Kriging Regression Imputation Method to Semiparametric Model with Missing Data

MINGXING ZHANG
Guizhou University of Fiance and Economics
School of Mathematics and Statistics
Luchongguan Road 276, Guiyang
CHINA
zmxgraduate@163.com

ZIXIN LIU
Guizhou University of Fiance and Economics
School of Mathematics and Statistics
Luchongguan Road 276, Guiyang
CHINA
xinxin905@163.com

Abstract: This paper investigates a class of estimation problems of the semiparametric model with missing data. In order to overcome the robust defect of traditional complete data estimation method and regression imputation estimation technique, we propose a modified imputation estimation approach called Kriging-regression imputation. Compared with previous method used in the references cited therein , the new proposed method not only makes more use of the data information, but also has better robustness. Model estimation and asymptotic distribution of the estimators are also derived theoretically. In order to improve the robustness, LASSO technique is further introduced into Kriging-regression imputation. Numerical experiment is also provided to show the effectiveness and superiority of our method.

Key–Words: semiparameter model, data missing, imputation techniques, asymptotic normality, consistency.

1 Introduction

With the rapid development of computing techniques, statistic inference theories of parametric and nonparametric regression model have gradually become mature, See the work of Conover [1], Green and Silverman [2], Midi and Mohammed [3], Pinto et al. [4], Adeogun [5], Fan and Gijbels [6] and Mood [7], among others. In recent three decades, semiparametric model has attained more and more popularity in various areas because of parametric model is prone to subjective error as well as nonparametric method can not avoid the curse of dimensionality. Similarly, semiparametric model has various forms, such as additive model, partially linear regression model, varying-coefficient model and so on. For example, Hastie and Tibshirani [8] have received the corresponding fitting method of the additive model. As a useful extension of partially linear model and varying-coefficient model, semiparametric partially linear vary-coefficient model has widely application.

Consider the semiparametric partially linear varying-coefficient model as follows:

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

where Y is response variable, and (X^T, Z^T, U) is the associated covariates. For simplicity, we assume U is univariate. ε is an independent random error with $E(\varepsilon|X, Z, U) = 0$ and $Var(\varepsilon|X, Z, U) = \delta^2$. $\beta = (\beta_1, \ldots \beta_q)^T$ is a q-dimensional vector of unknown parametric component, $\alpha(U) = (\alpha_1(U), \ldots \alpha_p(U))^T$ is a p-dimensional vector of unknown coefficient function.

Obviously, when $Z = 0$, model (1) reduces to varying-coefficient model, which has been widely studied in the literature, see the work of Cai et al. [9], Chen and Tsay [10], Hastile and Tibshirani [11], Huang et al. [12], Fan and Zhang [13], Xia and Li [14], Hoover et al.[15], Hu and Xia [16] and Huang et al.[17], among others. When $p = 1$ and $Z = 1$ model (1) becomes partially linear regression model, which was proposed by Engle et al. [18] when they researched the influence of weather on electricity demand. A series of literature (Chen [19], Yatchew [20], Liang [21], Speckman [22], Liang et al. [23]) regarding partially linear regression model have provided corresponding statistics inference.

Recently, model (1) has been widely studied by Fan and Huang [24], Zhou and You [25], Wei and Wu [26], Zhang and Lee [27], You and Chen [28] and so on. In [28], You and Chen studied the estimation of partially linear varying-coefficient model under the circumstance that some covariates were measured with additive errors. In [24], Fan and Huang proposed a profile least squares technique for estimating parametric component and put forth the generalized likelihood ratio test for the testing problem. Furthermore, they proved asymptotic normality of the profile least squares estimator and demonstrated the generalized likelihood ratio statistics followed an asymptotically chi-squared distribution under the null hypothesis.

It is worth pointing out that, in practice, data may often not be available completely because of some inevitable factors. This means that, in actual statistical analysis, response variable Y may be missing. As is well known that, data missing can cause certain influence for the estimation accuracy of parameter component β and function $\alpha(U)$. In this case, a commonly used technique is to introduce a new variable δ. If $\delta = 0$, means that Y is missing, and $\delta = 1$ otherwise. If Y is missing at random, δ and Y are conditionally independent, then we have

$$P(\delta = 1|Y, X, Z, U) = P(\delta = 1|X, Z, U).$$

Due to the importance and practicability of the missing data estimation, semiparametric partially linear varying-coefficient model with missing response has attracted many authors' attention, such as Wei [29], Little and Rubin [30], Chu and Cheng [31], Wang el at. [32] and so on. The simplest method for dealing with the missing data is to delete the missing data, in other words, we just adopt the data when $\delta = 1$. This technique is so called the method of complete-case data. However, deleting the missing data means the loss of data information. In order to better utilize the data information, Chu and Cheng [31] adopted the techniques of regression imputation. The main idea of regression imputation is to use the simplest local linear smoother to impute a prediction value for the missing Y at each X with $\delta = 0$. For the method of complete-case data, it has the advantage in statistical computation, but it can not make full use of data information, so that it can not capture the relation between the responses and their associated covariates well. Compared with complete-case data method, regression imputation approach has the better estimation efficiency, thus it has attracted more and more researchers' attention.

However, the traditional classical regression imputation approach dose not consider the factor of response variable. Thus it is not an excellent method. Another point needs to be pointed out that, traditional regression imputation approach utilizes the ordinary least squares estimation to replace the missing response values. As is well known that, ordinary least squares estimator has poor robustness. If there exists abnormal data in independent variables, the imputed value of response variable Y must be far away from the true value. This naturally leads to that the parameter estimator has poor robustness.

To inherit the advantage of regression imputation approach, and overcome the defects of the two aspects mentioned in the above discussion, a natural idea is to introduce Kriging imputation and Lasso approach into traditional regression imputation technique. Since regression imputation can effectively utilize the information of independent variable, and Kriging imputation method can effectively utilize the information of response variable, thus, the combination between regression imputation and Kriging imputation can sufficiently make use of data information and improve the efficiency of estimator. Additionally, noticed the effect of Lasso method to eliminate the cumulative effect in model selection and imputation estimator. In this paper, we further derive some improved results by using least absolute shrinkage and selection operator technique. In order to show the effectiveness and superiority of our technique proposed in this paper, one numerical simulation is also provided.

The rest of the paper is arranged as follows. We will introduce the complete-case data method and classical regression imputation approach of handling the semiparametric model with missing responses respectively in section 2 and section 3, respectively. In section 4, we use the Kriging regression imputation technique to receive the estimators for the parametric and nonparametric component. And then we provide statistics inference of asymptotic normality of profile least-squares estimator and strong convergence rate of nonparametric component. Further discussion is provided in section 5. Some simulation studies are conducted in section 6. The proofs of the main results are relegated to the Appendix.

2 Complete-case Data Estimation

As for the estimation method of complete-case data, we focus on the case where $\delta = 1$. We begin with the following assumptions:

Assumption 1. The random variable U has a bounded support Ω. Its density function $f(.)$ is Lipschitz continuous and bounded away from 0 on its support.

Assumption 2. For each $E(U \in \Omega)$, $E(XX^T|U)$ is non-singular. $E(XX^T|U)$, $E(XX^T|U)^{-1}$ and $E(XZ^T|U)$ are all Lipschitz continuous.

Assumption 3. There is an $s > 2$ such that $E||X||^{2s} < \infty$ and $E||Z||^{2s} < \infty$ and for some $\varepsilon < 2 - s^{-1}$ such that $n^{2\varepsilon-1}h \to \infty$.

Assumption 4. $\{\alpha_j(.), j = 1, \ldots p\}$ have continuous second derivatives in $U \in \Omega$.

Assumption 5. The function $K(.)$ is a symmetric density function with compact support and the bandwidth satisfies $nh^8 \to 0$ and $nh^2/(\log n)^2 \to \infty$.

Denote $c_{ni} = h_i^2 + \{\log(1/h_i)/nh_i\}^{1/2}, i = 1, 2,$

which will be used in the proof of the lemmas and theorems.

Let $\{X_i, Y_i, Z_i, U_i, \delta_i\}_{i=1}^n$ be a set of random sample with size n, from model (1), we have

$$\delta_i Y_i = \delta_i Z_i^T \beta + \delta_i X_i^T \alpha(U_i) + \delta_i \varepsilon_i. \qquad (2)$$

If the parametric component β is given, model (2) can be written as

$$\delta_i(Y_i - Z_i^T \beta) = \sum_{j=1}^p \delta_i \alpha_j(U_i) X_{ij} + \delta_i \varepsilon_i. \qquad (3)$$

According to Fan and Huang [24], we can apply the local linear regression technique to estimate the coefficient $\alpha(U)$. For u in a small neighborhood of u_0, by Taylor expansion, we have

$$\alpha_j(u) \approx \alpha_j(u_0) + \alpha_j'(u_0)(u-u_0) \triangleq a_j + b_j(u-u_0).$$

This leads to the following weighted least-squares problem: find $\{(a_j, b_j), j = 1, \ldots p\}$ to minimize

$$\sum_{i=1}^n [Y_i^* - \sum_{j=1}^p \{a_j + b_j(U_i - u_0)\} X_{ij}]^2 K_{h_1}(U_i - u_0)\delta_i, \quad (4)$$

where $Y_i^* = Y_i - Z_i^T \beta, K_{h_1}(.) = K(./h_1)/h_1$; $K(.)$ is a kernel function and h_1 is a bandwidth. From equation (4), the weighted least-square estimation of $\hat{\alpha}_c(u)$ is given by

$$\hat{\alpha}_c(u) = (I_p, 0_p)\{D_{u_0}^T W_{u_0}^\delta D_{u_0}\}^{-1} D_{u_0}^T W_{u_0}^\delta (Y - Z\beta),$$

where

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \ D_{u_0} = \begin{pmatrix} X_1^T & \frac{U_1-u_0}{h_1} X_1^T \\ X_2^T & \frac{U_2-u_0}{h_1} X_2^T \\ \vdots & \vdots \\ X_n^T & \frac{U_n-u_0}{h_1} X_n^T \end{pmatrix},$$

$$Z = \begin{pmatrix} Z_1^T \\ Z_2^T \\ \vdots \\ Z_n^T \end{pmatrix} = \begin{pmatrix} Z_{11} & Z_{12} & \ldots & Z_{1q} \\ Z_{21} & Z_{22} & \ldots & Z_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ Z_{n1} & Z_{n2} & \ldots & Z_{nq} \end{pmatrix},$$

$W_{u_0}^\delta = \mathrm{diag}(K_{h1}(U_1 - u_0)\delta_1, \ldots K_{h1}(U_n - u_0)\delta_n)$. Replace $\alpha(U_i)$ by $\hat{\alpha}_c(U_i)$, model (3) can be simplified as

$$\delta_i(Y_i - \hat{Y}_i) = \delta_i(Z_i - \hat{Z}_i)^T \beta + \delta_i \varepsilon_i, \qquad (5)$$

where $\hat{Y} \triangleq (\hat{Y}_1, \ldots \hat{Y}_n) = S_c Y$, $\hat{Z} \triangleq (\hat{Z}_1, \ldots \hat{Z}_n) = S_c Z$ with

$$S_c = \begin{pmatrix} (X_1^T, 0)\{D_{u_1}^T W_{u_1}^\delta D_{u_1}\}^{-1} D_{u_1}^T W_{u_1}^\delta \\ (X_2^T, 0)\{D_{u_2}^T W_{u_2}^\delta D_{u_2}\}^{-1} D_{u_2}^T W_{u_2}^\delta \\ \vdots \\ (X_n^T, 0)\{D_{u_n}^T W_{u_n}^\delta D_{u_n}\}^{-1} D_{u_n}^T W_{u_n}^\delta \end{pmatrix}.$$

Applying the least squares to model (5), we obtain the profile least squares estimator of β as follows

$$\hat{\beta}_c = \{\sum_{i=1}^n \delta_i (Z_i - \hat{Z}_i)(Z_i - \hat{Z}_i)^T\}^{-1} \sum_{i=1}^n \delta_i (Z_i - \hat{Z}_i)(Y_i - \hat{Y}_i). \qquad (6)$$

Substituting $\hat{\beta}_c$ into the expression of $\hat{\alpha}_c(u)$, we can further get the final estimator of $\hat{\alpha}_c(u)$ as follows

$$\hat{\alpha}_c(u) = (I_p \,, 0_p)\{D_{u_0}^T W_{u_0}^\delta D_{u_0}\}^{-1} D_{u_0}^T W_{u_0}^\delta (Y - Z\hat{\beta}_c).$$

Similar to Wei [29], the properties of $\hat{\beta}_c$, $\hat{\alpha}_c(u)$ can be shown as follows.

Theorem 1 *Suppose that the assumptions 1-5 hold, the estimator of parametric component β is asymptotically normal, that is*

$$\sqrt{n}(\hat{\beta}_c - \beta) \longrightarrow N(0, \Sigma_1^{-1} \Omega_1 \Sigma_1^{-1}),$$

where

$$\Sigma_1 = E\{\delta[Z - \Phi_c(U)\Gamma_c^{-1} X]^{\otimes 2}\}, \Omega_1 = \{E(ZZ^T) - E[E(\delta(ZX^T|U))E(\delta(XX^T|U)^{-1})E(\delta(ZX^T|U))]\},$$

$A^{\otimes 2}$ *means* AA^T.

Theorem 2 *Suppose that the assumptions 1-5 hold, if $h_1 = cn^{-1/5}$, where c is constant, then*

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

a.s.

Remark 1 *Obviously, complete-case data estimation technique is easy to see, and the estimators are easy to be obtained. It has the advantage in statistical computation. However, deleting the missing data directly means the loss of data information, this makes the method of complete-case data can not make full use of data information, so it can not capture the real relations between the responses and their associated covariates well.*

3 Regression Imputation Approach

In order to overcome the flaws of complete-case technique, a commonly used method for handing missing data is regression imputation. Since regression imputation can utilize more data information, it has received a lot of researchers' interest. The main idea of regression imputation technique is to impute a plausible value for each missing response. Chu and Cheng

[31] used the simplest local linear smoother to impute a prediction value for each missing Y on the basis of complete-case data estimation approach. For a random sample $\{Y_i, X_i, Z_i, U_i\}_{i=1}^n$ and the imputation values, the new observed values can be denoted as $\{\tilde{Y}_i, X_i, Z_i, U_i\}_{i=1}^n$, where

$$\tilde{Y}_i = \delta_i Y_i + (1-\delta_i)(X_i^T \hat{\alpha}_c(U_i) + Z_i^T \hat{\beta}_c). \quad (7)$$

Substituting the expression of $\tilde{Y}_i$ for Y in (1), we have

$$\tilde{Y}_i = X_i^T \alpha(U_i) + Z_i^T \beta + e_i, \quad (8)$$

where $e_i = \tilde{Y}_i - Y_i + \varepsilon_i$. Similar to the derivation method used in section 2, the profile least squares estimator of β by the classical regression imputation technique can be given as follows

$$\hat{\beta}_I = \{\sum_{i=1}^n (Z_i - \tilde{Z}_i)(Z_i - \tilde{Z}_i)^T\}^{-1} \sum_{i=1}^n (Z_i - \tilde{Z}_i)(\tilde{Y}_i - \tilde{Y}_i^*), \quad (9)$$

where $\tilde{Y}^* = (\tilde{Y}_1^*, \ldots \tilde{Y}_n^*) \triangleq S_I \tilde{Y}$, $\tilde{Z} \triangleq (\tilde{Z}_1, \ldots \tilde{Z}_n) = S_I Z$,

$$S_I = \begin{pmatrix} (X_1^T, 0)\{D_{u_1}^T W_{u_1} D_{u_1}\}^{-1} D_{u_1}^T W_{u_1} \\ (X_2^T, 0)\{D_{u_2}^T W_{u_2} D_{u_2}\}^{-1} D_{u_2}^T W_{u_2} \\ \vdots \\ (X_n^T, 0)\{D_{u_n}^T W_{u_n} D_{u_n}\}^{-1} D_{u_n}^T W_{u_n} \end{pmatrix},$$

$W_{u_0} = \mathrm{diag}(K_{h_2}(U_1 - u_0), \ldots K_{h_2}(U_n - u_0))$, and h_2 is different from h_1.

Furthermore, the estimator of $\alpha_I(u)$ can be expressed by

$$\hat{\alpha}_I(u) = (I_p, 0_p)\{D_{u_0}^T W_{u_0} D_{u_0}\}^{-1} \times D_{u_0}^T W_{u_0}(\tilde{Y} - Z\hat{\beta}_I).$$

The following theorems illustrate the asymptotic normality and consistency of corresponding estimators as follows.

Theorem 3 *Suppose that the assumptions 1-5 hold. The estimator of parametric component β is asymptotically normal, that is*

$$\sqrt{n}(\hat{\beta}_I - \beta) \longrightarrow N(0, \Sigma^{-1}\Omega_2\Sigma^{-1}),$$

where

$$\Sigma = E\{[Z - \Phi(U)\Gamma^{-1}(U)X][Z - \Phi(U)\Gamma^{-1}(U)X]^T\},$$

$$\Omega_2 = (\Sigma_2 + \Sigma_1)\Sigma_1^{-1}\Omega_1\Sigma_1^{-1}(\Sigma_2 + \Sigma_1),$$

$$\Sigma_2 = E\{(1-\delta)[Z - \Phi(U)\Gamma^{-1}(U)X] \times [Z - \Phi(U)\Gamma^{-1}(U)X]^T\}.$$

Theorem 4 *Suppose that the assumptions 1-5 hold.If $h_1 = b_1 n^{-1/5}$, $h_2 = b_2 n^{-1/5}$, where b_1 and b_2 are constants, then*

$$\max_{1\le j\le p} \sup_{u\in\Omega} |\hat{\alpha}_{Ij}(u) - \alpha_j(u)| = O\{n^{-2/5}(logn)^{1/2}\},$$

a.s.

Remark 2 *From the analysis of classical regression imputation approach, one can see that the estimation efficient is improved comparing with the method of complete-case data. It can make more use of the data information if we substitute missing Y_i with $X_i^T \hat{\alpha}_c(U_i) + Z_i^T \hat{\beta}_c$. However, it is worth pointing out that, the imputation process of imputation approach mainly considers the information of independent variable. If the information of response variable can be sufficiently utilized, the accuracy of estimator may be further improved.*

Remark 3 *As is well known that, least squares estimator has poor robustness. If there exists abnormal data in independent variables, the imputed value of response variable Y must be far away from the true value. In this case, the imputed value may destroy the estimation efficiency of regression imputation approach. One solution to overcome this defect is to further introduce Kriging technique into imputation process. Kriging technique can sufficiently utilize the response variable information, thus may be useful to improve the estimation efficiency and robustness. The related discussion can be seen in section 4.*

Remark 4 *Considering the function of Lasso technique in eliminating cumulative effect for model selection and imputation estimator, another method to improve the estimation robustness is to introduce Lasso technique into regression imputation approach, and the related further discussion can be seen in section 5.*

4 Kriging Regression Imputation Technique

Combined Kriging imputation idea with the approach of classical regression imputation, in this section, we propose a modified imputation technique called Kriging regression imputation. Kriging imputation aims at assigning a weight to each non-lost response and imputes response $\hat{Y}_i = (X_i^T \hat{\alpha}_c(U_i) + Z_i^T \hat{\beta}_c)$. Based on the theory of Kriging imputation, the result is more precise and has less error due to the non-bias condition together with minimum estimation variance are required.

For the convenience of discussion, we first discuss the case where the weight function is previous given. This is a special Kriging imputation technique. Suppose that $\{Y_i, X_i, Z_i, U_i\}_{i=1}^n$ is a random sample, and the imputation value is $(X_i^T\hat{\alpha}_c(U_i) + Z_i^T\hat{\beta}_c)$, then the new observed values can be noted as $\{Y_i^0, X_i, Z_i, U_i\}_{i=1}^n$, where

$$Y_i^0 = \bar{\delta}_i Y_i + (1-\bar{\delta}_i)\sum_{j=1}^n K_h^{ij}(U_j - u_i)\tilde{Y}_j, \quad (10)$$

$\tilde{Y}_j$ is defined in (7). If $\bar{\delta} = 0$, means that Y is missing, and $\bar{\delta} = 1$ otherwise. Here $\bar{\delta}$ can be independent with δ, and we always assume that

$$\sum_{j=1}^n K_h^{ij}(U_j - u_i) = 1$$

Similar to the discussions in section 3, we can obtain

$$Y_i^0 = X_i^T\alpha(U_i) + Z_i^T\beta + er_i, \quad (11)$$

where $er_i = Y_i^0 - Y_i + \varepsilon_i$. For model (11), the estimator of nonparametric component $\alpha(u)$, we follows the idea of locally linear smoothing by the Taylor expansion, and find the optimal estimator similar to section 2. The profile least squares estimator of β by the Kriging regression imputation technique is

$$\hat{\beta}_{KI} = \{\sum_{i=1}^n (Z_i - \check{Z}_i)(Z_i - \check{Z}_i)^T\}^{-1}\sum_{i=1}^n (Z_i - \check{Z}_i)(Y_i^0 - \check{Y}_i^0), \quad (12)$$

where $\check{Y}^0 \triangleq (\check{Y}_1^0, \ldots \check{Y}_n^0) = S_{KI}Y^0$, $\check{Z} \triangleq (\check{Z}_1 \ldots, \check{Z}_n) = S_{KI}Z$,

$$S_{KI} = \begin{pmatrix} (X_1^T, 0)\{D_{u_1}^T W_{u_1}^0 D_{u_1}\}^{-1} D_{u_1}^T W_{u_1}^0 \\ (X_2^T, 0)\{D_{u_2}^T W_{u_2}^0 D_{u_2}\}^{-1} D_{u_2}^T W_{u_2}^0 \\ \vdots \\ (X_n^T, 0)\{D_{u_n}^T W_{u_n}^0 D_{u_n}\}^{-1} D_{u_n}^T W_{u_n}^0 \end{pmatrix},$$

$W_{u_0}^0 = \text{diag}(K_{h3}(U_1 - u_0), \ldots K_{h3}(U_n - u_0))$, and h_3 can be different from h_1 and h_2.

Furthermore, the estimator of $\alpha_{KI}(U)$ can be expressed by

$$\hat{\alpha}_{KI}(u) = (I_p\ , 0_p)\{D_{u_0}^T W_{u_0}^0 D_{u_0}\}^{-1} \times$$
$$D_{u_0}^T W_{u_0}(Y^0 - Z\hat{\beta}_{KI}).$$

The theorems illustrate the asymptotic normality and consistency of corresponding estimators are as follows.

Theorem 5 *Under the assumptions 1-5, the profile least square estimator of β is asymptotically normal, that is*

$$\sqrt{n}(\hat{\beta}_{KI} - \beta) \longrightarrow N(0, \Xi^{-1}\Omega_3\Xi^{-1}),$$

where

$$\begin{aligned} \Xi &= E\{[Z - \Phi_{KI}(U)\Gamma_{KI}^{-1}(U)X] \times \\ &\quad [Z - \Phi_{KI}(U)\Gamma_{KI}^{-1}(U)X]^T\}, \\ \Omega_3 &= (\Sigma_3 + \check{\Sigma}_1)\Sigma_1^{-1}\Omega_1\Sigma_1^{-1}(\Sigma_3 + \check{\Sigma}_1), \\ \Sigma_3 &= \sum_{j=1}^n K_h^{ij}(U_j - u_0)(1-\delta_j)E\{(1-\bar{\delta}) \\ &[Z - \Phi_{KI}(U)\Gamma_{KI}^{-1}(U)X][Z - \Phi_{KI}(U)\Gamma_{KI}^{-1}(U)X]^T\}, \\ \check{\Sigma}_1 &= \sum_{j=1}^n K_h^{ij}(U_j - u_0)(1-\delta_j)\Sigma_1. \end{aligned}$$

Theorem 6 *Suppose that the assumptions 1-5 hold. If $h_1 = c_1 n^{-1/5}$, $h_3 = c_2 n^{-1/5}$, where c_1 and c_2 are constants, then we have*

$$\max_{1\le j\le p}\sup_{u\in\Omega}|\hat{\alpha}_{KI_j}(u) - \alpha_j(u)| = O\{n^{-2/5}(logn)^{1/2}\},$$

a.s.

Theorem 7 *It is easy to see that the Kriging regression imputation technique is more efficient than the the traditional solution of handling missing response. Based on theorem 3 and theorem 5 we have*

$$Var\hat{\beta}_{KI} \le Var\hat{\beta}_I.$$

When the kernel function $K_h^{ij}(U_j - u_i)$ is not previous given, we can obtain the weights by establishing Kriging equations. Note that variable v_i is the observations of (Y_i, U_i) geographic coordinates of point locations. If we denote λ as weight coefficient matrix and then we can deduce Kriging equation set by decomposing variation function which aims at minimizing variance of estimation errors. Further, we can obtain the weight coefficient matrix λ and conduct statistics inference further. If the Kriging equations are expressed in matrix form, then

$$K\lambda = M, \quad (13)$$

where $K =$

$$\begin{pmatrix} \gamma(U_1,U_1) & \gamma(U_1,U_2) & \ldots & \gamma(U_1,U_n) & 1 \\ \gamma(U_2,U_1) & \gamma(U_2,U_2) & \ldots & \gamma(U_2,U_n) & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma(U_n,U_1) & \gamma(U_n,U_2) & \ldots & \gamma(U_n,U_n) & 1 \\ 1 & 1 & \ldots & 1 & 0 \end{pmatrix},$$

$$M = \begin{pmatrix} \gamma(U_1, U) \\ \gamma(U_2, U) \\ \vdots \\ \gamma(U_n, U) \\ 1 \end{pmatrix}, \qquad \lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{pmatrix}.$$

Here, $\gamma(.)$ means the variance function, $\gamma(U_1, U)$ means the variance function value between the first point and the unknown point. In that case, from equation (13), the weight coefficient matrix can be expressed by

$$\lambda = K^{-1}M.$$

5 Kriging Regression Imputation with LASSO Technique

In order to improve the robustness of estimator, in this section, we will further introduce lasso technique into the semiparametric model. Since LASSO technique can complete the parameter estimation and model selection simultaneously, thus, it attracts more and more researchers' attention. Here, our purpose is to utilize LASSO technique to identify unimportant regression independent variables, and remove them. In this way, the estimation error may become smaller, thus can improve the robustness of estimator.

Firstly, in the case of complete-case data, we adopt the responses where $\delta_i = 1, i = 1, \ldots, n$. For $\alpha(u)$, by Taylor expansion, we have

$$\alpha_j(u) \approx \alpha_j(u_0) + \alpha_j'(u_0)(u - u_0)$$

Thus, model (2) can be rewritten as

$$Y^{\delta_i} = X_i^{\delta^T}\alpha(u_0) + (U_i - u_0)X_i^{\delta^T}\alpha'(u_0) + Z_i^{\delta^T}\beta + \varepsilon_i^{\delta}, \quad (14)$$

where $Y^{\delta_i} = \delta_i Y_i$, $X_i^{\delta} = \delta_i X_i$, $Z_i^{\delta} = \delta_i Z_i$, $\varepsilon_i^{\delta} = \delta_i \varepsilon_i$.

Let us consider the following lasso model,

$$Q_\lambda(\beta) = \sum_{i=1}^{n} \delta_i \{Y_i - \hat{Y}_i - (Z_i - \hat{Z}_i)^T\beta)\}^2 + \lambda \sum_{j=1}^{q} |\beta_j|, \quad (15)$$

where $\lambda > 0$ is an arbitrary constant.

The lasso estimator can be given by

$$\hat{\beta}^{lasso} = arg\ min\ Q_\lambda(\beta).$$

Notice that, if $\hat{\beta}_c \geq \tilde{\lambda}$, where $\tilde{\lambda} = \frac{1}{2}\lambda$, $\hat{\beta}^{lasso} = \hat{\beta}_c - \tilde{\lambda}$. If $\hat{\beta}_c \leq -\tilde{\lambda}$, $\hat{\beta}^{lasso} = \hat{\beta}_c + \tilde{\lambda}$. If $-\tilde{\lambda} \leq \hat{\beta}_c \leq \tilde{\lambda}$, $\hat{\beta}^{lasso} = 0$, thus the relationship between $\hat{\beta}_c$ with $\hat{\beta}^{lasso}$ is

$$\hat{\beta}^{lasso} = \hat{\beta}_c - sign(\hat{\beta}_c)\tilde{\lambda}.$$

Then equation (7) can be rewritten as

$$\tilde{Y}_i^{lasso} = \delta_i Y_i + (1 - \delta_i)(X_i^T \hat{\alpha}_c^{lasso}(U_i) + Z_i^T \hat{\beta}_c^{lasso}), \quad (16)$$

where $\hat{\alpha}_c^{lasso}(u_i) = (I_p \quad, 0_p)\{D_{u_0}^T W_{u_0}^{\delta} D_{u_0}\}^{-1} D_{u_0}^T W_{u_0}^{\delta}(Y - Z\hat{\beta}_c^{lasso})$.

Substituting $\tilde{Y}_i$ with $\tilde{Y}_i^{lasso}$ in model (7), we can obtain the new form of Y_i^0 as follows

$$Y_i^{0lasso} = \bar{\delta}_i Y_i + (1 - \bar{\delta}_i)\sum_{j=1}^{n} K_h^{ij}(U_j - u_i)\tilde{Y}_j^{lasso}. \quad (17)$$

Similar to the analysis process in section 4, we can get the following statistical inference results.

Theorem 8 *Suppose that the assumptions 1-5 hold. The profile least square estimator of β satisfies*

$$\sqrt{n}(\hat{\beta}_{KIL} - \beta) \longrightarrow N(sign(\hat{\beta}_c)\tilde{\lambda}\Xi^{-1}, \Xi^{-1}(\Omega_3 + \tilde{\lambda}^2)\Xi^{-1}),$$

where

$$\begin{aligned} \Xi &= E\{[Z - \Phi_{KI}(U)\Gamma_{KI}^{-1}(U)X] \times [Z - \Phi_{KI}(U)\Gamma_{KI}^{-1}(U)X]^T\}, \\ \Omega_3 &= (\Sigma_3 + \check{\Sigma}_1)\Sigma_1^{-1}\Omega_1\Sigma_1^{-1}(\Sigma_3 + \check{\Sigma}_1), \\ \Sigma_3 &= \sum_{j=1}^{n} K_h^{ij}(U_j - u_i)(1 - \delta_j)E\{(1 - \bar{\delta}) \\ &[Z - \Phi_{KI}(U)\Gamma_{KI}^{-1}(U)X][Z - \Phi_{KI}(U)\Gamma_{KI}^{-1}(U)X]^T\}, \\ \check{\Sigma}_1 &= \sum_{j=1}^{n} K_h^{ij}(U_j - u_i)(1 - \delta_j)\Sigma_1. \end{aligned}$$

Theorem 9 *Suppose that the assumptions 1-5 hold. If $h_1 = d_1 n^{-1/5}$, $h_4 = d_2 n^{-1/5}$, where d_1 and d_2 are constants, then we have*

$$\max_{1 \leq j \leq p} \sup_{u \in \Omega} |\hat{\alpha}_{KIL_j}(u) - \alpha_j(u)| = O\{n^{-2/5}(logn)^{1/2}\},$$

a.s.

In order to further improve the robustness of the estimators, according to Zou [33], we can also consider the following adaptive lasso model

$$Q_\lambda(\theta) = \sum_{i=1}^{n}\{Y_i^{\delta} - Z_i^{\delta^T}\beta - X_i^{\delta^T}\alpha(u_0) - (U_i - u_0)X_i^{\delta^T}\alpha'(u_0))\}^2 + \lambda \sum_{j=1}^{2p+q} w_j|\theta_j|, \quad (18)$$

where $\theta = (\alpha_1(u_0), \dots \alpha_p(u_0), \alpha'_{p+1}(u_0) \dots \alpha'_{2p}(u_0), \beta_1, \dots \beta_q), w_j = 1/|\hat{\theta}|^\tau, \tau > 0$.

The adaptive lasso estimator can be given by

$$\hat{\theta}^{alasso} = arg\ min\ Q_\lambda(\theta).$$

Then model (7) can be rewritten as

$$\widetilde{Y}_i^{alasso} = \delta_i Y_i + (1-\delta_i)(X_i^T \hat{\alpha}_c^{alasso}(U_i) + Z_i^T \hat{\beta}_c^{alasso}). \quad (19)$$

Similar to the proof of theorem8 and theorem 9, we can further derive the asymptotically properties. Since the expressions are the same as to theorem 8 and theorem 9, thus they are omitted here.

6 Numerical experiments

6.1 Simulation Studies

In this section, we carried a numerical simulation example to show the effectiveness and superiority of our method.

Considering the semiparameter model as follows:

$$Y = \beta_1 Z_1 + \beta_2 Z_2 + \alpha_1(u) X_1 + \alpha_2(u) X_2 + \varepsilon,$$

where $\beta_1 = 3, \beta_2 = 2$, $\alpha_1(u) = \sin(2\pi u)$, $\alpha_2(u) = \cos(2\pi u)$. Let $Z_1 \sim N(0,1)$, $Z_2 \sim N(0,1.5)$, $X_1 \sim N(0,1.5)$, $X_2 \sim N(0,1)$, $u \sim U(0,1)$, $\varepsilon \sim N(0,\sigma)$. By using the techniques described in previous sections, our aim is to estimate β_1, β_2 and compare the estimation efficiency.

In all simulations, we consider a random sample with 30 percent for missing. Let the kernel function is standard Gauss kernel, and the bandwidth $h = 0.1$. Table 1 shows the compared results with different methods. From Table 1, one can see that, for different σ, the methods established in this paper have smaller variance, which means our method is more superior than traditional complete-case method and classical regression imputation approach.

Table 1: Estimators of β_1 and β_2 for different σ

σ	1	1.5	2.0	2.5
mean(β_{C1})	3.0197	3.0335	3.0723	3.0429
std(β_{C1})	0.6704	0.8481	0.7375	0.7729
mean(β_{C2})	2.0144	1.9809	1.9342	2.0557
std(β_{C2})	0.4299	0.5376	0.8101	0.5813
mean(β_{I1})	3.0695	3.0956	3.1363	3.0392
std(β_{I1})	0.6358	0.8127	0.6412	0.7396
mean(β_{I2})	1.9727	1.9345	2.0352	2.0455
std(β_{I2})	0.3605	0.4491	0.5972	0.5547
mean(β_{KI1})	3.0690	3.0952	3.1366	3.0397
std(β_{KI1})	0.6349	0.8109	0.6411	0.7387
mean(β_{KI2})	1.9728	1.9348	2.0351	2.0456
std(β_{KI2})	0.3598	0.4487	0.5968	0.5546
mean(β_{KIL1})	3.0690	3.0952	3.1366	3.0397
std(β_{KIL1})	0.6349	0.8109	0.6411	0.7387
mean(β_{KIL2})	1.9727	1.9348	2.0351	2.0456
std(β_{KIL2})	0.3598	0.4487	0.5968	0.5546

Table 2: Estimators of β_1 and β_2 in the application of Boston housing data

σ	2		coefficient
mean(β_{C1})	0.8858	$\alpha_{c1}(u_0)$	-1.5189
std(β_{C1})	1.4084	$\alpha_{c2}(u_0)$	0.32815
mean(β_{C2})	0.4108	$\alpha_{c3}(u_0)$	-0.0503
std(β_{C2})	0.2164	$\alpha_{c4}(u_0)$	30.9534
mean(β_{I1})	0.9595	$\alpha_{I1}(u_0)$	-1.14285
std(β_{I1})	1.3111	$\alpha_{I2}(u_0)$	1.2548
mean(β_{I2})	0.4039	$\alpha_{I3}(u_0)$	-0.09895
std(β_{I2})	0.2114	$\alpha_{I4}(u_0)$	54.52215
mean(β_{KI1})	0.9609	$\alpha_{KI1}(u_0)$	-1.1427
std(β_{KI1})	1.3090	$\alpha_{KI2}(u_0)$	1.2549
mean(β_{KI2})	0.4036	$\alpha_{KI3}(u_0)$	-0.09895
std(β_{KI2})	0.2111	$\alpha_{KI4}(u_0)$	54.53585
mean(β_{KIL1})	0.9609	$\alpha_{KIL1}(u_0)$	-1.14285
std(β_{KIL1})	1.3090	$\alpha_{KIL2}(u_0)$	1.2548
mean(β_{KIL2})	0.4036	$\alpha_{KIL3}(u_0)$	-0.09895
std(β_{KIL2})	0.2111	$\alpha_{KIL4}(u_0)$	54.5221

6.2 Application to Boston housing data

To further illustrate the efficiency of the proposed methods, we take the application of Boston housing data for example. Following Fan and Huang[24], we take MEDV (median value of owner-occupied homes in 1,000 United States dollar) as the responses, $\sqrt{LSTAT}$(the percentage of lower status of the population) as the index variable, and the following predictors as the covariates: CRIM(per-captia crime rata by town), RM(average number of rooms per dweling), TAX(full-value property-tax rate per \$10000), NOX (nitric oxide concentration in parts per 10 million), PTRATIO(pupil-teacher ratio by town), AGE(proportion of owner-occupied units built prior to 1940). Consider the semiparametric model as follows:

$$Y = \beta_1 Z_1 + \beta_2 Z_2 + \sum_{k=1}^{4} \alpha_k(U) X_k + \varepsilon.$$

Let PTRATIO, AGE be the variables of Z_1,Z_2 respectively, and CRIM, RM, TAX, NOX are denoted respectively by X_1,$\dots X_4$. The main results are provided in the form of Table 2. From Table 2, one can see that, for different σ, the methods established in this paper have smaller variance, which means our method is more superior than traditional complete-case method and classical regression imputation approach.

7 Conclusion

In this paper, we propose a modified regression imputation approach called Kriging regression imputation on the basis of the estimation method of complete-case data and classical regression imputation. Compared with the method of complete-case data, Kriging regression imputation technique not only makes more use of the data information, but also improves the estimation efficiency. Compared with the classical regression imputation approach, the estimation efficiency of Kriging regression imputation technique with LASSO is better as the imputed values not only consider the effect of covariates but also takes the influence of responses into account. Numerical experiment shows that the technique established in this paper is more effective than the results obtained in the references cited there in, and have better robustness.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (61472093).

Appendix

Lemma 7.1 *Let $(X_1, Y_1)\ldots(X_n, X_n)$, be independent and identically distributed random vectors, where the $Y_i, i = 1, 2, \ldots n$. are scalar random variables. Further assume that $E|y|^s < \infty$ and $\sup_x \int |y|^s f(x, y)dy < \infty$, where f means the joint density of (X,Y). Let K be a bounded positive function with a bounded support, satisfying a Lipschitz condition. Given that $n^{2\varepsilon-1}h \to \infty$ for some $\varepsilon < 1 - s^{-1}$, then*

$$\sup_x |\frac{1}{n}\sum_{i=1}^n [K_h(X_i - x)Y_i - E\{K_h(X_i - x)Y_i\}]|$$

$$= O_p(\{\frac{\log(1/h)}{nh}\}^{1/2}).$$

The proof of Lemma 7.1 can be found in Mack and Silverman [34].

Lemma 7.2 *Suppose that the Assumptions 1-5 hold. Then it shows that*

$$n^{-1}\sum_{i=1}^n \delta_i(Z_i - \hat{Z}_i)(Z_i - \hat{Z}_i)^T \to \Sigma_1,$$

$$n^{-1}\sum_{i=1}^n (Z_i - \tilde{Z}_i)(Z_i - \tilde{Z}_i)^T \to \Sigma.$$

$$n^{-1}\sum_{i=1}^n (Z_i - \check{Z}_i)(Z_i - \check{Z}_i)^T \to \Xi.$$

The proof of Lemma 7.2 is similar to Lemma A.2 in Fan and Huang [24], which is omitted here.

Proof of Theorem 1. By the definition of $\hat{\beta}_c$ and let $\Lambda_n = n^{-1}\sum_{i=1}^n \delta_i(Z_i - \hat{Z}_i)(Z_i - \hat{Z}_i)^T$, then we have

$$\begin{aligned} & \sqrt{n}(\hat{\beta}_c - \beta) \\ = & \frac{1}{\sqrt{n}}\Lambda_n^{-1}\sum_{i=1}^n \delta_i(Z_i - \hat{Z}_i)(X_i^T\alpha(U_i) - S_{ci}^T M) \\ & + \frac{1}{\sqrt{n}}\Lambda_n^{-1}\sum_{i=1}^n \delta_i(Z_i - \hat{Z}_i)(\varepsilon_i - \hat{\varepsilon}_i) \\ = & I_1 + I_2. \end{aligned} \tag{20}$$

By lemma 7.1 and similar to the proof of Theorem 4.1 in Fan and Huang [24], it is easy to show $I_1 = O_p(\sqrt{n}c_n^2)$.

Denote

$$D_{u_0}^T W_{u_0}^\delta D_{u_0} = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix},$$

where

$$\begin{aligned} D_{11} &= \sum_{i=1}^n X_i X_i^T K_{h_1}(U_{iu})\delta_i, U_{iu} = U_i - u_0, \\ D_{12} &= \sum_{i=1}^n X_i X_i^T \frac{U_{iu}}{h_1} K_{h_1}(U_{iu})\delta_i, \\ D_{21} &= \sum_{i=1}^n X_i X_i^T \frac{U_{iu}}{h_1} K_{h_1}(U_{iu})\delta_i, \\ D_{22} &= \sum_{i=1}^n X_i X_i^T (\frac{U_{iu}}{h_1})^2 K_{h1}(U_{iu})\delta_i. \end{aligned}$$

Notice that

$$D_{u_0}^T W_{u_0}^\delta Z = \begin{pmatrix} \sum_{i=1}^n X_i Z_i^T K_{h_1}(U_i - u_0)\delta_i \\ \sum_{i=1}^n X_i Z_i^T \frac{U_i - u_0}{h_1} K_{h_1}(U_i - u_0)\delta_i \end{pmatrix},$$

by lemma 7.1, it is easy to see that

$$D_{u_0}^T W_{u_0}^\delta D_{u_0} = nf(U)\Gamma_c(U) \otimes \begin{pmatrix} 1 & 0 \\ 0 & u_2 \end{pmatrix}$$

$$\times \{1 + O_p(c_n)\},$$

$$D_{u_0}^T W_{u_0}^\delta Z = nf(U)\Phi_c(U) \otimes (1, 0)^T\{1 + O_p(c_n)\},$$

where

$$\Gamma_c(U) = E(\delta X X^T \mid U), \Phi_c(U) = E(\delta X Z^T \mid U).$$

By simple calculation, we have

$$\begin{aligned} & [X^T, 0]\{D_{u_0}^T W_{u_0}^\delta D_{u_0}\}^{-1} D_{u_0}^T W_{u_0}^\delta Z \\ & \qquad = X^T \Gamma_c(U)^{-1}\Phi_c(U)\{1 + O_p(c_n)\}, \end{aligned}$$

$$[X^T,0]\{D_{u_0}^T W_{u_0}^\delta D_{u_0}\}^{-1} D_{u_0}^T W_{u_0}^\delta M = X^T\alpha(U)\{1+O_p(c_n)\}.$$

Thus, we have

$$\begin{aligned}
I_1 &= \frac{1}{\sqrt{n}}\Lambda_n^{-1}\sum_{i=1}^n \delta_i(Z_i-\hat{Z}_i)(X_i^T\alpha(U_i)-S_{ci}^T M)\\
&= \frac{1}{\sqrt{n}}\Lambda_n^{-1}\sum_{i=1}^n \delta_i(Z_i-\hat{Z}_i)(X_i^T\alpha(U_i)-[X_i^T,0]\\
&\quad\times\{D_{u_i}^T W_{u_i}^\delta D_{u_i}\}^{-1}D_{u_i}^T W_{u_i}^\delta M\}\\
&= \frac{1}{\sqrt{n}}\Lambda_n^{-1}\sum_{i=1}^n \delta_i(Z_i-X_i^T\Gamma_c^{-1}(U_i)\Phi_c(U_i))\\
&\quad\times X_i^T\alpha(U_i)\{1+O_p(c_n)\}O_p(c_n)\\
&= O_p(\sqrt{n}c_n^2).
\end{aligned} \tag{21}$$

Similarly,

$$I_2=\Lambda_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_i(Z_i-X_i^T\Gamma_c^{-1}(U_i)\Phi_c(U_i))(\varepsilon_i-\hat{\varepsilon}_i),$$

where

$$\begin{aligned}
\hat{\varepsilon} &= [X^T,0]\{D_{u_0}^T W_{u_0}^\delta D_{u_0}\}^{-1}D_{u_0}^T W_{u_0}^\delta \varepsilon\\
&= X^T\Gamma_c^{-1}(U)E(\delta X\varepsilon^T|U)O_p(c_n).
\end{aligned}$$

By lemma 7.2, we obtain

$$I_2=\Sigma_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_i(Z_i-X_i^T\Gamma_c^{-1}(U_i)\Phi_c(U_i))\varepsilon_i+O_p(1).$$

From Slutsky theorem, and center limit theorem, one can obtain the results described in theorem 1.

Proof of Theorem 2. By the definition of $\hat{\alpha}_c(u)$, we have

$$\begin{aligned}
&\hat{\alpha}_c(u)\\
&= (I_p,0_p)\{D_{u_0}^T W_{u_0}^\delta D_{u_0}\}^{-1}D_{u_0}^T W_{u_0}^\delta (Y-Z\hat{\beta}_c)\\
&= (I_p,0_p)\{D_{u_0}^T W_{u_0}^\delta D_{u_0}\}^{-1}D_{u_0}^T W_{u_0}^\delta M\\
&\quad+(I_p,0_p)\{D_{u_0}^T W_{u_0}^\delta D_{u_0}\}^{-1}D_{u_0}^T W_{u_0}^\delta (\varepsilon-Z\beta)\\
&\quad+(I_p,0_p)\{D_{u_0}^T W_{u_0}^\delta D_{u_0}\}^{-1}D_{u_0}^T W_{u_0}^\delta Z(\beta-\hat{\beta}_c\}.
\end{aligned} \tag{22}$$

Similar to the proof of Theorem 3.1 in Xia and Li [14], we show that

$$\max_{1\le j\le p}\sup_{U\in\Omega}|\hat{\alpha}_{cj}(u)-\alpha_j(u)|=O\{h_1^2+(logn/nh_1)^{1/2}\},$$

a.s.

If we let $h_1=cn^{-1/5}$, where c is constant, then it satisfies,

$$\max_{1\le j\le p}\sup_{U\in\Omega}|\hat{\alpha}_{cj}(u)-\alpha_j(u)|=O\{n^{-2/5}(logn)^{1/2}\}$$

a.s., which completes the proof.

Proof of Theorem 3. Denote

$$\Pi_n=\frac{1}{n}\sum_{i=1}^n (Z_i-\tilde{Z}_i)(Z_i-\tilde{Z}_i)^T.$$

From equation (9), we have

$$\begin{aligned}
&\sqrt{n}(\hat{\beta}_I-\beta)\\
&= \frac{1}{\sqrt{n}}\Pi_n^{-1}\sum_{i=1}^n (Z_i-\tilde{Z}_i)(\tilde{Y}_i-\tilde{Y}_i^*\\
&\quad-(Z_i-\tilde{Z}_i)^T\beta)\\
&= \frac{1}{\sqrt{n}}\Pi_n^{-1}(1-\delta_i)\sum_{i=1}^n (Z_i-\tilde{Z}_i)(Z_i-\\
&\qquad\hat{Z}_i)^T(\hat{\beta}_c-\beta)+\frac{1}{\sqrt{n}}\Pi_n^{-1}\sum_{i=1}^n (Z_i-\tilde{Z}_i)\\
&\qquad\times(X_i^T\alpha(U_i)+\tilde{Z}_i^T\beta-S_{I_i}^T\tilde{Y}_i)+\frac{1}{\sqrt{n}}\Pi_n^{-1}\\
&\qquad\times\sum_{i=1}^n (Z_i-\tilde{Z}_i)\delta_i\varepsilon_i+\frac{1}{\sqrt{n}}\Pi_n^{-1}\sum_{i=1}^n (1-\delta_i)\\
&\qquad\times(Z_i-\tilde{Z}_i)(S_{ci}^T M-X_i^T\alpha(U_i))+\frac{1}{\sqrt{n}}\Pi_n^{-1}\\
&\qquad\times(1-\delta_i)\sum_{i=1}^n (Z_i-\tilde{Z}_i)S_{c_i}^T\varepsilon_i\\
&= \Pi_n^{-1}(I_1+I_2+I_3+I_4+I_5).
\end{aligned} \tag{23}$$

Let $E(XZ^T|U)=\Phi(U), E(XX^T|U)=\Gamma(U)$, then $\tilde{Z}_i=\Phi(U_i)\Gamma^{-1}(U_i)X_i$. By Lemma 7.1 and the law of large numbers, we have

$$\begin{aligned}
I_1 &= \frac{1}{\sqrt{n}}\sum_{i=1}^n (1-\delta_i)(Z_i-\Phi(U_i)\Gamma^{-1}(U_i)X_i)(Z_i\\
&\quad-\Phi_c(U_i)\Gamma_c^{-1}(U_i)X_i)^T\sqrt{n}(\hat{\beta}_c-\beta)+O_p(1)\\
&= \Sigma_2\Sigma_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \delta_i(Z_i\\
&\quad-X_i^T\Gamma_c^{-1}(U_i)\Phi_c(U_i))\varepsilon_i+O_p(1).
\end{aligned} \tag{24}$$

Consequently,

$$\begin{aligned}
I_2 &= \frac{1}{\sqrt{n}}\sum_{i=1}^n (Z_i-\tilde{Z}_i)(X_i^T\alpha(U_i)-S_{Ii}M)\\
&\quad-\frac{1}{\sqrt{n}}\sum_{i=1}^n (Z_i-\tilde{Z}_i)\delta S_{I_i}\varepsilon
\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i-\tilde{Z}_i)(I-\delta)S_{Ii}(\hat{M}_c-M)\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i-\tilde{Z}_i)\tilde{Z}_i^T(I-\delta)(\hat{\beta}_c-\beta)\\
=\;& I_{21}+I_{22}+I_{23}+I_{24}, \qquad (25)
\end{aligned}
$$

where $\hat{M}_c=[X_1^T\hat{\alpha}_c(U_1),\ldots,X_n^T\hat{\alpha}_c(U_n)]^T$.

By lemma 7.1, one can obtain $I_{21}=O_p(1)$, $I_{22}=O_p(1)$. Notice that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(Z_i-\tilde{Z}_i)\tilde{Z}_i^T=O_p(1),\\
&\beta-\hat{\beta}_c=O_p(n^{-1/2}),\\
&\frac{1}{n}\sum_{i=1}^{n}(Z_i-\tilde{Z}_i)S_{Ii}^TX=O_p(1).
\end{aligned}
$$

As a result it has that $I_{23}=O_p(1)$, $I_{24}=O_p(1)$. Thus, we have $I_2=O_p(1)$. By Lemma 7.1, we have $I_3=O_p(\sqrt{n}c_n^2)$.

As for I_5, one can see that

$$
\begin{aligned}
I_5 =\;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)Z_i(X_i^T,0)\{D_{u_i}^TW_{u_i}^{\delta}D_{u_i}\}^{-1}\\
&\times D_{u_i}^TW_{u_i}^{\delta}\varepsilon_i-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\tilde{Z}_i(X_i^T,0)\\
&\times\{D_{u_i}^TW_{u_i}^{\delta}D_{u_i}\}^{-1}D_{u_i}^TW_{u_i}^{\delta}\varepsilon_i,\\
=\;& I_{51}-I_{52}. \qquad (26)
\end{aligned}
$$

Furthermore, it follows that

$$
\begin{aligned}
I_{51} =\;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)Z_iX_i^T(nf(u_i)\Gamma_c^{-1}(U_i))^{-1}\\
&\times\sum_{j=1}^{n}K_{h1}(U_j-u_i)X_j\delta_j\varepsilon_j+O_p(1)\\
=\;& \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\{\sum_{i=1}^{n}(1-\delta_i)Z_iX_i^T(nf(u_i)\\
&\times\Gamma_c^{-1}(U_i))^{-1}K_{h1}(U_j-u_i)\}X_j\delta_j\varepsilon_j+O_p(1)\\
=\;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi(U_i)\Gamma_c^{-1}(U_i)X_i\delta_i\varepsilon_i\\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi_c(U_i)\Gamma_c^{-1}(U_i)X_i\delta_i\varepsilon_i+O_p(1), \qquad (27)
\end{aligned}
$$

$$
\begin{aligned}
I_{52} =\;& \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\{\sum_{i=1}^{n}(1-\delta_i)\Phi(U_i)\Gamma^{-1}(U_i)X_iX_i^T\\
&\times(nf(u_i)\Gamma_c^{-1}(U_i))^{-1}K_{h1}(U_j-u_i)\}\\
&\times X_j\delta_j\varepsilon_j+O_P(1)\\
=\;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi(U_i)\Gamma_c^{-1}(U_i)X_i\delta_i\varepsilon_i-\\
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi(U_i)\Gamma^{-1}(U_i)X_i\delta_i\varepsilon_i+O_p(1). \qquad (28)
\end{aligned}
$$

From equation (27) and (28), we obtain

$$
\begin{aligned}
I_5 =\;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\Phi(U_i)\Gamma^{-1}(U_i)X_i\delta_i\varepsilon_i\\
&-\Phi_c(U_i)\Gamma_c^{-1}(U_i)X_i\delta_i\varepsilon_i)+O_p(1).
\end{aligned}
$$

Combined with the given results, we further conclude that

$$
\begin{aligned}
\sqrt{n}(\hat{\beta}_I-\beta) =\;& \Sigma^{-1}(\Sigma_1+\Sigma_2)\Sigma^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i(Z_i\\
&-X_i^T\Gamma_c^{-1}(U_i)\Phi_c(U_i))\varepsilon_i+O_p(1).
\end{aligned}
$$

Proof of Theorem 5. Let $\nabla_n=n^{-1}\sum_{i=1}^{n}\delta_i(Z_i-\check{Z}_i)(Z_i-\check{Z}_i)^T$, from the definition of $\hat{\beta}_{KI}$, we obtain

$$
\begin{aligned}
&\sqrt{n}(\hat{\beta}_{KI}-\beta)\\
&=\frac{1}{\sqrt{n}}\nabla_n^{-1}\sum_{i=1}^{n}(Z_i-\check{Z}_i)\\
&\quad\times(Y_i^0-\check{Y}_i^0-(Z_i-\check{Z}_i)^T\beta)\\
&=\frac{1}{\sqrt{n}}\nabla_n^{-1}\sum_{i=1}^{n}(Z_i-\check{Z}_i)(1-\bar{\delta}_i)\sum_{j=1}^{n}\\
&\quad K_h^{ij}(U_j-u_i)(1-\delta_j)(Z_i-\hat{Z}_i)^T(\hat{\beta}_c\\
&\quad-\beta)+\frac{1}{\sqrt{n}}\nabla_n^{-1}\sum_{i=1}^{n}(Z_i-\check{Z}_i)(1-\bar{\delta}_i)\\
&\quad K_h^{ij}\sum_{j=1}^{n}(U_j-u_i)(1-\delta_j)(S_{ci}M-\\
&\quad X_i^T\alpha(U_i))+\frac{1}{\sqrt{n}}\nabla_n^{-1}\sum_{i=1}^{n}(Z_i-\check{Z}_i)\\
&\quad(X_i^T\alpha(U_i)+\check{Z}_i^T\beta-S_{KI}Y_i^0)+\frac{1}{\sqrt{n}}\\
&\quad\nabla_n^{-1}\sum_{i=1}^{n}(Z_i-\check{Z}_i)(1-\bar{\delta}_i)\sum_{j=1}^{n}K_h^{ij}\\
&\quad(U_j-u_i)(1-\delta_j)\hat{\varepsilon}_j+\frac{1}{\sqrt{n}}\nabla_n^{-1}\sum_{i=1}^{n}\\
&\quad(Z_i-\check{Z}_i)(\bar{\delta}_i\varepsilon_i+(1-\bar{\delta})\sum_{j=1}^{n}K_h^{ij}(U_j\\
&\quad-u_i)\delta_j\varepsilon_j)\\
&=\nabla_n^{-1}(I_1+I_2+I_3+I_4+I_5). \qquad (29)
\end{aligned}
$$

Notice that $\sum_{j=1}^{n} K_h^{ij}(U_j-u_i)=1$, one can obtain

$$\begin{aligned}
&(Z_i-\hat{Z}_i)^T\beta \\
= \;& (1-\bar{\delta}_i)\sum_{j=1}^{n}(1-\delta_j)K_h^{ij}(U_j-u_i) \\
& \times (Z_i-\hat{Z}_i)^T\beta+(1-\bar{\delta}_i)\sum_{j=1}^{n}K_h^{ij}(U_j \\
& -u_i)\delta_j(Z_i-\hat{Z}_i)^T\beta+\bar{\delta}_i(Z_i-\hat{Z}_i)^T\beta.
\end{aligned}$$

Let $\Phi_{KI}(U_i)=\Phi_I(U_i)$,$\Gamma_{KI}=\Gamma_I$. By simple calculation, we have

$$\begin{aligned}
I_1 \;=\; & \frac{1}{n}\sum_{i=1}^{n}(1-\bar{\delta}_i)\sum_{j=1}^{n}K_h^{ij}(U_j-u_i) \\
& \times(1-\delta_j)(Z_i-\Phi_{KI}(U_i)\Gamma_{KI}^{-1}(U_i)X_i)(Z_i- \\
& \Phi_c(U_i)\Gamma_c^{-1}(U_i)X_i)^T\sqrt{n}(\hat{\beta}_c-\beta)+O_p(1) \\
= \;& \Sigma_3\Sigma_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i(Z_i \\
& -X_i^T\Gamma_c^{-1}(U_i)\Phi_c(U_i))\varepsilon_i+O_p(1). \qquad (30)
\end{aligned}$$

Similarly, $I_2=O_p(\sqrt{n}c_n^2)$. The expression of I_3 can be rewritten as

$$\begin{aligned}
I_3 \;=\; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i-\check{Z}_i)(X_i^T\alpha(U_i)-S_{KI_i}M) \\
& -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i-\check{Z}_i)(1-\bar{\delta}_i) \\
& \times\sum_{j=1}^{n}K_h^{ij}(U_j-u_i)(1-\delta_j)S_{KI_j}(\hat{M}_c-M) \\
& -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i-\check{Z}_i)(\bar{\delta}_i+ \\
& (1-\bar{\delta}_i)\sum_{j=1}^{n}K_h^{ij}(U_j-u_i)\delta_j)S_{KI_i}\varepsilon_i \\
& -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i-\check{Z}_i)(1-\bar{\delta}_i) \\
& \times\sum_{j=1}^{n}K_h^{ij}(U_j-u_i)(1-\delta_j)\check{Z}_j(\hat{\beta}_c-\beta), \\
= \;& I_{31}+I_{32}+I_{33}+I_{34}. \qquad (31)
\end{aligned}$$

Since $I_{31}=O_p(1)$, $I_{32}=O_p(1)$, $\frac{1}{n}\sum_{i=1}^{n}(Z_i-\check{Z}_i)S_{KIi}^TX=O_p(1)$, $\frac{1}{n}\sum_{i=1}^{n}(Z_i-\check{Z}_i)\check{Z}_i^T=O_p(1)$, $\beta-\hat{\beta}_c=O_p(n^{-1/2})$. Thus, we have $I_{33}=O_p(1)$, $I_{34}=O_p(1)$, which means that $I_3=O_p(1)$.

For I_4, it follows that

$$\begin{aligned}
I_4 \;=\; & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\bar{\delta}_i)\sum_{j=1}^{n}K_h^{ij}(U_j-u_i)(1-\delta_j)Z_i \\
& \times(X_i^T,0)\{D_{u_i}^TW_{u_i}^{\delta}D_{u_i}\}^{-1}D_{u_i}^TW_{u_i}^{\delta}\varepsilon_i \\
& -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\bar{\delta}_i)\sum_{j=1}^{n}K_h^{ij}(U_j-u_i)(1-\delta_j)\check{Z}_i \\
& \times(X_i^T,0)\{D_{u_i}^TW_{u_i}^{\delta}D_{u_i}^T\}^{-1}D_{u_i}^TW_{u_i}^{\delta}\varepsilon_i, \\
= \;& I_{41}-I_{42}. \qquad (32)
\end{aligned}$$

Similar to the analysis of I_3, we have

$$\begin{aligned}
I_{41} \;=\; & \sum_{j=1}^{n}K_h^{ij}(U_j-u_i)(1-\delta_j) \\
& \times\frac{1}{\sqrt{n}}\{\sum_{i=1}^{n}\Phi_{KI}(U_i)\Gamma_c^{-1}(U_i)X_i\delta_i\varepsilon_i \\
& -\sum_{i=1}^{n}\Phi_c(U_i)\Gamma_c^{-1}(U_i)X_i\delta_i\varepsilon_i\}+O_p(1).
\end{aligned}$$

$$\begin{aligned}
I_{42} \;=\; & \sum_{j=1}^{n}K_h^{ij}(U_j-u_i)(1-\delta_j) \\
& \times\frac{1}{\sqrt{n}}\{\sum_{i=1}^{n}\Phi_{KI}(U_i)\Gamma_c^{-1}(U_i)X_i\delta_i\varepsilon_i \\
& -\sum_{i=1}^{n}\Phi_{KI}(U_i)\Gamma_{KI}^{-1}(U_i)X_i\delta_i\varepsilon_i\}+O_p(1).
\end{aligned}$$

Thus, it satisfies

$$\begin{aligned}
I_4 \;=\; & \sum_{j=1}^{n}K_h^{ij}(U_j-u_i)(1-\delta_j) \\
& \times\frac{1}{\sqrt{n}}\{\sum_{i=1}^{n}(\Phi_{KI}(U_i)\Gamma_{KI}^{-1}(U_i)X_i\delta_i\varepsilon_i, \\
& -\sum_{i=1}^{n}\Phi_c(U_i)\Gamma_c^{-1}(U_i)X_i\delta_i\varepsilon_i\}+O_p(1).
\end{aligned}$$

For I_5, it can be given as

$$\begin{aligned}
I_5 = \;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i-\check{Z}_i)(\bar{\delta}_i\varepsilon_i \\
& +(1-\bar{\delta}_i)\sum_{j=1}^{n}K_h^{ij}(U_j-u_i)\delta_j\varepsilon_j) \\
= \;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i-\check{Z}_i)(\varepsilon_i \qquad (33) \\
& -(1-\bar{\delta}_i)\sum_{j=1}^{n}K_h^{ij}(U_j-u_0)(1-\delta_j)\varepsilon_j) \\
= \;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i-\check{Z}_i)\bar{\delta}_i\sum_{j=1}^{n}K_h^{ij}(U_j-u_i) \\
& \times(1-\delta_j)\varepsilon_j+O_p(1).
\end{aligned}$$

Hence, it shows that

$$\sqrt{n}(\hat{\beta}_{KI} - \beta) = \Xi^{-1}(\Sigma_3 + \check{\Sigma}_1)\Sigma_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i(Z_i - X_i^T\Gamma_c^{-1}(U_i)\Phi_c(U_i))\varepsilon_i + O_p(1). \quad (34)$$

Similar to the proof of the Theorem 3, by the Theorem of Slutsky and the center limit Theorem, we obtain the results of Theorem 5.

Simultaneously, similar to the proof of theorem 5 and theorem 2, we can obtain theorem 8, theorem 6 and theorem 9, respectively, which are omitted here.

Proof of Theorem 7. By the theorem 3 and theorem 5, we have

$$\begin{aligned}
&\Xi^{-1}\Omega_3\Xi^{-1} - \Sigma^{-1}\Omega_2\Sigma^{-1} \\
&= \Sigma^{-1}(\Omega_3 - \Omega_2)\Sigma^{-1} \\
&= \Sigma^{-1}(\sum_{j=1}^{n} K_h^{ij}(U_j - u_i)(1-\delta_j)(\Sigma_2 + \Sigma_1) \\
&\quad \times \Sigma_1^{-1}\Omega_1\Sigma_1^{-1}\sum_{j=1}^{n} K_h^{ij}(U_j - u_i)(1-\delta_j) \\
&\quad \times (\Sigma_2 + \Sigma_1) - (\Sigma_2 + \Sigma_1)\Sigma_1^{-1}\Omega_1\Sigma_1^{-1}(\Sigma_2 + \Sigma_1))\Sigma^{-1} \\
&= \Sigma^{-1}\{[(\sum_{j=1}^{n} K_h^{ij}(U_j - u_i)(1-\delta_j))^2 I - I](\Sigma_2 \\
&\quad + \Sigma_1)\Sigma_1^{-1}\Omega_1\Sigma_1^{-1}(\Sigma_2 + \Sigma_1)\}\Sigma^{-1}.
\end{aligned}$$

As we have denoted $\sum_{j=1}^{n} K_h^{ij}(U_j - u_i) = 1$ in section 4 and the value of δ_j is 1 or 0, thus it is obviously that $(\sum_{j=1}^{n} K_h^{ij}(U_j - u_i)(1-\delta_j))^2 I - I < 0$. Additionally, as we know, the covariance matrix is positive definite. Furtherly, the result of $\Xi^{-1}\Omega_3\Xi^{-1} - \Sigma^{-1}\Omega_2\Sigma^{-1}$ is a negative definite matrix and the theorem 7 is proved.

References:

[1] W. J. Conover, *Practical Nonparametric Statistics*, John–Wiley– Sons–New York 1980

[2] P. J. Green, B. W. Silverman, *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*, Chapman–Hall 1994

[3] H. Midi, M. A. Mohammed ,The Performance of Robust Latent Root Regression Based on M-M and modified GM estimators, *WSEAS. Math. Trans.* 13, 2014, pp. 916-924.

[4] P. Pinto, L. Pinto, The Use of Parametric Statistical Tests to Detect Network Traffic Anomalies on SDN Architectures, *WSEAS. Commu. Trans.* 13, 2014, pp. 659-664.

[5] R. Adeogun, Multipath Parameter Estimation and Channel Prediction for Wideband Mobile to Mobile Wireless Channel. *WSEAS. Commu. Trans.* 13, 2014, pp. 201-208

[6] J. Q. Fan, I. Gijbels, *Local Polynomial Modelling and Its Applications*, Chapman–Hall 1986

[7] A. M. Mood, The Distribution Theory of Certain Nonparametric Two-sample Tests, *Ann. Math. Statist.* 25, 1954, pp. 367–392.

[8] T. J,Hastile, R. J.Tibshirani, Generalized Additive Models, *Monographs on Statistics and Applied Probability*, Chapman–Hall–London 1990

[9] Z. W. Cai, J. Q Fan, Q. W. Yao, Efficient Estimation and Inference for Varying Coefficient Model, *J. Amer. Statist. Assoc.* 95, 2000, pp. 888–902.

[10] R. Chen, R. J. Tasy, Functional-coefficient Autoregressive Models. *J. Amer. Statist. Assoc.* 88, 1993, pp. 298–308.

[11] T. J,Hastile, R. J.Tibshirani, Varying-coefficient Models(with Discussion), *J. Roy. Statist. Soc. Ser. B.* 55,1993, pp. 757–796.

[12] J. Z. Huang, C. O. Wu, L. Zhou, Varying-coefficient Models and Basis Function Approximations for the Analysis of Repeated Measurements, *Biometrika.* 89, 2002, pp. 111–128.

[13] J. Q. Fan, W. Y. Zhang, Statistical Estimation in Varying-coefficient Models, *J. Roy. Statist. Soc. Ser. B.* 27(5), 1999, pp. 1491–1581.

[14] Y. C. Xia, W. K. Li, On the Estimation and Testing of Functional-coefficient Linear Models, *Statist. Sinica.* 9, 1999, 737–757.

[15] D. R. Hoover, J. A. Rice, C. O. Wu, L. P.Yang, Nonparametric Smoothing Estimates of Time-varying Coefficient Models With Longitudinal Data. *Biometrika.*85(4), 1998, pp. 809–822.

[16] T. Hu, Y. C.Xia, Adaptive Semivarying Coefficient Model Selection, *Statistica. Sinica* 22, 2010, pp. 575-599.

[17] J. Z. Huang, C. O. Wu, L. Zhou, Polynominal Spline Estimation and Inference for Varying Coefficient Models with Longitudinal Data, *Statistic Sinica.* 14, 2004, pp. 763–788.

[18] R. F. Engle, W. J. Granger, J. Rice, A. Weiss, Semiparametric Estimates of the Relation Between Weather and Electricity Techniques, *J. Amer. Statist. Assoc.* 80, 1986, pp. 310–319.

[19] H. Chen, Convergence Rates for Parametric Components in a Partially Linear Model, *Ann. Statist.* 88, 1988, pp. 298–308.

[20] A. Yatchew, An Elementary Estimator of the Partial Linear Model, *Economics. Lett.* 57, 1997, pp. 135–143.

[21] H. Liang, Estimation in Partially Linear Models and Numerical Comparisons, *Comput. Statist. Data. Anal.* 50, 2006, pp. 675–687.

[22] P. Speckman, Kernel Smoothing in Partial Linear Models, *J. Roy. Statist. Soc. Ser. B.* 50, 1988, PP. 413–416.

[23] H. Liang, W. Hardle, R. J. Carroll, Estimation in a Semiparametric Partially Linear Errors-in-variables Model, *Ann. Statist.* 27, 1999, pp. 1519–1535.

[24] J. Fan, T. Huang, Profile Likelihood Inferences on Semiparametric Varying-coefficient Partially Linear Models, *Bernoulli* . 11, 2005, pp. 1031–1057.

[25] X. Zhou, J. You, Wavelet Estimation in Varying-coefficient Partially Linear Regression Models, *Statist. Probab. Lett.* 68, 2004, pp. 91–104.

[26] C. H. Wei, X. Z.Wu, Profile Lagrange Multiplier Test for Partially Linear Varying-coefficient Regression Models, *Sys. Sci. Math. Scis.* 28, 2008, pp. 416–424.

[27] W. Zhang, S. Y. Lee, X. Song, Local Polynomial Fitting in Semivarying Coefficient Models, *J. Multi. Anal.* 82, 2002, pp. 166–188.

[28] J. H. You, G. Chen, Estimation of a Semiparametric Varying-coefficient Partially Linear Errors-in-variables Model, *J. Multivar. Anal.* 97, 2004, pp. 324–341.

[29] C. H. Wei, Estimation in Partially Linear Varying-coefficient Errors-in-Variables Models with Missing Responses, *Acta. Mathematic. Scientia.* 30A(4), 2010, pp. 1042–1054.

[30] R. Little, D. Rubin, Statistical Analysis with Missing Data, *Technometrics.* 45(4), 2003, pp. 364–365.

[31] C. Chu, P. Cheng, Nonparametric Regression Estimation with Missing Data, *J. Statist. Planning. Inference.* 48, 1995, PP. 85–99.

[32] Q. Wang, O. Linton, W, Hardle, Semiparametric Regression Analysis with Missing Response at Random.*J. Multivar. Anal.* 98, 2007, pp. 334–345.

[33] H. Zou, The Adaptive Lasso and Its Oracle Properties, *J. Amer. Statist. Assoc.* 2006, 101, pp.1418–1429.

[34] Y. P. Mack, B. W. Silverman, Weak and Strong Uniform Consistency of Kernel Regression Estimates, *Z. Wahrsch. Verw. Gebiete.* 61, 1982, pp. 405–415.