinear model.

The rest of this paper is organized as follows. Section 2 introduces a robust estimation procedure based on modal regression and inverse probability weighting. Section 3 establishes its theoretical properties. In Section 4, we discuss the selection of bandwidth and the specific estimation algorithm. Simulation studies and a real data are then conducted to evaluate the performances of the proposed estimation procedure in Sections 5 and 6. We make our concluding remarks in Section 7 and leave the proofs of the main theorems to the Appendix.

2. Estimation methodology

2.1. Modal regression procedure with nonignorable missing response

Suppose that $\{Y_i, \mathbf{X}_i, \mathbf{Z}_i, U_i\}_{i=1}^n$ is a random sample from model (1.1); that is,

$$Y_i = \mathbf{X}_i^T \boldsymbol{\alpha}(U_i) + g(\mathbf{Z}_i, \boldsymbol{\beta}) + \varepsilon_i, i = 1, 2, \dots, n, \quad (2.1)$$

where covariates X_i, Z_i, U_i are completely observed and response Y_i has nonignorable missing values. Let δ_i be a binary response indicator for Y_i , where $\delta_i = 1$ if Y_i is observed, and $\delta_i = 0$ otherwise. Furthermore, we define the propensity of missing data as follows

$$\pi(Y_i, T_i) = Pr(\delta_i = 1 | Y_i, X_i, Z_i, U_i),$$

where $T_i = (X_i^T, Z_i^T, U_i)^T$. In the MNAR assumption, the propensity of missing data depends on Y_i , regardless of whether Y_i is observed or missing.

To simplify model (2.1), we apply B-spline functions to approximate the nonparametric function $\alpha(u)$ instead of utilizing the local polynomial estimation. This choice is motivated by the fact that B-spline functions approximation possesses bounded support and numerical stability. In addition, B-spline functions approximation avoids the issue of high computational complexity associated with local polynomial estimation, as elaborated in [21]. More specifically, let $\mathbf{B}(u) = (B_1(u), \dots, B_L(u))^T$ be the B-spline basis function with the order of M , where $L = K + M$ and K is the number of interior knots. Therefore, the nonparametric function $\alpha_k(u)$ can be approximated by

$$\alpha_k(u) \approx \mathbf{B}^T(u)\boldsymbol{\gamma}_k, k = 1, 2, \dots, p, \quad (2.2)$$

where $\boldsymbol{\gamma}_k = (\gamma_{k1}, \gamma_{k2}, \dots, \gamma_{kL})^T$. Then, we substitute (2.2) into model (2.1) to obtain

$$Y_i \approx \mathbf{W}_i^T \boldsymbol{\gamma} + g(\mathbf{Z}_i, \boldsymbol{\beta}) + \varepsilon_i, i = 1, 2, \dots, n, \quad (2.3)$$

where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T, \boldsymbol{\gamma}_2^T, \dots, \boldsymbol{\gamma}_p^T)^T$ and $\mathbf{W}_i = \mathbf{I}_p \otimes \mathbf{B}(U_i) \cdot X_i$, with \mathbf{I}_p being a p -dimensional identity matrix.

For fixed $\boldsymbol{\beta}$, model (2.3) can be reexpressed as

$$\mathbf{Y} - g(\mathbf{Z}, \boldsymbol{\beta}) \approx \mathbf{W}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad (2.4)$$

where $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^T$, $g(\mathbf{Z}, \boldsymbol{\beta}) = (g(\mathbf{Z}_1, \boldsymbol{\beta}), g(\mathbf{Z}_2, \boldsymbol{\beta}), \dots, g(\mathbf{Z}_n, \boldsymbol{\beta}))^T$, $\mathbf{W} = (\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_n)^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$.

Clearly, model (2.4) is a standard linear model. We can obtain the initial estimators of $\boldsymbol{\gamma}$ with the ordinary least squares method as follows

$$\hat{\boldsymbol{\gamma}}^{(1)} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T (\mathbf{Y} - g(\mathbf{Z}, \boldsymbol{\beta})).$$

We replace $\boldsymbol{\gamma}$ in (2.4) by $\hat{\boldsymbol{\gamma}}^{(1)}$, and model (2.4) becomes

$$\mathbf{P}\mathbf{Y} \approx \mathbf{P}g(\mathbf{Z}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}, \quad (2.5)$$

where $\mathbf{P} = \mathbf{I}_n - \mathbf{W}(\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T$, with \mathbf{I}_n being an n -dimensional identity matrix.

Then, the inverse probability weighting technique is adopted to reduce the estimation bias when Y_i is missing and modal regression is involved to attained the robust estimators. Motivated by the idea of Yao et al. [18], the modal estimator for $\boldsymbol{\beta}$ is constructed by maximizing the following objective function

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^n \Delta_i \phi_h [\mathbf{P}_i (\mathbf{Y} - g(\mathbf{Z}, \boldsymbol{\beta}))], \quad (2.6)$$

where $\Delta_i = \delta_i/\pi(Y_i, \mathbf{T}_i)$, $\phi_h(t) = \phi(t/h)/h$ with kernel density function $\phi(\cdot)$ and bandwidth h . \mathbf{P}_i denotes the i th row of the matrix \mathbf{P} . According to Yao et al. [18], we adopt the Gaussian kernel density function for $\phi(\cdot)$ throughout the paper. The specific selection of h is described in Section 4.1.

However, the propensity score $\pi(Y_i, \mathbf{T}_i)$ is usually unknown in practice, and it is common to use the estimator $\hat{\pi}(Y_i, \mathbf{T}_i)$ to replace $\pi(Y_i, \mathbf{T}_i)$. The estimation procedure of the propensity $\pi(Y_i, \mathbf{T}_i)$ is given in the next subsection.

2.2. Estimation for semiparametric nonignorable propensity

According to Shao and Wang [17], we assume the following semiparametric model for the propensity

$$\pi(Y_i, \mathbf{T}_i) = Pr(\delta_i = 1 | Y_i, \mathbf{X}_i, \mathbf{Z}_i, U_i) = \frac{1}{1 + \psi(\mathbf{T}_i)q(Y_i, \boldsymbol{\zeta})}, \quad (2.7)$$

where $q(Y_i, \boldsymbol{\zeta})$ is a known function of $\boldsymbol{\zeta}$, which is a d -dimensional unknown parameter, and $\psi(\cdot)$ is an unknown function. Several common models are special cases of the semiparametric model (2.7). For example, (2.7) reduces to the model for ignorable missing data when $q(Y_i, \boldsymbol{\zeta})$ does not depend on Y_i . When $q(Y_i, \boldsymbol{\zeta}) = \exp(\boldsymbol{\zeta}Y_i)$, (2.7) simplifies to the exponential tilting model for nonignorable missing data, as described in Kim and Yu [14].

Without any further assumptions, Shao and Wang [17] showed that $\psi(\cdot)$ and $\boldsymbol{\zeta}$ are not identifiable in propensity score function (2.7). To overcome this difficulty, we assume that \mathbf{T}_i can be decomposed as two parts $\mathbf{T}_i = (\mathbf{V}_i^T, \mathbf{S}_i^T)^T$ and \mathbf{S}_i can be excluded from the propensity model, then (2.7) can be reduced as

$$\pi(Y_i, \mathbf{T}_i) = \pi(Y_i, \mathbf{V}_i) = \frac{1}{1 + \psi(\mathbf{V}_i)q(Y_i, \boldsymbol{\zeta})}, \quad (2.8)$$

where $\psi(\cdot)$ is an unspecified function of \mathbf{V}_i . The covariate \mathbf{S}_i is referred to the instrument variable by Wang et al. [16], which aids in identifying and estimating all unknown parameters.

To estimate the propensity defined by (2.8), for any fixed $\boldsymbol{\zeta}$, we have

$$\psi_{\boldsymbol{\zeta}}(\mathbf{V}) = \frac{E\{(1 - \delta) | \mathbf{V}\}}{E\{\delta q(Y, \boldsymbol{\zeta}) | \mathbf{V}\}}.$$

Thus, a nonparametric kernel estimator of $\psi_{\boldsymbol{\zeta}}(\mathbf{V})$ is given by

$$\hat{\psi}_{\boldsymbol{\zeta}}(\mathbf{V}) = \frac{\sum_{i=1}^n (1 - \delta_i) L_b(\mathbf{V} - \mathbf{V}_i)}{\sum_{i=1}^n \delta_i q(Y_i, \boldsymbol{\zeta}) L_b(\mathbf{V} - \mathbf{V}_i)}, \quad (2.9)$$

where $L_b(\cdot) = b^{-1}L(\cdot/b)$, $L(\cdot)$ is a kernel function and b is a bandwidth. Note that $\hat{\psi}_{\boldsymbol{\zeta}}(\mathbf{V})$ is not an estimator of $\psi(\mathbf{V})$ as $\boldsymbol{\zeta}$ is an unknown parameter value. However, $\hat{\psi}_{\boldsymbol{\zeta}}(\mathbf{V})$ is useful in the following estimating equations for estimating unknown $\boldsymbol{\zeta}$. Define

$$f(Y, \mathbf{T}, \delta, \psi_{\boldsymbol{\zeta}}, \boldsymbol{\zeta}) = \left\{ \frac{\delta}{\pi(Y, \mathbf{T}, \psi_{\boldsymbol{\zeta}}, \boldsymbol{\zeta})} - 1 \right\} h(\mathbf{S}), \quad (2.10)$$

where $\pi(Y, \mathbf{T}, \psi_\zeta, \zeta) = \{1 + \psi(\mathbf{V})q(Y, \zeta)\}^{-1}$ and $h(\mathbf{S})$ is a known vector-valued function with dimension $R \geq d + 1$. As in Wang et al. [16], we recommend to use $h(\mathbf{S}) = (1, \mathbf{V}^T, \mathbf{S}^T)^T$ throughout this paper. It is easy to verify that $E[f(Y, \mathbf{T}, \delta, \psi_{\zeta_0}, \zeta_0)] = \mathbf{0}$ under the true parameter value ζ_0 . The estimating equations are over-identifies because of $R > d$, and we employ the two-step generalized method of moments (GMM).

Let $F_n(\hat{\psi}_\zeta, \zeta) = \frac{1}{n} \sum_{i=1}^n f(Y_i, \mathbf{T}_i, \delta_i, \hat{\psi}_\zeta, \zeta)$ and $\hat{\zeta}^{(1)} = \arg \min_{\zeta} F_n(\hat{\psi}_\zeta, \zeta)^T F_n(\hat{\psi}_\zeta, \zeta)$. The two-step GMM estimator of ζ is

$$\hat{\zeta} = \arg \min_{\zeta} F_n(\hat{\psi}_\zeta, \zeta)^T \hat{W}_n^{-1} F_n(\hat{\psi}_\zeta, \zeta), \quad (2.11)$$

where $\hat{W}_n = \frac{1}{n} \sum_{i=1}^n f(Y_i, \mathbf{T}_i, \delta_i, \hat{\psi}_{\hat{\zeta}^{(1)}}) f(Y_i, \mathbf{T}_i, \delta_i, \hat{\psi}_{\hat{\zeta}^{(1)}})^T$. Eventually, the propensity model can be consistently estimated by

$$\hat{\pi}(Y_i, \mathbf{T}_i) : \hat{\pi}(Y_i, \mathbf{T}_i, \hat{\zeta}) = \frac{1}{1 + \hat{\psi}_{\hat{\zeta}}(\mathbf{V}_i)q(Y_i, \hat{\zeta})}. \quad (2.12)$$

Then, the modal estimator $\hat{\beta}$ for β is constructed by maximizing the following objective function

$$\hat{Q}(\beta) = \sum_{i=1}^n \hat{\Delta}_i \phi_h[\mathbf{P}_i(\mathbf{Y} - g(\mathbf{Z}, \beta))], \quad (2.13)$$

where $\hat{\Delta}_i = \delta_i / \hat{\pi}(Y_i, \mathbf{T}_i)$.

Once modal estimator $\hat{\beta}$ is obtained, model (2.3) can be transformed to

$$Y_i^* \approx \mathbf{W}_i^T \boldsymbol{\gamma} + \varepsilon_i, i = 1, \dots, n, \quad (2.14)$$

where $Y_i^* = Y_i - g(\mathbf{Z}_i, \hat{\beta})$.

Similarly, the modal estimators of $\boldsymbol{\gamma}$ are obtained by maximizing the following objective function

$$\hat{Q}(\boldsymbol{\gamma}) = \sum_{i=1}^n \hat{\Delta}_i \phi_h[(Y_i^* - \mathbf{W}_i^T \boldsymbol{\gamma})]. \quad (2.15)$$

Therefore, we have the estimator of coefficient functions by $\hat{\alpha}_k(u) = \mathbf{B}(u)^T \hat{\boldsymbol{\gamma}}_k, k = 1, \dots, p$.

3. Theoretical properties

In this section, we discuss the asymptotic properties of the proposed modal regression estimators for both parametric and nonparametric. Denote the true values of β and $\alpha_k(\cdot)$ as β_0 and $\alpha_{0k}(\cdot)$ in model (2.1), respectively. Correspondingly, $\boldsymbol{\gamma}_{0k}$ is the optimal approximation coefficient of $\boldsymbol{\gamma}_k$ in (2.2). Let $F(x, z, u, h) = E(\phi_h''(\varepsilon)|X = x, Z = z, U = u)$ and $G(x, z, u, h) = E(\phi_h'(\varepsilon)^2|X = x, Z = z, U = u)$. To facilitate the presentation, we denote by $g'(z, \beta) = \partial g(z, \beta) / \partial \beta$ the $q \times 1$ vector of the first order derivatives of $g(z, \beta)$ with respect to β .

The required assumption conditions of the main results are derived in the following. These assumption are common and can be easily satisfied.

A1: The random variable U has a bounded support \mathcal{U} . Its density function $f(\cdot)$ is Lipschitz continuous, bounded away from 0 to infinite on its support.

A2: All the true coefficient functions $\alpha_{0k}(u), k = 1, \dots, p$ are all r th continuously differentiable on the interval $(0, 1)$ with $r \geq 2$.

A3: For any z , $g(z, \beta)$ is a continuous function of β , and the 2th derivative of $g(z, \beta)$ with respect to β is continuous.

A4: There exists an $s > 0$ such that $E\|\mathbf{X}\|^{2s} < \infty, E\|g'(z, \beta)\|^{2s} < \infty$.

A5: The bandwidth h satisfies as $n \rightarrow \infty, nh^{2r} \rightarrow \infty, nh/\log n \rightarrow \infty$.

A6: Suppose that ζ_0 is the unique solution to $E[f(Y, \mathbf{T}, \delta, \psi_{\zeta}, \zeta)] = 0, \Gamma = E[\partial f(Y, \mathbf{T}, \delta, \psi_{\zeta_0}, \zeta_0)/\partial \zeta]$ is of full rank and $D = E[f(Y, \mathbf{T}, \delta, \psi_{\zeta_0}, \zeta_0)^{\otimes 2}]$ is positive definite.

A7: Let c_1, \dots, c_K be the interior knots on $[0, 1]$. Set $c_0 = 0, c_{K+1} = 1$ and $h_i = c_i - c_{i-1}$. Then, there exists a constant C_0 such that

$$\frac{\max\{h_i\}}{\min\{h_i\}} \leq C_0, \quad \max\{|h_{i+1} - h_i|\} = o(K^{-1}).$$

A8: $F(x, z, u, h)$ and $G(x, z, u, h)$ are continuous with respect to (x, z, u) . In addition, $F(x, z, u, h) < 0$ for any $h > 0$.

A9: $E(\phi'_h(\varepsilon)|x, z, u) = 0$ and $E(\phi''_h(\varepsilon)^2|x, z, u), E(\phi'_h(\varepsilon)^3|x, z, u)$ and $E(\phi'''_h(\varepsilon)|x, z, u)$ are continuous with respect to (x, z, u) .

A10: The matrix A and Σ defined in Theorem 1 are positive definite.

These assumptions are commonly adopted in the previous semiparametric model. Assumptions A1–A4 are generally required in the varying-coefficient partially nonlinear model. Assumption A5 is common with bandwidth in kernel function. Assumption A6 is necessary for proving the asymptotic properties of two-step GMM estimator ζ . Assumption A7 indicates that c_1, \dots, c_K is a quasi-uniform sequence of partitions on $[0, 1]$. Assumptions A8 and A9 are commonly used in the modal regression. Assumption A10 ensures the existence of the asymptotic variance of the estimator $\hat{\beta}$.

The following Theorem 1 gives the asymptotic normality for the parameter estimator $\hat{\beta}$.

Theorem 1. Suppose that assumptions A1–A10 hold and the number of interior knots is $K = O(n^{1/(2r+1)})$. As $n \rightarrow \infty$, we have

$$\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, C^{-1}),$$

where $C = \Sigma A^{-1} \Sigma^T, \Sigma = E(F(x, z, u, h)g'(z, \beta)g'(z, \beta)^T), A = B + H\Pi H^T, B = E[\pi(Y, \mathbf{T}, \zeta_0)^{-1}G(x, z, u, h)g'(z, \beta)g'(z, \beta)^T], \Pi$ is the asymptotic variance of $\hat{\zeta}$ defined in remark 1, $H = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E[\pi(Y, \mathbf{T}, \zeta_0)^{-1} \phi'_h(\varepsilon)(\mathbf{P}_i g'(\beta))^T \{\partial \pi(Y, \mathbf{T}, \zeta_0)/\partial \zeta\}^T]$ and \mathbf{P}_i is the i th row of $\mathbf{P} = \mathbf{I}_n - \mathbf{W}(\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T, g'(\beta) = (g'(\mathbf{Z}_1, \beta), \dots, g'(\mathbf{Z}_n, \beta))^T$.

Remark 1. As discussed in [22], $\hat{\zeta}$ is a consistent estimator of ζ , and

$$n^{1/2}(\hat{\zeta} - \zeta) \xrightarrow{d} N(0, \Pi),$$

where $\Pi = (\Xi^T D^{-1} \Xi)^{-1} \Xi^T D^{-1} \Omega D^{-1} \Xi (\Xi^T D^{-1} \Xi)^{-1}, \Xi = E\{(1 - \delta)[h(\mathbf{S}) - m_s(\mathbf{V})][\eta(Y, \zeta_0) - m_\eta(\mathbf{V})]\}, m_\eta(\mathbf{V}) = E\{\delta \eta(Y, \zeta_0)q(Y, \zeta_0)|\mathbf{V}\}/E\{\delta q(Y, \zeta_0)|\mathbf{V}\}, m_s(\mathbf{V}) = E\{\delta h(\mathbf{S})q(Y, \zeta_0)|\mathbf{V}\}/E\{\delta q(Y, \zeta_0)|\mathbf{V}\}, \eta(Y, \zeta_0) =$

$\{q(Y, \zeta_0)\}^{-1} \partial q(Y, \zeta_0) / \partial \zeta^\top$, Ω is the covariate matrix of $\Lambda = \{h(\mathbf{S}) - m_s(\mathbf{V})\} [\delta\{1 + \psi_{\zeta_0}(\mathbf{V})q(Y, \zeta_0)\} - 1] + \mathcal{M}$, $\mathcal{M} = (\mathcal{M}_1, \dots, \mathcal{M}_R)^\top$, $\mathcal{M}_r = \text{vec}(\Phi_r)^\top \ell(\mathbf{T}, Y, \delta)$, $r = 1, \dots, R$, $\ell(\mathbf{T}, Y, \delta)$ is an influence function, R is the dimension of $h(\mathbf{S})$, $\Phi_r = E\{\psi_{\zeta_0}(\mathbf{V})[\nabla C_r(\mathbf{V})f_v(\mathbf{V}) + C_r(\mathbf{V})\nabla f_v(\mathbf{V})]/f_v(\mathbf{V})\}$, $C_r(\mathbf{V}) = \text{Cov}(\mathbf{T}, \tilde{\mathbf{W}}_r | \mathbf{V})$, $\nabla C_r(\mathbf{V}) = \partial C_r(\mathbf{V}) / \partial \mathbf{V}^\top$, $\nabla f_v(\mathbf{V}) = \partial f_v(\mathbf{V}) / \partial \mathbf{V}^\top$, $f_v(\mathbf{V})$ is the density of \mathbf{V} and $\tilde{\mathbf{W}}_r$ is the r th coordinate of $\delta q(Y, \zeta_0)[h(\mathbf{S}) - m_s(\mathbf{V})]$.

The following Theorem 2 gives the consistency of estimator $\hat{\alpha}_k(\cdot)$, $k = 1, \dots, p$.

Theorem 2. *Suppose that assumptions A1–A10 hold, and the number of interior knots is $K = O(n^{1/(2r+1)})$. As $n \rightarrow \infty$, we have*

$$\|\hat{\alpha}_k(\cdot) - \alpha_{0k}(\cdot)\| = O_p(n^{-\frac{r}{2r+1}}), k = 1, \dots, p.$$

4. Bandwidth selection and estimation algorithm

In this section, we discuss the selection of the bandwidth and estimation procedure based on the modal expectation maximization (MEM) algorithm proposed by Li et al. [23].

4.1. Bandwidth selection

To obtain the modal estimators $\hat{\beta}$ and $\hat{\gamma}$ of parameters β and γ , it is necessary to select an appropriate bandwidth h . For simplicity, we assume that the error variable is independent of \mathbf{X} , \mathbf{Z} and U . According to the thought of Yao et al. [18], the ratio of the asymptotic variance of our proposed estimator to that of the least square estimator is given by

$$R(h_1) = \frac{G(h_1)}{F(h_1)^2 \sigma_1^2}, \quad (4.1)$$

where $\sigma_1^2 = E(\varepsilon^2)$, $F(h_1) = E[\phi'_{h_1}(\varepsilon)]$ and $G(h_1) = E[\phi'_{h_1}(\varepsilon)^2]$. In practice, we do not know the error distribution, hence, we cannot obtain $F(h_1)$ and $G(h_1)$. A feasible method is with estimator $F(h_1)$ and $G(h_1)$ by $\hat{F}(h_1) = \frac{1}{n} \sum_{i=1}^n \phi'_{h_1}(\hat{\varepsilon}_{1i})$, and $\hat{G}(h_1) = \frac{1}{n} \sum_{i=1}^n [\phi'_{h_1}(\hat{\varepsilon}_{1i})]^2$, respectively.

Then, the estimator for $R(h_1)$ is given by

$$\hat{R}(h_1) = \frac{\hat{G}(h_1)}{\hat{F}(h_1)^2 \hat{\sigma}_1^2}, \quad (4.2)$$

where $\hat{\varepsilon}_{1i} = \hat{\Delta}_i \mathbf{P}_i [\mathbf{Y} - g(\mathbf{Z}, \hat{\beta}^{\text{int}})]$ and $\hat{\sigma}_1^2 = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{1i}^2$, $\hat{\beta}^{\text{int}}$ are the pilot estimators.

Similarly, to obtain the modal estimator $\hat{\gamma}$, the estimator for the ratio of the asymptotic variance of our proposed estimator to that of the least square estimator is given by

$$\hat{R}(h_2) = \frac{\hat{G}(h_2)}{\hat{F}(h_2)^2 \hat{\sigma}_2^2}, \quad (4.3)$$

where $\hat{F}(h_2) = \frac{1}{n} \sum_{i=1}^n \phi'_{h_2}(\hat{\varepsilon}_{2i})$, $\hat{G}(h_2) = \frac{1}{n} \sum_{i=1}^n [\phi'_{h_2}(\hat{\varepsilon}_{2i})]^2$, $\hat{\varepsilon}_{2i} = \hat{\Delta}_i (Y_i - g(\mathbf{Z}_i, \hat{\beta}) - \mathbf{X}_i^T \hat{\alpha}^{\text{int}}(U_i))$ and $\hat{\sigma}_2^2 = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_{2i}^2$, $\hat{\beta}$ are the modal estimators and $\hat{\alpha}^{\text{int}}(\cdot)$ are the pilot estimators.

Furthermore, we use the grid search method to choose appropriate bandwidth. h_{opt1} and h_{opt2} can be obtained by minimizing (4.2) and (4.3), respectively. The optimal choice of h_1 and h_2 are

$$h_{opt1} = \arg \min_{h_1} \hat{R}(h_1), h_{opt2} = \arg \min_{h_2} \hat{R}(h_2).$$

Following Yao et al. [18], the possible grid points for bandwidth can be $h_1 = 0.5\hat{\sigma}_1 \times 1.02^j$, $h_2 = 0.5\hat{\sigma}_2 \times 1.02^j$, $j = 0, 1, \dots, 100$.

4.2. The MEM algorithm of parametric components

In this subsection, we adopt the following MEM algorithm by Li et al. [23] to obtain the estimators of parameters β by maximizing (2.13). Let $\beta^{(0)}$ be initial estimators of β with $m = 0$.

Step 1 (E-step): We update $\pi(i|\beta^{(m)})$ by

$$\pi(i|\beta^{(m)}) = \frac{\hat{\Delta}_i \phi_h\{\mathbf{P}_i[\mathbf{Y} - g(\mathbf{Z}, \beta^{(m)})]\}}{\sum_{i=1}^n \hat{\Delta}_i \phi_h\{\mathbf{P}_i[\mathbf{Y} - g(\mathbf{Z}, \beta^{(m)})]\}}, i = 1, 2, \dots, n.$$

Step 2 (M-step): We update $\beta^{(m+1)}$ by

$$\begin{aligned} \beta^{(m+1)} &= \arg \max_{\beta} \sum_{i=1}^n \pi(i|\beta^{(m)}) \hat{\Delta}_i \log \phi_h\{\mathbf{P}_i[\mathbf{Y} - g(\mathbf{Z}, \beta)]\} \\ &= \arg \max_{\beta} \sum_{i=1}^n \pi(i|\beta^{(m)}) \hat{\Delta}_i \log \phi_h\{\mathbf{P}_i[\mathbf{Y} - g(\mathbf{Z}, \beta^{(m)}) - g'(\mathbf{Z}, \beta^{(m)})(\beta - \beta^{(m)})]\} \\ &= (\mathbf{G}^T \mathbf{D} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{D} \hat{\mathbf{Y}}, \end{aligned}$$

where $\hat{\mathbf{Y}} = (\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n)^T$, $\hat{Y}_i = \hat{\Delta}_i \mathbf{P}_i [\mathbf{Y} - g(\mathbf{Z}, \beta^{(m)}) + g'(\mathbf{Z}, \beta^{(m)})\beta^{(m)}]$, $\mathbf{G} = (\hat{\Delta}_1 \mathbf{P}_1 g'(\mathbf{Z}, \beta^{(m)}), \hat{\Delta}_2 \mathbf{P}_2 g'(\mathbf{Z}, \beta^{(m)}), \dots, \hat{\Delta}_n \mathbf{P}_n g'(\mathbf{Z}, \beta^{(m)}))^T$, \mathbf{D} is $n \times n$ diagonal matrix with $\mathbf{D} = \text{diag}(\pi(1|\beta^{(m)}), \pi(2|\beta^{(m)}), \dots, \pi(n|\beta^{(m)}))$.

Step 3: Let $m = m + 1$ and iterate the E-step and M-step until the algorithm converges. Denote the final estimator of β as $\hat{\beta}$.

4.3. The MEM algorithm of nonparametric components

Based on the MEM algorithm, we can obtain the estimators of parameters γ by maximizing (2.15). Let $\gamma^{(0)}$ be initial estimators of γ with $m = 0$.

Step 1 (E-step): We update $\pi(i|\gamma^{(m)})$ by

$$\pi(i|\gamma^{(m)}) = \frac{\hat{\Delta}_i \phi_h[(Y_i^* - \mathbf{W}_i^T \gamma^{(m)})]}{\sum_{i=1}^n \hat{\Delta}_i \phi_h[(Y_i^* - \mathbf{W}_i^T \gamma^{(m)})]}, i = 1, 2, \dots, n.$$

Step 2 (M-step): We update $\gamma^{(m+1)}$ by

$$\gamma^{(m+1)} = \arg \max_{\gamma} \sum_{i=1}^n \pi(i|\gamma^{(m)}) \hat{\Delta}_i \log \phi_h[(Y_i^* - \mathbf{W}_i^T \gamma)]$$

$$= (\tilde{W}^T \tilde{D} \tilde{W})^{-1} \tilde{W}^T \tilde{D} \tilde{Y}, \quad (4.4)$$

where $\tilde{Y} = \Delta Y^*$, $\tilde{W} = \Delta W$, $\tilde{D} = \text{diag}(\pi(1|\gamma^{(m)}), \dots, \pi(n|\gamma^{(m)}))$ and $\Delta = \text{diag}(\hat{\Delta}_1, \dots, \hat{\Delta}_n)$.

Step 3: Let $m = m + 1$ and iterate the E-step and M-step until the algorithm converges. Denote the final estimator of γ as $\hat{\gamma}$. The corresponding estimator of nonparametric function is $\hat{\alpha}_k(u) = \mathbf{B}(u)^T \hat{\gamma}_k, k = 1, \dots, p$.

5. Simulation studies

In this section, we conduct some simulations to evaluate the finite sample performances of the proposed estimation procedures. We generate the data from the varying-coefficient partially nonlinear model described as follows:

$$Y = \mathbf{X}^T \boldsymbol{\alpha}(U) + g(\mathbf{Z}, \boldsymbol{\beta}) + \varepsilon, \quad (5.1)$$

where the nonlinear function $g(\mathbf{Z}, \boldsymbol{\beta}) = \exp(Z_1 \beta_1 + Z_2 \beta_2)$ with parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2)^T = (1, 1.5)^T$, the coefficient functions $\boldsymbol{\alpha}(U) = (\alpha_1(U), \alpha_2(U))^T$ with $\alpha_1(U) = \sin(2\pi U)$ and $\alpha_2(U) = 3.5\{\exp[-(4U - 1)^2] + \exp[-(4U - 3)^2] - 1.5\}$. The variable U is generated from uniform distribution $U(0, 1)$. The variable $\mathbf{X} = (X_1, X_2)^T$ follows $(0, \Sigma)$ with $\Sigma_{ij} = 0.5^{|i-j|}, 1 \leq i, j \leq 2$ and $\mathbf{Z} = (Z_1, Z_2)^T$ and it is the same generated as \mathbf{X} . To illustrate the robustness of the proposed method, the following three different distribution of model errors are considered: (1) The normal distribution: $\varepsilon \sim N(0, 1)$; (2) the t distribution: $\varepsilon \sim t(3)$; (3) the mixed normal distribution(MN): $\varepsilon \sim 0.9N(0, 1^2) + 0.1N(0, 9^2)$. The sample sizes n are 200 and 400, separately, and the simulations are based on 500 replications. Furthermore, we generate δ_i from the Bernoulli distribution with probability $\pi(Y_i, \mathbf{V}_i) = \{1 + \psi(\mathbf{V}_i) \exp\{\zeta Y_i\}\}^{-1}$ and consider four choices of the function $\psi(\cdot)$ and parameter ζ :

Case1. $\psi(\mathbf{V}_i) = \exp\{-0.1 - 1.5X_{1i} - 1.5Z_{1i} - 1.5U_i\}$ and $\zeta = 0$;

Case2. $\psi(\mathbf{V}_i) = \exp\{-0.1 + 0.5X_{1i} + 0.5Z_{1i} + 0.5U_i\}$ and $\zeta = -0.8$;

Case3. $\psi(\mathbf{V}_i) = \exp\{-0.1 + 0.5 \sin(X_{1i}) + 0.5 \sin(Z_{1i}) + 0.5 \sin(U_i)\}$ and $\zeta = -0.8$;

Case4. $\psi(\mathbf{V}_i) = \exp\{0.6 - 0.3 \exp(X_{1i}) - 0.3 \exp(Z_{1i}) - 0.3 \exp(U_i)\}$ and $\zeta = -0.1$.

While Case 1 is an ignorable missing mechanism, Cases 2–4 represent three different kinds of nonignorable missing mechanisms. The coefficients in the aforementioned four probability models are chosen so that the average proportions of missing data are between 30% and 40%. It can be seen that $\mathbf{V}_i = (X_{1i}, Z_{1i}, U_i)^T$ and the instrument variable $\mathbf{S}_i = (X_{2i}, Z_{2i})^T$. As in Shao and Wang [17], we select the Gaussian kernel function $L(\cdot) = \frac{1}{\sqrt{2\pi}} \exp(-(\cdot)^2/2)$ to compute the nonparametric kernel estimator with $L(X_{1i}, Z_{1i}, U_i) = L(X_{1i})L(Z_{1i})L(U_i)$, and choose the bandwidth to be $b_1 = 1.5\hat{\sigma}_{X_1}n^{-1/3}$, $b_2 = 1.5\hat{\sigma}_{Z_1}n^{-1/3}$ and $b_3 = 1.5\hat{\sigma}_U n^{-1/3}$, where $\hat{\sigma}_{X_1}, \hat{\sigma}_{Z_1}, \hat{\sigma}_U$ is the standard deviation of datasets $\{X_{1i}\}, \{Z_{1i}\}, \{U_i\}, i = 1, 2, \dots, n$.

To evaluate the efficiencies of the proposed estimation procedure, four estimators of $\boldsymbol{\beta}$ are considered as follows:

(i) The proposed modal regression (MR) estimator is based on (2.13), which is denoted as MNAR-MR.

(ii) The MR estimator is based on (2.13) in which $\hat{\pi}$ is estimated with MAR assumption, which is denoted as MAR-MR.

(iii) The MR estimator is based on complete-case(CC) data, which is denoted as CC-MR.

(iv) The MR estimator is based on full sample data, which is denoted as FULL-MR.

Further, in order to assess the robustness of the proposed procedure, we also compare the profile least square (PLS) method in four scenarios: (i) The PLS estimator based on (2.13) which is denoted as MNAR-PLS, (ii) The PLS estimator based on (2.13) in which $\hat{\pi}$ is estimated with MAR assumption, which is denoted as MAR-PLS, (iii) The PLS estimator based on complete-case data, which is denoted as CC-PLS and (iv) The PLS estimator based on full sample data, which is denoted as FULL-PLS. The absolute Bias and standard deviation (SD) with the proposed MR method and the PLS method are reported in Tables 1–4, respectively.

Table 1. Bias and SD of four estimators under proposed MR procedure with $n = 200$.

ε	π		MNAR-MR		MAR-MR		CC-MR		FULL-MR	
			β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
$N(\mathbf{0}, \mathbf{1})$	Case1	Bias	-0.0041	0.0057	-0.0042	0.0032	-0.0078	-0.0089	-0.0036	0.0002
		SD	0.0363	0.0485	0.0742	0.0808	0.0624	0.0710	0.0082	0.0072
	Case2	Bias	-0.0017	-0.0006	0.0999	0.0114	-0.0038	-0.0015	0.0006	0.0002
		SD	0.0316	0.0305	0.0935	0.0861	0.0704	0.0196	0.0103	0.0091
	Case3	Bias	-0.0007	-0.0008	-0.0190	0.0223	-0.0019	0.0112	-0.0003	0.0001
		SD	0.0309	0.0299	0.1962	0.1218	0.0959	0.0933	0.0088	0.0075
	Case4	Bias	-0.0004	0.0001	0.0275	0.0962	0.0015	-0.0072	-0.0002	0.0001
		SD	0.0286	0.0273	0.1339	0.1074	0.0598	0.0752	0.0095	0.0080
$t(3)$	Case1	Bias	-0.0019	0.0018	-0.0024	0.0015	-0.0060	0.0038	-0.0003	0.0001
		SD	0.0250	0.0245	0.0407	0.0358	0.0383	0.0347	0.0056	0.0057
	Case2	Bias	0.0003	-0.0007	0.0176	0.0011	-0.0082	-0.0018	-0.0002	-0.0001
		SD	0.0148	0.0129	0.0897	0.0483	0.0406	0.0385	0.0075	0.0067
	Case3	Bias	0.0007	-0.0009	0.0149	0.0131	0.0013	-0.0032	0.0003	-0.0003
		SD	0.0152	0.0139	0.2315	0.0889	0.0479	0.0662	0.0059	0.0058
	Case4	Bias	-0.0004	-0.0006	-0.0103	0.0141	0.0048	-0.0031	0.0001	0.0001
		SD	0.0122	0.0109	0.2143	0.0966	0.0572	0.0444	0.0063	0.0051
MN	Case1	Bias	-0.0004	0.0003	-0.0027	-0.0001	-0.0025	-0.0053	-0.0001	-0.0001
		SD	0.0263	0.0247	0.0498	0.0489	0.0423	0.0482	0.0052	0.0044
	Case2	Bias	-0.0016	0.0008	-0.0045	0.0172	-0.0027	-0.0013	-0.0007	0.0004
		SD	0.0171	0.0141	0.1186	0.0949	0.0510	0.0479	0.0063	0.0049
	Case3	Bias	-0.0005	-0.0003	-0.0113	0.0140	-0.0027	-0.0021	-0.0001	-0.0001
		SD	0.0160	0.0138	0.0973	0.0682	0.0505	0.0483	0.0065	0.0057
	Case4	Bias	-0.0001	-0.0005	0.0232	0.0119	-0.0012	-0.0028	-0.0002	-0.0001
		SD	0.0170	0.0209	0.1681	0.1512	0.0336	0.0535	0.0055	0.0046

Table 2. Bias and SD of four estimators under proposed MR procedure with $n = 400$.

ε	π		MNAR-MR		MAR-MR		CC-MR		FULL-MR	
			β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
$N(\mathbf{0}, \mathbf{1})$	Case1	Bias	0.0001	-0.0005	-0.0005	-0.0007	0.0052	-0.0012	-0.0002	0.0001
		SD	0.0597	0.0675	0.0544	0.0932	0.0937	0.0932	0.0087	0.0093
	Case2	Bias	-0.0007	0.0005	0.0418	0.0275	-0.0016	-0.0029	0.0001	0.0003
		SD	0.0064	0.0055	0.0875	0.0813	0.1004	0.1084	0.0042	0.0036
	Case3	Bias	0.0008	0.0005	0.0660	0.0870	-0.0018	-0.0331	-0.0004	0.0005
		SD	0.0094	0.0051	0.0989	0.1022	0.0625	0.0833	0.0039	0.0034
	Case4	Bias	-0.0013	-0.0064	0.0705	0.0924	0.0108	-0.0093	-0.0002	0.0002
		SD	0.0061	0.0085	0.1084	-0.0989	0.0873	0.0607	0.0034	0.0030
$t(3)$	Case1	Bias	0.0004	-0.0003	-0.0013	0.0003	0.0082	-0.0018	0.0001	-0.0001
		SD	0.0197	0.0176	0.0173	0.0165	0.0342	0.0331	0.0037	0.0032
	Case2	Bias	-0.0006	0.0003	0.0106	0.0070	-0.0013	-0.0019	0.0001	-0.0001
		SD	0.0073	0.0062	0.0523	0.0759	0.0397	0.0504	0.0035	0.0034
	Case3	Bias	-0.0001	-0.0002	0.0241	0.0187	-0.0054	-0.0030	-0.0001	-0.0001
		SD	0.0079	0.0066	0.0852	0.0708	0.0482	0.0583	0.0034	0.0031
	Case4	Bias	-0.0006	0.0002	0.0559	0.0106	-0.0056	-0.0048	-0.0001	0.0001
		SD	0.0058	0.0052	0.0982	0.0816	0.0611	0.0542	0.0034	0.0031
MN	Case1	Bias	0.0001	0.0002	0.0021	-0.0001	-0.0032	-0.0027	0.0003	-0.0003
		SD	0.0109	0.0148	0.0207	0.0165	0.0382	0.0434	0.0029	0.0027
	Case2	Bias	0.0005	-0.0007	0.0573	-0.0044	-0.0013	-0.0010	0.0001	-0.0002
		SD	0.0072	0.0062	0.0896	0.0876	0.0497	0.0480	0.0026	0.0024
	Case3	Bias	-0.0003	0.0001	0.0146	0.0037	-0.0033	-0.0017	-0.0002	0.0001
		SD	0.0043	0.0042	0.0187	0.0230	0.0104	0.0100	0.0027	0.0024
	Case4	Bias	-0.0001	0.0003	-0.0115	0.0106	0.0081	-0.0019	-0.0001	0.0001
		SD	0.0067	0.0062	0.1307	0.1195	0.0599	0.0477	0.0028	0.0026

Table 3. Bias and SD of four estimators under PLS method with $n = 200$.

ε	π		MNAR-PLS		MAR-PLS		CC-PLS		FULL-PLS	
			β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
$N(\mathbf{0}, \mathbf{1})$	Case1	Bias	-0.0040	0.0005	-0.0040	-0.0018	-0.006324	0.0026	-0.0003	-0.0001
		SD	0.0297	0.0275	0.0404	0.0382	0.0394	0.0317	0.0041	0.0040
	Case2	Bias	-0.0004	0.0004	-0.0994	0.0112	-0.0004	-0.0015	0.0002	0.0002
		SD	0.0096	0.0086	0.0764	0.0846	0.0208	0.0183	0.0039	0.0033
	Case3	Bias	0.0006	-0.0006	-0.0134	0.0150	-0.0019	-0.0011	-0.0003	-0.0002
		SD	0.0121	0.0111	0.1942	0.1140	0.0180	0.0149	0.0048	0.0042
	Case4	Bias	0.0003	-0.0002	0.0116	0.0609	0.0002	-0.0029	0.0002	-0.0002
		SD	0.0125	0.0106	0.1236	0.0877	0.0184	0.0153	0.0058	0.0046
$t(3)$	Case1	Bias	-0.0021	0.0021	0.0026	-0.0021	-0.0196	-0.0096	0.0008	-0.0004
		SD	0.0305	0.0286	0.0899	0.1050	0.1255	0.0581	0.0195	0.0220
	Case2	Bias	-0.0011	-0.0017	0.0684	0.1159	0.0090	-0.0104	-0.0007	0.0003
		SD	0.0444	0.0613	0.1062	0.1131	0.0921	0.0884	0.0158	0.0131
	Case3	Bias	-0.0085	-0.0017	-0.0207	0.0123	-0.0032	-0.0081	0.0014	-0.0011
		SD	0.0408	0.0324	0.2315	0.1011	0.0980	0.0762	0.0111	0.0104
	Case4	Bias	-0.0086	0.0037	-0.0101	0.0103	-0.0094	-0.0094	0.0004	-0.0005
		SD	0.0542	0.0447	0.2347	0.1103	0.0915	0.0967	0.0106	0.0093
MN	Case1	Bias	0.0013	-0.0018	-0.0026	0.0005	-0.0092	-0.0118	-0.0006	0.0004
		SD	0.0495	0.0397	0.1036	0.0989	0.1090	0.0961	0.0207	0.0181
	Case2	Bias	0.0019	-0.0065	-0.0156	0.0174	-0.0039	-0.0070	-0.0014	0.0006
		SD	0.0425	0.0381	0.2588	0.1277	0.1122	0.1108	0.0095	0.0077
	Case3	Bias	0.0020	-0.0043	0.0595	0.0194	-0.0996	0.0068	-0.0003	-0.0005
		SD	0.0591	0.0385	0.1131	0.0959	0.0981	0.1118	0.0114	0.0094
	Case4	Bias	-0.0010	-0.0055	0.0865	0.0246	-0.0031	0.0108	-0.0008	0.0007
		SD	0.0466	0.0547	0.1801	0.1546	0.1091	0.1001	0.0086	0.0075

Table 4. Bias and SD of four estimators under proposed PLS method with $n = 400$.

ε	π	MNAR-PLS		MAR-PLS		CC-PLS		FULL-PLS		
		β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	
$N(\mathbf{0}, \mathbf{1})$	Case1	Bias	-0.0005	-0.0001	0.0004	-0.0004	-0.0043	0.0017	-0.0001	-0.0001
		SD	0.0151	0.0121	0.0183	0.0162	0.0338	0.0339	0.0024	0.0020
	Case2	Bias	-0.0001	0.0008	0.0047	0.0024	-0.0017	-0.0006	0.0002	0.0001
		SD	0.0063	0.0054	0.0240	0.0220	0.0101	0.0849	0.0023	0.0019
	Case3	Bias	-0.0003	0.0005	0.0232	0.0054	-0.0014	-0.0086	-0.0001	0.0002
		SD	0.0074	0.0059	0.0622	0.0205	0.0127	0.0107	0.0028	0.0023
	Case4	Bias	-0.0002	0.0001	0.0741	0.0566	0.0003	-0.0016	0.0001	0.0001
		SD	0.0055	0.0049	0.0919	0.0821	0.0175	0.0167	0.0023	0.0020
$t(3)$	Case1	Bias	0.0014	0.0043	-0.0030	0.0042	0.0062	-0.0061	-0.0002	-0.0004
		SD	0.0239	0.0223	0.0287	0.0287	0.0945	0.0958	0.0116	0.0110
	Case2	Bias	-0.0016	0.0053	0.0286	0.0302	-0.0061	-0.0287	-0.0001	-0.0002
		SD	0.0281	0.0445	0.0944	0.0980	0.0261	0.0850	0.0085	0.0075
	Case3	Bias	0.0012	-0.0050	0.0412	0.0310	-0.0071	-0.0041	0.0004	-0.0005
		SD	0.0152	0.0140	0.1119	0.0951	0.0944	0.0834	0.0058	0.0052
	Case4	Bias	-0.0046	-0.0033	0.0780	0.0934	-0.0085	-0.0075	-0.0008	0.0005
		SD	0.0188	0.0129	0.1065	0.1088	0.0954	0.0734	0.0055	0.0063
MN	Case1	Bias	0.0043	-0.0011	-0.0029	-0.0039	-0.0053	-0.0431	0.0005	-0.0007
		SD	0.0399	0.0191	0.0942	0.0890	0.0405	0.0508	0.0099	0.0082
	Case2	Bias	-0.0024	0.0040	0.0682	0.0098	-0.0042	-0.0041	0.0007	-0.0009
		SD	0.0123	-0.0100	0.1341	0.0710	0.0954	0.0995	0.0045	0.0038
	Case3	Bias	0.0076	0.0082	-0.0212	0.0437	-0.0258	-0.0097	-0.0006	0.0004
		SD	0.0225	0.0216	0.1043	0.0990	0.0949	0.0626	0.0073	0.0074
	Case4	Bias	-0.0008	-0.0014	-0.0520	0.0494	-0.0075	-0.0068	-0.0002	0.0002
		SD	0.0107	0.0543	0.1525	0.1453	0.0955	0.0968	0.0040	0.0036

Simulation results of Tables 1–4 observe the following conclusions: (1) According to Tables 1 and 2, the proposed MNAR-MR estimation procedure behaves better than the CC-MR and MAR-MR in terms of giving smaller Bias and SD in most of the considered scenarios, even under ignorable missing Case 1. The CC-MR method only focuses on the fully observed samples and ignores the information of the covariates corresponding to missing response, which causes the biased estimators. As expected, the estimators of MAR-MR perform well under Case 1 when the missing mechanism is ignorable, while it has large bias under Cases 2–4 with nonignorable missing mechanism. (2) Comparing to the proposed MR procedure and PLS method, it is clear to see that Bias and SD of the estimators obtained by the PLS method are smaller than the proposed MR procedure when the error distribution is normal. However, when the model error yields to t distribution or mixed normal distribution, the estimators obtained by the proposed MR procedure are more accurate than those obtained by the PLS method. This indicates that the proposed MR estimation procedure is less sensitive to the heavy-tailed distribution or distribution of containing outliers. (3) As the sample size increases, there is an evident tendency that the performances of all estimation procedures are getting better and better in most of the considered scenarios.

Furthermore, to evaluate the performance of the proposed estimation procedure for nonparametric functions, we present the estimated curves of $\hat{\alpha}_1(\cdot)$ and $\hat{\alpha}_2(\cdot)$ with different methods, respectively. Specifically, Figure 1 shows the estimated curves of $\hat{\alpha}_1(\cdot)$ with the proposed MR procedure and the PLS method when the sample size n is 200 and the missing mechanism is Case 2. Similar to Figure 1, Figure 2 displays the estimated curves of $\hat{\alpha}_2(\cdot)$. Meanwhile, we consider the root of average square error (RASE) value of the estimation results for the nonparametric functions, where $RASE = \sqrt{\frac{1}{M} \sum_{j=1}^2 \sum_{i=1}^M [\hat{\alpha}_j(u_i) - \alpha_j(u_i)]^2}$ and $M = 200$ and u_i takes the equality points on the interval

(0, 1). Figure 3 shows the boxplots of the RASE for four estimators based on the proposed MR procedure when the sample size n is 200 and 400 with the model error $\varepsilon \sim t(3)$, respectively.

According to Figures 1–3, a few conclusions can be drawn as follows: (1) It is noted that the curve-fitting results of MNAR-MR are more effective than MNAR-PLS in that it is much closer to the real curve, regardless of whether the model error yields to normal distribution or mixed normal distribution. This indicates that the advantages of modal regression was apparent when data contains outliers. (2) It can be seen from Figure 3 that the RASE value of MAR-MR is closest to the RASE value of FULL-MR under Case 1, while it has large bias under the nonignorable missing mechanism. However, the RASE value of MNAR-MR is closest to the RASE value of FULL-MR with nonignorable missing mechanism with evidence that the proposed procedure has better performance on the nonparametric part. (3) The RASE value of four estimators based on the proposed MR procedure decreases with the increasing of sample size.

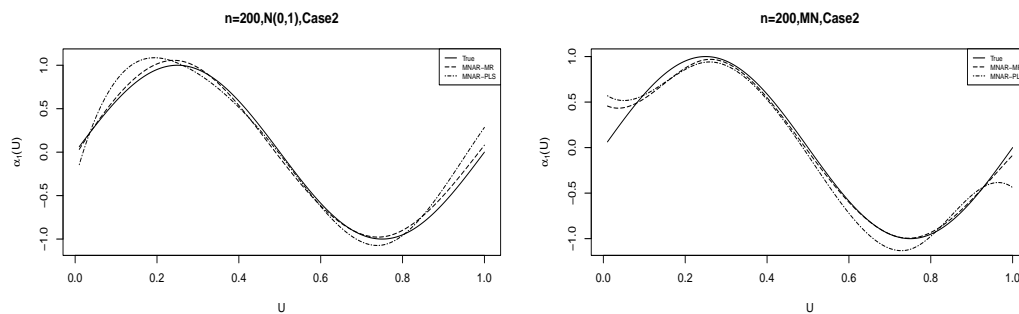


Figure 1. The estimated curves of the nonparametric function $\alpha_1(\cdot)$.

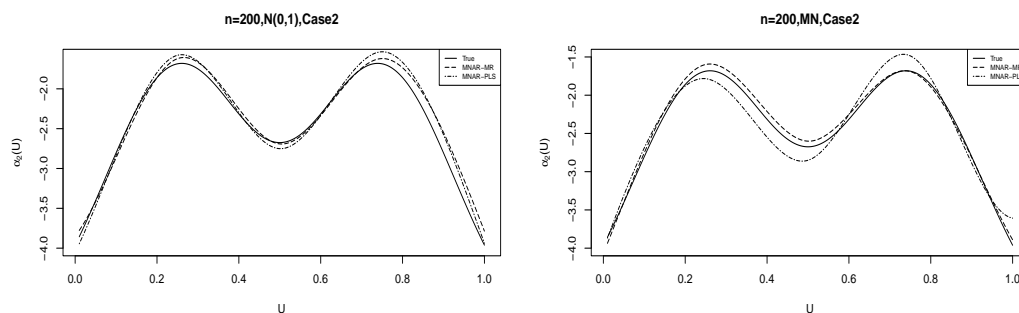


Figure 2. The estimated curves of the nonparametric function $\alpha_2(\cdot)$.

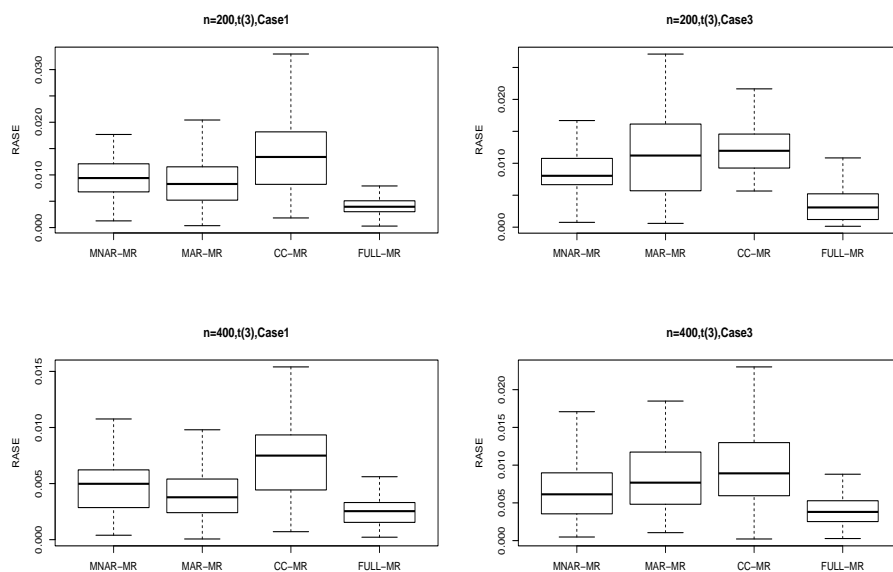


Figure 3. The boxplots of the RASE for the nonparametric functions.

6. A real data analysis

In this section, we illustrate the proposed method using HIV-CD4 data, which is collected on 2139 HIV positive patients enrolled in AIDS Clinical Trials Group Protocol 175 (ACTG 175) (Hammer et al. [24]). In this data, the patients were randomized into four groups receiving their respective antiretroviral treatment regimen: (1) Zidovudine or ZDV with 532 subjects, (2) didanosine or ddi with 522 subjects, (3) ZDV + ddi with 524 subjects and (4) ZDV + zalcitabine with 561 subjects. Let response Y be the CD4 count at 96 ± 5 weeks under each antiretroviral treatment regimen. There are six continuous baseline covariates: Age (U), weight (Z_1), CD4 cell counts at baseline and 20 ± 5 weeks (Z_2 and Z_3) and CD8 cell counts at baseline and 20 ± 5 weeks (Z_4 and Z_5). Due to adverse events, death, dropout and some other reasons, Y has missing values, but the values of covariates are fully observed. Specifically, the proportions of missing data under Y in four regimens are about 39.66%, 36.21%, 35.69% and 37.43%, respectively.

Typically, a sharp decline in CD4 cell count is indicative of disease progression, and patients with low CD4 cell count are more likely to drop out from the scheduled study visits compared to patients with normal CD4 (Yuan and Yin [25]). Therefore, the missing values of the CD4 count Y at 96 ± 5 weeks is likely related to the CD4 value count itself, and Y has nonignorable missing values. Following Zhang and Wang [26], given the six baseline covariates and CD4 count at 96 ± 5 weeks, the missing data propensity does not depend on age and weight. That is, to apply the proposed method, we use the age and weight as the instrument variable S .

The scatterplot in Figure 4 indicates that there is no obvious linear relationship between Y and the covariates. According to [24], it is generally believed that the influence of CD4 cell counts and CD8 cell counts on Y may not be immediate. There is a lagging effect, which is similar to the impact of acceleration on displacement. Based on the above discussions, we establish the following varying-

coefficient partially nonlinear model:

$$Y = \alpha(U) + \exp(Z_1\beta_1 + Z_2\beta_2 + Z_3\beta_3 + Z_4\beta_4 + Z_5\beta_5) + \varepsilon. \quad (6.1)$$

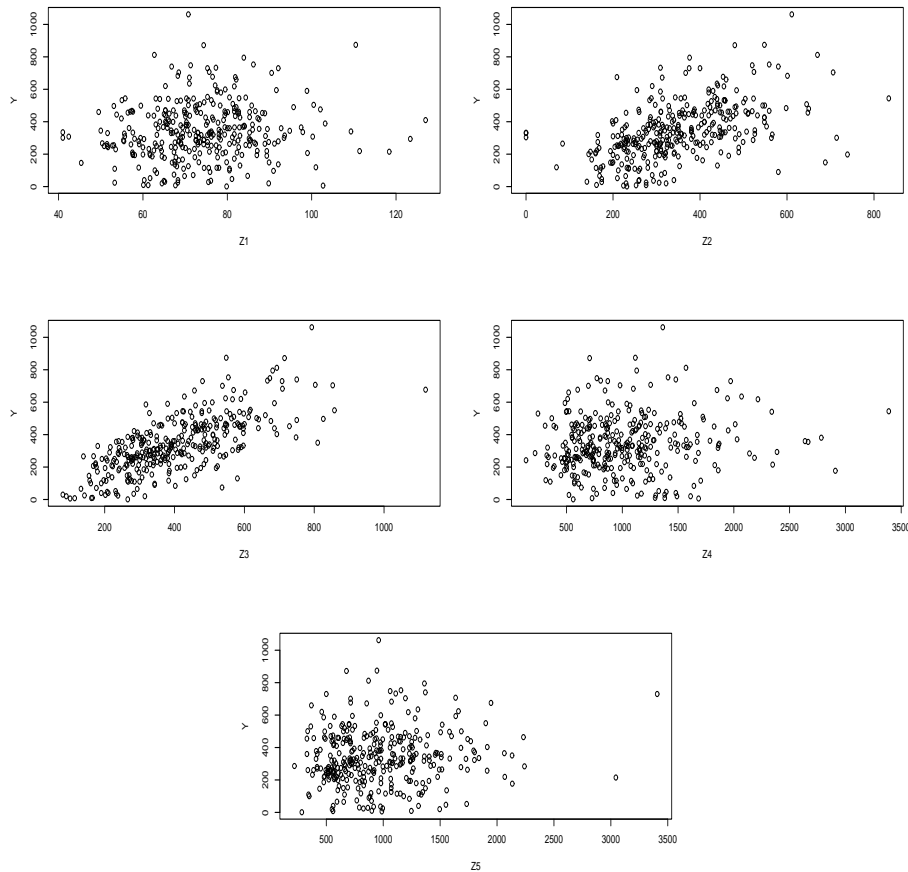


Figure 4. The scatterplot of CD4 count at 96 ± 5 weeks with baseline covariates.

The estimators of parameter based on the MNAR-MR procedure and fitting curves for the nonparametric function are given in Table 5 and Figure 5 (From left to right, top to bottom, they are regimen (1) to regimen (4)).

Table 5. The MNAR-MR estimators for HIV-CD4 data.

	β_1	β_2	β_3	β_4	β_5
Regimen(1)					
Estimator	0.025479	0.122334	0.324853	-0.013147	-0.013933
Regimen(2)					
Estimator	0.032869	0.160647	0.255346	0.004643	-0.065893
Regimen(3)					
Estimator	0.077097	0.284075	0.465495	0.032198	-0.073453
Regimen(4)					
Estimator	0.032597	0.199342	0.417049	-0.011851	-0.037176

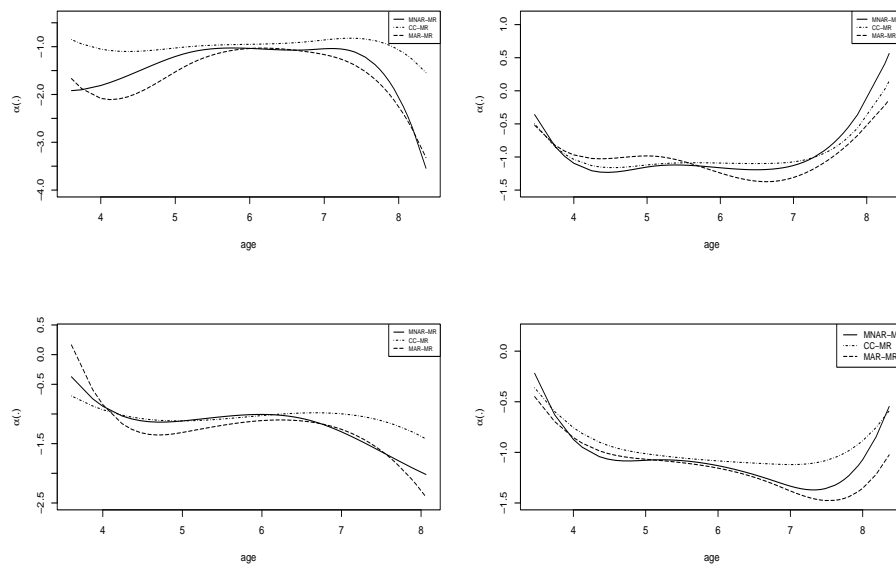


Figure 5. The fitting curves of coefficient function for HIV-CD4 data.

It can be seen that: (1) For all the regimens, the estimated coefficients of Z_4 and Z_5 are close to zero, while the estimated coefficients of Z_2 and Z_3 are relatively large. This may indicate that CD8 cell counts at baseline and 20 ± 5 weeks have no significant impact on the CD4 count at 96 ± 5 weeks, but CD4 cell counts at baseline and 20 ± 5 weeks appear to affect the CD4 count at 96 ± 5 weeks. (2) Under four regimens, the coefficients of Z_2 and Z_3 under regimen (3) are greater than the other regimens, which may indicate that regimen (3) gives smaller relative hazard ratios than other regimens. (3) The estimation curve for the nonparametric function is different under the assumptions of MNAR, MAR and CC, which suggests that the assumption of a nonignorable missing propensity is reasonable, and also that the effect of age on the response differs under four scenarios. In the future study, the method of distributed fault estimation based on the fault estimation observer for multi-agent systems [27, 28, 29] may be applied to improve the estimation performance.

7. Conclusions

This article studied the statistical inference for the varying-coefficient partially nonlinear model with a nonignorable missing response. A robust estimation procedure was proposed to estimate the parametric and nonparametric parts separately based on modal regression. Mainly due to the identifiability of the nonresponse propensity, we considered a semiparametric propensity model and applied the GMM approach, making use of an instrument variable to identify unknown parameters in the propensity. Further, the coefficient function was approximated by the B-spline function to simplify the model structure, and the inverse probability weighted technique and modal regression were applied to construct the efficient MNAR-MR estimators. Compared to mean regression, modal regression is robust against outliers or heavy-tail error distribution, and it performs no worse than the least-square-based method for normal error distribution. Inverse probability weighted technique can increase the estimation efficiency by the complete-case-data procedure in which the missing subject is

excluded. Under some mild conditions, the asymptotic properties of the proposed procedure were established. Some simulation studies and a real example were carried out to demonstrate that the proposed procedure has good performance in the finite samples. In our assumptions, nonlinear function $g(\cdot)$ is pre-specified. A more valuable model is that of relaxing $g(\cdot)$, which is unknown. It would be an interesting topic in the future research.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work is supported by the Natural Science Basic Research Plan in Shaanxi Province of China(2022JM-027).

Conflict of interest

The authors declare no conflict of interest.

References

1. R. Li, L. Nie, Efficient statistical inference procedures for partially nonlinear models and their applications, *Biometrics*, **64** (2008), 904–911. <https://doi.org/10.1111/J.1541-0420.2007.00937.X>
2. N. S. Tang, P. Y. Zhao, Empirical likelihood semiparametric nonlinear regression analysis for longitudinal data with responses missing at random, *Ann. I. Stat. Math.*, **65** (2013), 639–665. <https://doi.org/10.1007/s10463-012-0387-4>
3. T. Z. Li, C. L. Mei, Estimation and inference for varying coefficient partially nonlinear models, *J. Stat. Plan Infer.*, **143** (2013), 2023–2037. <https://doi.org/10.1016/j.jspi.2013.05.011>
4. X. S. Zhou, P. X. Zhao, X. L. Wang, Empirical likelihood inferences for varying coefficient partially nonlinear models, *J. Appl. Stat.*, **44** (2017), 474–492. <https://doi.org/10.1080/02664763.2016.1177496>
5. J. Yang, H. Yang, Smooth-threshold estimating equations for varying coefficient partially nonlinear models based on orthogonality-projection method, *J. Comput. Appl. Math.*, **302** (2016), 24–37. <https://doi.org/10.1016/j.cam.2016.01.038>
6. Y. T. Xiao, Z. S. Chen, Bias-corrected estimations in varying-coefficient partially nonlinear models with measurement error in the nonparametric part, *J. Appl. Stat.*, **45** (2018), 586–603. <https://doi.org/10.1080/02664763.2017.1288201>
7. X. R. Tang, P. X. Zhao, Y. P. Yang, W. M. Yang, Adjusted empirical likelihood inferences for varying coefficient partially non linear models with endogenous covariates, *Commun. Stat.-Theor. M.*, **51** (2022), 953–973. <https://doi.org/10.1080/03610926.2020.1747078>
8. Y. L. Jiang, Q. H. Ji, B. J. Xie, Robust estimation for the varying coefficient partially nonlinear models, *J. Comput. Appl. Math.*, **326** (2017), 31–43. <https://doi.org/10.1016/j.cam.2017.04.028>

9. J. Yang, F. Lu, H. Yang, Quantile regression for robust inference on varying coefficient partially nonlinear models, *J. Korean. Stat. Soc.*, **47** (2018), 172–184. <https://doi.org/10.1016/j.jkss.2017.12.002>
10. Y. T. Xiao, L. L. Liang, Robust estimation and variable selection for varying-coefficient partially nonlinear models based on modal regression, *J. Korean. Stat. Soc.*, **51** (2020), 692–715. <https://doi.org/10.1007/s42952-021-00158-w>
11. R. J. A. Little, D. B. Rubin, *Statistical analysis with missing data*, New York: Wiley, 2002.
12. X. L. Wang, P. X. Zhao, H. Y. Du, Statistical inferences for varying coefficient partially non linear model with missing covariates, *Commun. Stat-Theor. M*, **50** (2021), 2599–2618. <https://doi.org/10.1080/03610926.2019.1674870>
13. L. Q. Xia, X. L. Wang, P. X. Zhao, Y. Q. Song, Empirical likelihood for varying coefficient partially nonlinear model with missing responses, *AIMS Math.*, **6** (2021), 7125–7152. <https://doi.org/10.3934/MATH.2021418>
14. J. K. Kim, C. L. Yu, A semiparametric estimation of mean functionals with nonignorable missing data, *J. Am. Stat. Assoc.*, **106** (2011), 157–165. <https://doi.org/10.1198/jasa.2011.tm10104>
15. N. S. Tang, P. Y. Zhao, H. T. Zhu, Empirical likelihood for estimating equations with nonignorably missing data, *Stat. Sin.*, **24** (2014), 723–747. <https://doi.org/10.5705/ss.2012.254>
16. S. Wang, J. Shao, J. K. Kim, An instrumental variable approach for identification and estimation with nonignorable nonresponse, *Stat. Sin.*, **24** (2014), 1097–1116. <https://dx.doi.org/10.5705/ss.2012.074>
17. J. Shao, L. Wang, Semiparametric inverse propensity weighting for nonignorable missing data, *Biometrika*, **103** (2016), 175–187. <https://doi.org/10.1093/biomet/asv071>
18. W. X. Yao, B. G. Lindsay, R. Z. Li, Local modal regression, *J. Nonparametr. Stat.*, **24** (2012), 647–663. <https://doi.org/10.1080/10485252.2012.678848>
19. R. Q. Zhang, W. H. Zhao, J. C. Liu, Robust estimation and variable selection for semiparametric partially linear varying coefficient model based on modal regression, *J. Nonparametr. Stat.*, **25** (2013), 523–544. <https://doi.org/10.1080/10485252.2013.772179>
20. W. H. Zhao, R. Q. Zhang, J. C. Liu, Y. Z. Lv, Robust and efficient variable selection for semiparametric partially linear varying coefficient model based on modal regression, *Ann. Inst. Stat. Math.*, **66** (2014), 165–191. <https://doi.org/10.1007/s10463-013-0410-4>
21. L. L. Schumaker, *Spline functions: Basic theory*, New York: John Wiley and Son, 1981.
22. L. Wang, P. Y. Zhao, J. Shao, Dimension-reduced semiparametric estimation of distribution functions and quantiles with nonignorable nonresponse, *Comput. Stat. Data An.*, **156** (2021), 107142. <https://doi.org/10.1016/j.csda.2020.107142>
23. J. Li, S. Ray, B. G. Lindsay, A nonparametric statistical approach to clustering via mode identification, *J. Mach. Learn. Res.*, **8** (2007), 1687–1723.
24. S. M. Hammer, D. A. Katzenstein, M. D. Hughes, H. Gundacker, R. T. Schooley, R. H. Haubrich, et al., A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic millimeter, *N. Engl. J. Med.*, **335** (1996), 1081–1090. <https://doi.org/10.1056/NEJM199610103351501>

25. Y. Yuan, G. S. Yin, Bayesian quantile regression for longitudinal studies with nonignorable missing data, *Biometrics*, **66** (2010), 105–114. <https://doi.org/10.1111/j.1541-0420.2009.01269.x>
26. T. Zhang, L. Wang, Smoothed empirical likelihood inference and variable selection for quantile regression with nonignorable missing response, *Comput. Stat. Data An.*, **144** (2020), 106888. <https://doi.org/10.1016/j.scsda.2019.106888>
27. J. Han, X. H. Liu, X. J. Wei, H. F. Zhang, Adjustable dimension descriptor observer based fault estimation for switched nonlinear systems with partially unknown nonlinear dynamics, *Nonlinear Anal.-Hybri.*, **42** (2021), 101083. <https://doi.org/10.1016/j.nahs.2021.101083>
28. J. Han, X. H. Liu, X. J. Wei, S. X. Sun, A dynamic proportional-integral observer-based nonlinear fault-tolerant controller design for nonlinear system with partially unknown dynamic, *IEEE T. Syst. Man Cy.-S.*, **52** (2022), 5092–5104. <https://doi.org/10.1109/TSMC.2021.3114326>
29. J. Han, X. H. Liu, X. J. Xie, X. J. Wei, Dynamic output feedback fault tolerant control for switched fuzzy systems with fast time varying and unbounded faults, *IEEE. T. Fuzzy Syst.*, **31** (2023), 3185–3196. <https://doi.org/10.1109/TFUZZ.2023.3246061>
30. J. Lv, H. Yang, C. H. Guo, Variable selection in partially linear additive models for modal regression, *Commun. Stat.-Simul. C.*, **46** (2017), 5646–5665. <https://doi.org/10.1080/03610918.2016.1171346>

Appendix: Proofs of theorems

To facilitate the presentation, we denote by

$$\eta_i(\boldsymbol{\beta}) = \delta_i / \hat{\pi}(Y_i, \mathbf{T}_i, \hat{\boldsymbol{\zeta}}) \phi'_h[\mathbf{P}_i(\mathbf{Y} - g(\mathbf{Z}, \boldsymbol{\beta}))] \mathbf{P}_i g'(\boldsymbol{\beta}), \quad (7.1)$$

then the modal estimator $\hat{\boldsymbol{\beta}}$ satisfies $\sum_{i=1}^n \eta_i(\hat{\boldsymbol{\beta}}) = 0$, which is obtained by maximizing (2.13).

Lemma 1. *Suppose Assumptions (A1)–(A10) hold. As $n \rightarrow \infty$, we have*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i(\boldsymbol{\beta}_0) \xrightarrow{d} N(0, A).$$

Proof. Let $R_j(U) = \alpha_j(U) - \mathbf{B}^T(U)\boldsymbol{\gamma}_{0j}$, $\mathbf{R}(U) = (R_1(U), R_2(U), \dots, R_p(U))^T$. Model (2.4) can be reexpressed as

$$\mathbf{Y} - g(\mathbf{Z}, \boldsymbol{\beta}) = \mathbf{W}\boldsymbol{\gamma} + \mathbf{X}\mathbf{R}(U) + \boldsymbol{\varepsilon}. \quad (7.2)$$

For fixed $\boldsymbol{\beta}$, model (7.2) becomes

$$\mathbf{P}\mathbf{Y} = \mathbf{X}\mathbf{R}(U) + \mathbf{P}g(\mathbf{Z}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}. \quad (7.3)$$

Notice that

$$\phi'_h[\mathbf{P}_i(\mathbf{Y} - g(\mathbf{Z}, \boldsymbol{\beta}))] = \phi'_h[\mathbf{X}_i^T \mathbf{R}(U_i) + \boldsymbol{\varepsilon}_i] = \phi'_h(\boldsymbol{\varepsilon}_i) + \phi''_h(\boldsymbol{\varepsilon}_i) \mathbf{X}_i^T \mathbf{R}(U_i) + o_p(1). \quad (7.4)$$

By simple calculation,

$$\begin{aligned}
 \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i(\beta_0) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{\hat{\pi}(Y_i, \mathbf{T}_i, \hat{\zeta})} \phi'_h[\mathbf{P}_i(\mathbf{Y} - g(\mathbf{Z}, \beta_0))] \mathbf{P}_i g'(\beta_0) \\
 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{\pi(Y_i, \mathbf{T}_i, \zeta_0)} \phi'_h[\mathbf{P}_i(\mathbf{Y} - g(\mathbf{Z}, \beta_0))] \mathbf{P}_i g'(\beta_0) \\
 &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\frac{\delta_i}{\hat{\pi}(Y_i, \mathbf{T}_i, \hat{\zeta})} - \frac{\delta_i}{\pi(Y_i, \mathbf{T}_i, \zeta_0)} \right] \phi'_h[\mathbf{P}_i(\mathbf{Y} - g(\mathbf{Z}, \beta_0))] \mathbf{P}_i g'(\beta_0) \\
 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n I_{1i} + \frac{1}{\sqrt{n}} \sum_{i=1}^n I_{2i}. \tag{7.5}
 \end{aligned}$$

For I_{1i} , by (7.4) and the fact $\|R(U) = O(K^{-r})\|$, it can be shown that $E(I_{1i}) = 0$ and $\text{Var}(I_{1i}) = E(I_{1i}^2) = E[\pi(Y_i, \mathbf{T}_i, \zeta_0)^{-1} \phi'_h(\varepsilon_i)^2 (\mathbf{P}_i g'(\beta_0))^T \mathbf{P}_i g'(\beta_0)]$.

Therefore,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n I_{1i} \xrightarrow{d} N(0, B) \tag{7.6}$$

is obtained, where $B = E[\pi(Y, \mathbf{T}, \zeta_0)^{-1} G(x, z, u, h) g'(z, \beta) g'(z, \beta)^T]$.

In terms of I_{2i} , noticing that

$$\begin{aligned}
 \frac{1}{\sqrt{n}} \sum_{i=1}^n I_{2i} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\frac{\delta_i}{\hat{\pi}(Y_i, \mathbf{T}_i, \hat{\zeta})} - \frac{\delta_i}{\pi(Y_i, \mathbf{T}_i, \zeta_0)} \right] \phi'_h[\mathbf{P}_i(\mathbf{Y} - g(\mathbf{Z}, \beta_0))] \mathbf{P}_i g'(\beta_0) \\
 &= \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial \pi(Y_i, \mathbf{T}_i, \zeta_0) / \partial \zeta}{\pi(Y_i, \mathbf{T}_i, \zeta_0)^2} \right] \delta_i \phi'_h[\mathbf{P}_i(\mathbf{Y} - g(\mathbf{Z}, \beta_0))] \mathbf{P}_i g'(\beta_0) \sqrt{n}(\hat{\zeta} - \zeta_0) + o_p(n^{-1/2}) \tag{7.7}
 \end{aligned}$$

it is easy to show that

$$\frac{1}{n} \sum_{i=1}^n \left[\frac{\partial \pi(Y_i, \mathbf{T}_i, \zeta_0) / \partial \zeta}{\pi(Y_i, \mathbf{T}_i, \zeta_0)^2} \right] \delta_i \phi'_h[\mathbf{P}_i(\mathbf{Y} - g(\mathbf{Z}, \beta_0))] \mathbf{P}_i g'(\beta_0) \xrightarrow{p} H,$$

where H is defined in Theorem 1.

According to Wang et al. [22], we have $\hat{\zeta} - \zeta_0 = O_p(n^{-1/2})$ and $n^{-1/2}(\hat{\zeta} - \zeta_0) \xrightarrow{d} N(0, \Pi)$. Then, we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n I_{2i} \xrightarrow{d} N(0, H\Pi H^T). \tag{7.8}$$

By (7.5), (7.6) and (7.8), the result is proved. \square

Lemma 2. Suppose Assumptions (A1)–(A10) hold. As $n \rightarrow \infty$, we have

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial \eta_i(\beta_0)}{\partial \beta} \xrightarrow{p} \Sigma.$$

Proof.

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \frac{\partial \eta_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} &= \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}} \left\{ \frac{\delta_i}{\pi(Y_i, \mathbf{T}_i, \zeta_0)} \phi'_h[\mathbf{P}_i(\mathbf{Y} - g(\mathbf{Z}, \boldsymbol{\beta}_0))] \mathbf{P}_i g'(\boldsymbol{\beta}_0) \right\} \\ &+ \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\beta}} \left\{ \left[\frac{\delta_i}{\hat{\pi}(Y_i, \mathbf{T}_i, \hat{\zeta})} - \frac{\delta_i}{\pi(Y_i, \mathbf{T}_i, \zeta_0)} \right] \phi'_h[\mathbf{P}_i(\mathbf{Y} - g(\mathbf{Z}, \boldsymbol{\beta}_0))] \mathbf{P}_i g'(\boldsymbol{\beta}_0) \right\} \\ &= T_1 + T_2 \end{aligned} \quad (7.9)$$

it is easy to obtain $E(T_1) = E(\phi''_h(\varepsilon_i)(\mathbf{P}_i g'(\boldsymbol{\beta}_0))^T \mathbf{P}_i g'(\boldsymbol{\beta}_0))$. By the law of large numbers, we have $T_1 \xrightarrow{p} \Sigma$, where $\Sigma = E(F(x, z, u, h)g'(z, \boldsymbol{\beta})g'(z, \boldsymbol{\beta})^T)$.

For T_2 , applying the Taylor expansion, we get

$$T_2 = \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial \pi(Y_i, \mathbf{T}_i, \zeta_0) / \partial \zeta}{\pi(Y_i, \mathbf{T}_i, \zeta_0)^2} \right] \frac{\partial}{\partial \boldsymbol{\beta}} \{ \delta_i \phi'_h[\mathbf{P}_i(\mathbf{Y} - g(\mathbf{Z}, \boldsymbol{\beta}_0))] \mathbf{P}_i g'(\boldsymbol{\beta}_0) \} (\hat{\zeta} - \zeta_0) + o_p(1). \quad (7.10)$$

By the law of large numbers, we have

$$\begin{aligned} &\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial \pi(Y_i, \mathbf{T}_i, \zeta_0) / \partial \zeta}{\pi(Y_i, \mathbf{T}_i, \zeta_0)^2} \right] \frac{\partial}{\partial \boldsymbol{\beta}} \{ \delta_i \phi'_h[\mathbf{P}_i(\mathbf{Y} - g(\mathbf{Z}, \boldsymbol{\beta}_0))] \mathbf{P}_i g'(\boldsymbol{\beta}_0) \} \\ &= E \left\{ \frac{\partial \pi(Y_i, \mathbf{T}_i, \zeta_0) / \partial \zeta}{\pi(Y_i, \mathbf{T}_i, \zeta_0)} \frac{\partial}{\partial \boldsymbol{\beta}} \{ \phi'_h[\mathbf{P}_i(\mathbf{Y} - g(\mathbf{Z}, \boldsymbol{\beta}_0))] \mathbf{P}_i g'(\boldsymbol{\beta}_0) \} \right\}. \end{aligned} \quad (7.11)$$

As $\hat{\zeta}$ is the consistent estimator of ζ_0 , we conclude that $T_2 = o_p(1)$. This yields the result that

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial \eta_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \xrightarrow{p} \Sigma.$$

□

Proof of Theorem 1. Since the modal estimators $\hat{\boldsymbol{\beta}}$ satisfies $\sum_{i=1}^n \eta_i(\hat{\boldsymbol{\beta}}) = 0$, according to the Taylor expansion, we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i(\hat{\boldsymbol{\beta}}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i(\boldsymbol{\beta}_0) + \frac{1}{n} \sum_{i=1}^n \frac{\partial \eta_i(\boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}} \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0). \quad (7.12)$$

From Lemmas 1, 2 and (7.12), we have $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(0, \Sigma^{-1} A \Sigma^{-1})$. □

Proof of Theorem 2. Let $\delta_n = n^{-r/(2r+1)}$ and $\mathbf{v} = (v_1^T, \dots, v_p^T)^T$ with PL dimension. Define $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0 + \delta_n \mathbf{v}$.

We show that, for any given $\varepsilon > 0$, there exists a large enough constant C such that

$$P \left\{ \inf_{\|\mathbf{v}\|=C} \hat{Q}(\boldsymbol{\gamma}) > \hat{Q}(\boldsymbol{\gamma}_0) \right\} \geq 1 - \varepsilon, \quad (7.13)$$

where $\hat{Q}(\boldsymbol{\gamma})$ is defined in (2.15). Let $\varphi(\boldsymbol{\gamma}) = \hat{Q}(\boldsymbol{\gamma}) - \hat{Q}(\boldsymbol{\gamma}_0)$. By Taylor expansion we have

$$\varphi(\boldsymbol{\gamma}) = \delta_n \sum_{i=1}^n \phi'_h(\varepsilon_i + \mathbf{X}_i^T R(U_i) + g'(\mathbf{Z}_i, \boldsymbol{\beta})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})) \hat{\Delta}_i \mathbf{W}_i^T \mathbf{v}$$

$$\begin{aligned}
& + \delta_n^2 \sum_{i=1}^n \phi_h''(\varepsilon_i + \mathbf{X}_i^T R(U_i) + g'(\mathbf{Z}_i, \boldsymbol{\beta})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})) (\hat{\Delta}_i \mathbf{W}_i^T v)^2 \\
& + \delta_n^3 \sum_{i=1}^n \phi_h'''(\xi_i) (\hat{\Delta}_i \mathbf{W}_i^T v)^3 \\
& \triangleq I_1 + I_2 + I_3,
\end{aligned}$$

where ξ_i lies in $Y_i^* - \mathbf{W}_i^T \boldsymbol{\gamma}_0$ and $Y_i^* - \mathbf{W}_i^T \boldsymbol{\gamma}$. Then, for I_1 , using Taylor expansion, we obtain that

$$\begin{aligned}
I_1 & = \delta_n \sum_{i=1}^n \{ \phi_h'(\varepsilon_i) + \phi_h''(\varepsilon_i) \mathbf{X}_i^T R(U_i) + \phi_h''(\varepsilon_i) g'(\mathbf{Z}_i, \boldsymbol{\beta})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \} \hat{\Delta}_i \mathbf{W}_i^T v \\
& + \delta_n \sum_{i=1}^n \{ \phi_h'''(\zeta_i) [\mathbf{X}_i^T R(U_i) + g'(\mathbf{Z}_i, \boldsymbol{\beta})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})]^2 \} \hat{\Delta}_i \mathbf{W}_i^T v,
\end{aligned}$$

where ζ_i lies in ε_i and $\varepsilon_i + \mathbf{X}_i^T R(U_i) + g'(\mathbf{Z}_i, \boldsymbol{\beta})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})$.

By Assumption (A6), (A8) and some calculation results, we have $I_1 = O_p(nK^{-r} \delta_n \|v\|) = O_p(n\delta_n^2 \|v\|)$.

Similarly, we can prove that $I_2 = O_p(n\delta_n^2 \|v\|^2)$ and $I_3 = O_p(n\delta_n^3 \|v\|^3)$. Hence, by choosing a sufficiently large C , I_2 dominates I_1 uniformly $\|v\| = C$. Since $\delta_n \rightarrow 0$, it follows that $\delta_n \|v\| \rightarrow 0$ with $\|v\| = C$, which leads to $I_3 = O_p(I_2)$. Hence, (7.13) holds. There exists local maximizers $\hat{\boldsymbol{\gamma}}$, and we have

$$\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0\| = O_p(\delta_n) = O_p(n^{-\frac{r}{2r+1}}). \quad (7.14)$$

Similar to the proof of Theorem 2.1(b) in [30], we can obtain that

$$\|\hat{\alpha}_k(\cdot) - \alpha_{0k}(\cdot)\| = O_p(n^{-\frac{r}{2r+1}}), k = 1, \dots, p. \quad (7.15)$$

This completes the Theorem 2. □



© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)