ASYMPTOTIC PROPERTIES FOR ESTIMATORS IN A SEMI-PARAMETRIC MODEL WITH MEASUREMENT ERRORS

Ting-ting Hu & Jing-jing Zhang*
College of Science, University of Shanghai for Science and Technology, Shanghai 200093, PR China
W-mail: 401120716@qq.com, *jenny617@usst.edu.cn

ABSTRACT

In this article, we focus on the semi-parametric error-in-variables model with missing responses: $y_i = \xi_i\beta + g(t_i) + \epsilon_i$, $x_i = \xi_i + \mu_i$, where y_i are the response variables missing at random, (ξ_i, t_i) are design points, ξ_i are the potential variables observed with measurement errors μ_i, the unknown slope parameter β need to be estimate. Here we choose two different approaches to estimate β. Under appropriate conditions, we study the asymptotic normality for the proposed estimators.

Keywords: *Semi-parametric model; Error-in-variables; Missing responses; Asymptotic normality.*
Mathematics Subject Classifications (2020): 62J12; 62N02; 62E99.

1. INTRODUCTION

Consider the following semi-parametric error-in-variables(EV) model

$$\begin{cases} y_i = & \xi_i\beta + g(t_i) + \epsilon_i, \\ x_i = & \xi_i + \mu_i, \end{cases} \tag{1.1}$$

where y_i are the scalar response variables, (ξ_i, t_i) are design points, x_i are observable random variables with measurement errors μ_i, $E\mu_i = 0$, ϵ_i are statistical errors with $E\epsilon_i = 0$. $\beta \in R$ is an unknown parameter waiting to be estimated. $g(\cdot)$ is a unknown function taking values in [0,1], $h(\cdot)$ is a known function defined on [0,1] satisfying

$$\xi_i = h(t_i) + v_i, \tag{1.2}$$

where v_i are also design points.

Model (1.1) has been one of the important issues in statistical research. When $\mu_i \equiv 0$, ξ_i are observed exactly, the model (1.1) reduces to the general semi-parametric model, which was first introduced by [1]. [2] considered marginal generalized semi parametric partially linear models for clustered data. The extensive application of semi-parametric model is of great significance to the development of parameter estimator and estimator efficiency.

More and more attention has been paid to the statistical research of measurement error data. This is because it has appeared in many subjects, including medicine, economics and engineering. There is no doubt that the final results would be biased or inconsistent. When y_i are complete observed and $g(\cdot) \equiv 0$, the model (1.1) reduces to the usual linear EV model. [3] proposed a constrained empirical likelihood confidence region. When $g(\cdot) \neq 0$, the model (1.1) has also been studied by many scholars. [4] establish the consistency and $\sqrt{n}$-normality property of the estimator of the finite-dimensional parameters of the model. Therefore, it is necessary to use corresponding measurement error model and it has been developing constantly.

However, we often encounter incomplete data due to various reasons. Methods to deal with missing data have been widely studied, missing data imputation is the most popular one. Among others, one can impute a plausible value for each missing data, then analyze the results as if they are complete. In regression problems, commonly used imputation approaches include linear regression imputation by [5] nonparametric kernel regression imputation by [6] semi-parametric regression imputation by [7]. We here extend the methods to the estimation of β under the model (1.1). We obtain two approaches to estimate β with missing responses and study the asymptotic normality for the estimators.

In this paper, investigate parameter estimates for models with fixed designs, suppose we obtain a random sample of incomplete data $\{(y_i, \delta_i, x_i, t_i)\}$ from the model (1.1), where δ_i is a number, $\delta_i = 0$ if y_i is missing, otherwise $\delta_i = 1$. We assume that y_i is missing at random. This assumption is a common assumption for statistical analysis with missing data and is reasonable in many practical situations.

The paper is organized as follows. In Section 2, we list some assumptions. The main results are given in Section 3. Some preliminary lemmas are stated in Section 4. Proof of the main results is provided in Section 5.

2. ASSUMPTIONS

In this section, we list some assumptions which will be used in the main results. Here $a_n = O(b_n)$ means $|a_n| \leq |b_n|$, $a_n = o(b_n)$ means $a_n/b_n \to 0$ as $n \to \infty$, while a.s. is stand for almost sure.

(A0) • Let $\{\epsilon_i, 1 \leq i \leq n\}$ and $\{\mu_i, 1 \leq i \leq n\}$ be independent random variables satisfying

- $E\epsilon_i = 0$, $E\mu_i = 0$, $E\mu_i^2 = \Xi_\mu^2 > 0$ is known.
- $\sup_i E|\epsilon_i|^{r_1} < \infty$, $\sup_i E|\mu_i|^{r_2} < \infty$ for some $r_1 > 8/3, r_2 > 4$.
- $\{\epsilon_i, 1 \leq i \leq n\}$ and $\{\mu_i, 1 \leq i \leq n\}$ are independent of each other.

(A1) • Let $\{v_i, 1 \leq i \leq n\}$ in (1.2) be a sequence satisfying

- $\lim_{n\to\infty} n^{-1} \sum_{i=1}^n v_i^2 = \Sigma_0, \lim_{n\to\infty} n^{-1} \sum_{i=1}^n \delta_i v_i^2 = \Sigma_1$ $(0 < \Sigma_0, \Sigma_1 < \infty)$.
- $\lim_{n\to\infty} \sup_n (\sqrt{n}\log n)^{-1} \cdot \max_1 |\sum_{i=1}^m v_{j_i}| < \infty$, where $\{j_1, j_2, \ldots, j_n\}$ is a permutation of $(1,2,\ldots,n)$.
- $\max_1 |v_i| = O(n^{1/2}\log^{-1} n)$.
- $\max_1 |v_i| = O(n^{1/4})$.

(A2) • $g(\cdot)$ and $h(\cdot)$ are continuous functions satisfying the first-order Lipschitz condition on the close interval [0,1].

(A3) • Let $W_{nj}^c(t_i)$ $(1, j)$ be weight functions defined on [0, 1] and satisfy

- $\max_1 \sum_{i=1}^n \delta_j W_{nj}^c(t_i) = O(1)$
- $\max_1 \sum_{j=1}^n \delta_j W_{nj}^c(t_i) I(|t_i - t_j| > a \cdot n^{-1/4}\log^{-1} n) = o(n^{-1/4}\log^{-1} n)$ for any $a > 0$.
- $\max_{1\leq i,j\leq n} W_{nj}^c(t_i) = o(n^{-1/2}\log^{-2} n)$

(A4) • The probability weight functions $W_{nj}(t_i)$ $(1 \leq i, j \leq n)$ are defined on [0,1] and satisfy

- $\max_{1\leq i\leq n} \sum_{i=1}^n W_{nj}(t_i) = O(1)$.
- $\max_{1\leq i\leq n} \sum_{j=1}^n W_{nj}(t_i) I(|t_i - t_j| > a \cdot n^{-1/4}\log^{-1} n) = o(n^{-1/4}\log^{-1} n)$, for any $a > 0$.
- $\max_{1\leq i,j1\leq n} W_{nj}(t_i) = o(n^{-1/2}\log^{-1} n)$.

Remark 2.1 *Conditions (A0)-(A4) are standard regularity conditions and used commonly in the literature, see [8], [9] and [10].*

3. MAIN RESULTS

For model (1.1), we want to seek the estimators of β. The most natural idea is to delete all the missing data. Thus, one can get model $\delta_i y_i = \delta_i \xi_i \beta + \delta_i g(t_i) + \delta_i \epsilon_i$. If ξ_i can be observed, we can apply the least squares estimation(LSE) method to estimate the parameter β. If the parameter β is known, using the complete data $\{(\delta_i y_i, \delta_i x_i, \delta_i t_i), 1 \leq i \leq n\}$, we can define the estimator of $g(\cdot)$ to be

$$g_n^*(t, \beta) = \sum_{j=1}^n W_{nj}^c(t)(\delta_j y_j - \delta_j x_j \beta),$$

where $W_{nj}^c(t)$ are weight functions satisfying (A3). On the other hand, under this condition of the semi-parametric EV model, Liang et al. (1999) improved the LSE on the basis of the usual partially linear model, and employ the estimator of parameter β to minimize the following formula:

$$SS(\beta) = \sum_{i=1}^n \delta_i \{[y_i - x_i\beta - g_n^*(t_i, \beta)]^2 - \Xi_\mu^2 \beta^2\} = min!$$

Therefore, we can achieve the modified LSE of β as follow:

$$\hat{\beta}_c = [\sum_{i=1}^n (\delta_i \tilde{x}_i^{c\,2} - \delta_i \Xi_\mu^2)]^{-1} \sum_{i=1}^n \delta_i \tilde{x}_i^c \tilde{y}_i^c,$$

(3.1)

where $\tilde{x}_i^c = x_i - \sum_{j=1}^n \delta_j W_{nj}^c(t_i)x_j$, $\tilde{y}_i^c = y_i - \sum_{j=1}^n \delta_j W_{nj}^c(t_i)y_j$.

Apparently, the estimators $\hat{\beta}_c$ is formed without taking all sample information into consideration. Hence, in order to make up for the missing data, we imply an imputation method from [7], and let

$$U_i^{[I]} = \delta_i y_i + (1-\delta_i)[x_i\hat{\beta}_c + \hat{g}_n^c(t_i)]. \quad (3.2)$$

Therefore, Using complete data $\{(U_i^{[I]}, x_i, t_i), 1 \le i \le n\}$, similar to (3.1), one can get the other estimators for β,

$$\hat{\beta}_I = [\sum_{i=1}^n (\tilde{x}_i^2 - \Xi_\mu^2)]^{-1} \sum_{i=1}^n \tilde{x}_i \widetilde{U}_i^{[I]} \quad (3.3)$$

where $\widetilde{U}_i^{[I]} = U_i^{[I]} - \sum_{j=1}^n W_{nj}(t_i)U_j^{[I]}$, $\tilde{x}_i = x_i - \sum_{j=1}^n W_{nj}(t_i)x_j$, $W_{nj}(t)$ are weight functions satisfying (A4).

Based on the two estimators for β, we take some notations which will be used and have the following results.

$$\tilde{\xi}_i^c = \xi_i - \sum_{j=1}^n \delta_j W_{nj}^c(t_i)\xi_j, \tilde{\xi}_i = \xi_i - \sum_{j=1}^n W_{nj}^c(t_i)\xi_j, S_n^2 = \sum_{i=1}^n \delta_i(\tilde{\xi}_i^c)^2, R_n^2 = \sum_{i=1}^n (\tilde{\xi}_i)^2,$$

$$\Sigma_n^2 = \mathrm{V}ar\{\sum_{i=1}^n \delta_i[(\tilde{\xi}_i^c + \mu_i)(\epsilon_i - \mu_i\beta) + \Xi_\mu^2\beta]\}, S_{1n}^2 = \sum_{i=1}^n \delta_i(\tilde{x}_i^2 - \Xi_\mu^2), D_{in} = S_{1n}^{-2}(1-\delta_i)\tilde{\xi}_i^2$$

$$\Sigma_{1n}^2 = \mathrm{V}ar\{\sum_{i=1}^n \delta_i[(\tilde{\xi}_i + D_{in}\tilde{\xi}_i^c)(\epsilon_i - \mu_i\beta) + (1+D_{in})(\mu_i\epsilon_i - (\mu_i^2 - \Xi_\mu^2)\beta)]\}$$

Theorem 3.1 *Suppose that (A0)-(A4) are satisfied.*

- If $\Sigma_n^2 \ge Cn$ for all n, then $S_n^2(\hat{\beta}_c - \beta)/\Sigma_n \xrightarrow{\mathcal{D}} N(0,1)$

Theorem 3.2 *Suppose that (A0)-(A4) are satisfied.*

- If $\Sigma_{1n}^2 \ge Cn$ for all n, then $R_n^2(\hat{\beta}_I - \beta)/\Sigma_{1n} \xrightarrow{\mathcal{D}} N(0,1)$

4. PRELIMINARY LEMMAS

In the sequel, let $C, C_1, \cdots$ be some finite positive constants, whose values are unimportant and may change. Now, we introduce several lemmas, which will be used in the proof of the main results.

Lemma 4.1 (Baek ang Liang [12], Lemma 3.1) *Let $\alpha > 2$, $e_1, \cdots, e_n$ be independent random variables with $Ee_i = 0$. Assume that $\{a_{ni}, 1\}$ is a triangular array of numbers with $max_{1\le i\le n}|a_{ni}| = O(n^{-1/2})$ and $\sum_{i=1}^n a_{ni}^2 = o(n^{-2/\alpha}log^{-1}n)$. If $sup_i E|e_i|^p < \infty$ for some $p > 2\alpha/(\alpha-1)$. Then*

$$\sum_{i=1}^n a_{ni}e_i = o(n^{-1/\alpha}) \text{ a.s.}$$

Lemma 4.2 (Hardle et al. [8], Lemma A.3) *Let $V_1, \cdots, V_n$ be independent random variables with $EV_i = 0$, finite variances and $sup_{1\le i\le n}E|V_j|^r \le C < \infty$ $(r > 2)$. Assume that $\{a_{ki}, k, i = 1, \dots, n\}$ is a sequence of numbers such that $sup_{1\le i,k\le n}|a_{ki}| = O(n^{-p_1})$ for some $0 < p_1 < 1$ and $\sum_{j=1}^n a_{ji} = O(n^{p_2})$ for $p_2 max(0, 2/r - p_1)$. Then*

$$\max_{1\le i\le n}|\sum_{k=1}^n a_{ki}V_k| = O(n^{-s}\log n) \text{ a.s.} \quad for\ s = (p_1 - p_2)/2.$$

Lemma 4.3

(a) • Let $\tilde{A}_i = A(t_i) - \sum_{j=1}^n W_{nj}(t_i)A(t_j)$, where $A(\cdot) = g(\cdot)$ or $h(\cdot)$. Let $\tilde{A}_i^c = A(t_i) - \sum_{j=1}^n \delta_j W_{nj}^c(t_i)A(t_j)$, where $A(\cdot) = g(\cdot)$ or $h(\cdot)$. Then, (A0)-(A4) imply that $\max_{1\le i\le n}|\tilde{A}_i| = o(n^{-1/4})$ and $\max_{1\le i\le n}|\tilde{A}_i^c| = o(n^{-1/4})$.

(b) • (A0)-(A4) imply that $n^{-1}\sum_{i=1}^n \tilde{\xi}_i^2 \to \Sigma_0$, $\sum_{i=1}^n |\tilde{\xi}_i| \le C_1 n$, $n^{-1}\sum_{i=1}^n \delta_i(\tilde{\xi}_i^c)^2 \to \Sigma_1$ and $\sum_{i=1}^n |\delta_i\tilde{\xi}_i^c| \le C_2 n$.

(c) • (A0), (A1)(i)(ii)(iii), (A2)-(A4) imply that $\max_1|\tilde{\xi}_i| = O(n^{1/2}\log^{-1}n)$ and

$\max_{1\le i\le n}|\tilde{\xi}_i^c| = O(n^{1/2}\log^{-1} n)$.
(d) • (A0), (A1)(i)(ii)(iv), (A2)-(A4) imply that $\max_{1\le i\le n}|\tilde{\xi}_i| = O(n^{1/4})$ and $\max_1|\tilde{\xi}_i^c| = O(n^{1/4})$.

Lemma 4.4 *Suppose that (A0)-(A4) are satisfied. Then one can deduce that*

$$\max_{1\le i\le n}|\hat{g}_n^c(t_i) - g(t_i)| = o(n^{-\frac{1}{4}}) \text{ a.s.}$$

5. PROOF OF MAIN RESULTS

Firstly, we introduce some notations, which will be used in the proofs below.

$$\tilde{\mu}_i^c = \mu_i - \sum_{j=1}^n \delta_j W_{nj}^c(t_i)\mu_j, \qquad \tilde{\mu}_i = \mu_i - \sum_{j=1}^n W_{nj}(t_i)\mu_j,$$

$$\tilde{g}_i^c = g(t_i) - \sum_{j=1}^n \delta_j W_{nj}^c(t_i) g(t_j), \qquad \tilde{g}_i = g(t_i) - \sum_{j=1}^n W_{nj}(t_i) g(t_j),$$

$$\tilde{\epsilon}_i^c = \epsilon_i - \sum_{j=1}^n \delta_j W_{nj}^c(t_i)\epsilon_j, \qquad \tilde{\epsilon}_i = \epsilon_i - \sum_{j=1}^n W_{nj}(t_i)\epsilon_j, \qquad S_{2n}^2 = \sum_{i=1}^n (\tilde{x}_i^2 - \Xi_\mu^2),$$

Proof of Theorem 3.1. From (1), one can write that

$$\hat{\beta}_c - \beta = S_{1n}^{-2}\{\sum_{i=1}^n [\delta_i(\tilde{\xi}_i^c + \tilde{\mu}_i^c)(\tilde{\epsilon}_i^c - \tilde{\mu}_i^c\beta) + \delta_i \Xi_\mu^2 \beta] + \sum_{i=1}^n \delta_i \tilde{\xi}_i^c \tilde{g}_i^c + \sum_{i=1}^n \delta_i \tilde{\mu}_i^c \tilde{g}_i^c\}$$

$$= S_{1n}^{-2}\{\sum_{i=1}^n \delta_i[(\tilde{\xi}_i^c + \mu_i)(\epsilon_i - \mu_i\beta) + \Xi_\mu^2\beta] + \sum_{i=1}^n \delta_i \tilde{\xi}_i^c \tilde{g}_i^c + \sum_{i=1}^n \delta_i \tilde{\mu}_i^c \tilde{g}_i^c$$

$$+ \sum_{i=1}^n \sum_{j=1}^n \delta_i\delta_j \tilde{\xi}_i^c W_{nj}^c(t_i)\mu_j\beta - \sum_{i=1}^n \sum_{j=1}^n \delta_i\delta_j W_{nj}^c(t_i)\xi_i^c\epsilon_j - \sum_{i=1}^n \sum_{j=1}^n \delta_i\delta_j W_{nj}^c(t_i)\mu_j\epsilon_i$$

$$- \sum_{i=1}^n \sum_{j=1}^n \delta_i\delta_j W_{nj}^c(t_i)\mu_i\epsilon_j + 2\sum_{i=1}^n \sum_{j=1}^n \delta_i\delta_j W_{nj}^c(t_i)\mu_i\mu_j\beta$$

$$+ \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \delta_i\delta_j\delta_k W_{nj}^c(t_i) W_{nk}^c(t_i)\mu_j\epsilon_k - \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \delta_i\delta_j\delta_k W_{nj}^c(t_i) W_{nk}^c(t_i)\mu_j\mu_k\beta\}$$

$$:= S_{1n}^{-2}\sum_{l=1}^{10} A_{ln}. \tag{5.1}$$

Thus, to prove $S_n^2(\hat{\beta}_c - \beta)/\Sigma_n \xrightarrow{\mathcal{D}} N(0,1)$. By Lemma 4.1 - Lemma 4.3 and A(0), one can get $S_{1n}^2/S_n^2 \xrightarrow{a.s.} 1$, so we only need to verify that $\sum_{l=1}^{10} A_{ln}/\Sigma_n \xrightarrow{\mathcal{D}} N(0,1)$.

Step 1. We verify that $A_{ln}/\Sigma_n \xrightarrow{P} 0$ for $l = 2,3,\cdots,10$ with Σ_n^2 . We only need to verify that $A_{ln} = o_p(n^{1/2})$ for $l = 2,3,\cdots,10$ **.** From the conditions of Theorem 3.1 , Lemma 4.3, one can achieve that

$$|A_{2n}| = |\sum_{i=1}^n \delta_i \tilde{\xi}_i^c \tilde{g}_i^c| \le 4|\sum_{i=1}^n \delta_i \tilde{\xi}_i^c \tilde{g}_i^c| = |\sum_{i=1}^n \delta_i \tilde{h}_i^c \tilde{g}_i^c + \sum_{i=1}^n \delta_i v_i \tilde{g}_i^c - \sum_{i=1}^n \sum_{j=1}^n \delta_i\delta_j W_{nj}^c(t_i) v_j \tilde{g}_i^c|$$

$$\leq Cn|\tilde{g}_i^c||\tilde{h}_i^c| + \max_1 \sum_{i=1}^{n} W_{nj}^c(t_i) \max_1 |\sum_{k=1}^{m} v_{j_k}| \max_1 |\tilde{g}_i^c|$$
$$+Cn \cdot \max_1 |\tilde{g}_i^c||\delta_i| \max_1 |\sum_{i=1}^{m} v_{k_i}| = o(n^{\frac{1}{2}}) \text{ a.s.}$$

$$E(A_{3n})^2 = E(\sum_{i=1}^{n} \delta_i \tilde{\mu}_i^c \tilde{g}_i^c)^2 \leq C \cdot \{E(\sum_{i=1}^{n} \delta_i \tilde{g}_i^c \mu_i)^2 + E[\sum_{i=1}^{n} \delta_i \tilde{g}_i^c \sum_{j=1}^{n} \delta_j W_{nj}(t_i)\mu_j]^2\}$$

$$\leq C \cdot \{\sum_{i=1}^{n} (\tilde{g}_i^c)^2 + \sum_{j=1}^{n} \delta_j^2 [\sum_{i=1}^{n} W_{nj}(t_i)\tilde{g}_i^c]^2\} = o(n^{\frac{1}{2}})$$

which yields that $A_{3n} = o_p(n^{1/2})$. On the other hand,note that,

$$A_{4n} = \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \delta_j \tilde{\xi}_i^c W_{nj}^c(t_i)\mu_j \beta$$
$$= \sum_{i=1}^{n} \delta_i \tilde{h}_i^c \sum_{j=1}^{n} \delta_j W_{nj}^c(t_i)\mu_j \beta + \sum_{i=1}^{n} \delta_i v_i \sum_{j=1}^{n} \delta_j W_{nj}^c(t_i)\mu_j \beta$$
$$- \sum_{i=1}^{n} \delta_i \sum_{s=1}^{n} W_{ns}^c(t_i)\delta_s v_s \sum_{j=1}^{n} \delta_j W_{nj}^c(t_i)\mu_j \beta$$
$$:= D_{1n} + D_{2n} + D_{3n}$$

According to the condition that Σ_n^2 ,we can achieve that

$$E(D_{1n})^2 \leq C \cdot E[\sum_{i=1}^{n} \delta_i \tilde{h}_i^c \sum_{j=1}^{n} \delta_j W_{nj}^c(t_i)\mu_j \beta]^2 \leq C \cdot \sum_{j=1}^{n} E[\sum_{i=1}^{n} \delta_i \tilde{h}_i^c \delta_j W_{nj}^c(t_i)]^2$$
$$\leq Cn \cdot [\sum_{i=1}^{n} \delta_i \tilde{h}_i^c W_{nj}^c(t_i)]^2 = o(n^{\frac{1}{2}})$$
$$E(D_{2n})^2 \leq C \cdot E[\sum_{i=1}^{n} \delta_i v_i \sum_{j=1}^{n} \delta_j W_{nj}^c(t_i)\mu_j \beta]^2 \leq C \cdot \sum_{j=1}^{n} [\delta_j \sum_{i=1}^{n} \delta_i v i W_{nj}^c(t_i)\beta]^2$$
$$\leq Cn \cdot \{\max_{1,j} W_{nj}^c(t_i) \cdot \max_{1\leq i\leq n} |\sum_{k=1}^{m} v_{k_i}|^2\} = o(n)$$
$$|D_{3n}| = |\sum_{i=1}^{n} \delta_i \sum_{s=1}^{m} W_{ns}(t_i)\delta_s v_s \sum_{j=1}^{n} \delta_j W_{nj}^c(t_i)\mu_j \beta|$$
$$\leq C \cdot \max_{1\leq i\leq n} |\sum_{s=1}^{m} v_{k_s}| \cdot \max_{1\leq i\leq n} |\sum_{i=1}^{n} \delta_i W_{ns}(t_i)| \cdot \max_{1\leq i\leq n} |\sum_{j=1}^{n} \delta_j W_{nj}^c(t_i)\mu_j|$$
$$= o(n^{\frac{1}{2}} \log n) \cdot O(1) \cdot o(n^{-\frac{1}{4}}) = o(n^{\frac{1}{4}} \log n) = o_p(n^{\frac{1}{2}})$$

It follows that $A_{4n} = o_p(n^{1/2})$. Similarly, one can achieve that

$$E(A_{6n})^2 \leq C \cdot E[\sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \delta_j W_{nj}^c(t_i)\mu_j \epsilon_i]^2 \leq \sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i^2 \delta_j^2 [W_{nj}^c(t_i)]^2 E\mu_j^2 E\epsilon_i^2 = o(n^{\frac{1}{2}} \log^{-1} n)$$
$$E(A_{8n})^2 \leq C \cdot E[\sum_{i=1}^{n} \sum_{j=1}^{n} \delta_i \delta_j W_{nj}^c(t_i)\mu_i \mu_j \beta]^2$$
$$\leq C \cdot \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} [\delta_{i_1}\delta_{i_2}\delta_{j_1}\delta_{j_2} W_{nj_1}(t_{i_1}) W_{nj_2}(t_{i_2})] E(\mu_{i_1}\mu_{i_2}\mu_{j_1}\mu_{j_2})$$

$$\leq C\cdot[\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_j^2W_{nj}^2(t_i)E(\mu_i^2\mu_j^2)+\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}W_{ni_1}(t_{i_1})W_{ni_2}(t_{i_2})E(\mu_{i_1}^2\mu_{i_2}^2)]=O(1)$$

Which leads to that $A_{6n}=o_p(n^{1/2})$ $A_{8n}=o_p(n^{1/2})$. Similarly, one achieve that $A_{kn}=o_p(n^{1/2})$ for $k=5,7,9,10$.

Therefore, we can get that $A_{ln}/\Sigma_n \xrightarrow{P} 0$ for $l=2,3,\cdots,10$

Step 2. We verify that $A_{1n}/\Sigma_n \xrightarrow{\mathcal{D}} N(0,1)$. According to (??), *Step 1*, we conclude that

$$S_{1n}^2(\hat{\beta}_c-\beta)=A_{1n}+\sum_{l=2}^{10}A_{ln}=\sum_{i=1}^{n}\eta_{in}+o_p(1).$$

From the condition that Σ_n^2, Lemma 4.3, taking $r=\min\{r_1,r_2/2\}>2$ in (A0) and arbitrary $\eta>0$, as $n\to\infty$, one can verify that

$$\frac{1}{n}\cdot\sum_{i=1}^{n}E[\eta_{in}^2\cdot I(|\eta_{in}|>\eta\cdot n^{\frac{1}{2}})]\leq\frac{C}{n}\cdot\sum_{i=1}^{n}E|\eta_{in}|^r\cdot I(|\eta_{in}|>\eta\cdot n^{\frac{1}{2}})(\eta\cdot n^{\frac{1}{2}})^{-(r-2)}$$
$$\leq\frac{C}{n}\cdot\sum_{i=1}^{n}[E|\delta_i\tilde{\xi}_i^c\epsilon_i|^r+E|\delta_i\tilde{\xi}_i^c\mu_i\beta|^r+E|\delta_i\mu_i\epsilon_i|^r+E|\delta_i(\mu_i^2-\Xi_\mu^2)\beta|^r](\eta\cdot n^{\frac{1}{2}})^{-(r-2)}$$
$$\leq\frac{C_1}{n}\sum_{i=1}^{n}(\tilde{\xi}_i^c)^2\max_{1\leq i\leq n}|\tilde{\xi}_i^c|^{r-2}n^{-\frac{r-2}{2}}+C_2n^{-\frac{r-2}{2}}=o(1).$$

This means that Lindeberg¡¯s Condition for the Central Limit Theorem is supportive. Denote that $\eta_{in}=\delta_i[(\tilde{\xi}_i^c+\mu_i)(\epsilon_i-\mu_i\beta)+\Xi_\mu^2\beta]$, $E(\eta_{in})=0$

$$E(\eta_{in}^2)=E\delta_i^2[(\tilde{\xi}_i^c+\mu_i)(\epsilon_i-\mu_i\beta)+\Xi_\mu^2\beta]^2$$
$$\leq C\cdot E\{(\tilde{\xi}_i^c\epsilon_i)^2+(\tilde{\xi}_i^c\mu_i\beta)^2+(\mu_i\epsilon_i)^2+\left(\mu_i^2-\Xi_\mu^2\right)^2\beta^2\}\leq\infty$$

Therefore, η_{in} is a dependent and random variables sequence with $E\eta_{in}=0$ and $Var(\sum_{i=1}^{n}\eta_{in})=\Sigma_n^2$. Thus, the proof of Theorem 3.1 is completed.

Proof of Theorem 3.2. From (3.2) and (5.1), write that

$$S_{2n}^2(\hat{\beta}_I-\beta)=\sum_{i=1}^{n}\tilde{x}_i[\tilde{U}_i^{[I]}-\tilde{x}_i\beta]+\sum_{i=1}^{n}\Xi_\mu^2\beta$$
$$=\sum_{i=1}^{n}\delta_i[(\tilde{\xi}_i+D_{in}\tilde{\xi}_i^c)(\epsilon_i-\mu_i\beta)+(1+D_{in})(\mu_i\epsilon_i-(\mu_i^2-\Xi_\mu^2)\beta)]+S_{1n}^{-2}\sum_{i=1}^{n}(1-\delta_i)\tilde{\xi}_i^2\sum_{l=2}^{10}A_{ln}$$
$$-\sum_{i=1}^{n}(1-\delta_i)\tilde{\xi}_i[g(t_i)-\hat{g}_n^c(t_i)]-\sum_{i=1}^{n}(1-\delta_i)\mu_i[g(t_i)-\hat{g}_n^c(t_i)]$$
$$-[\sum_{i=1}^{n}\mu_i\sum_{j=1}^{n}W_{nj}(t_i)\delta_j(\epsilon_j-\mu_j\beta)-(1-\delta_i)\Xi_\mu^2\beta]-\sum_{i=1}^{n}\tilde{\xi}_i\sum_{j=1}^{n}W_{nj}(t_i)\delta_j(\epsilon_j-\mu_j\beta)$$
$$+\sum_{i=1}^{n}\tilde{\xi}_i\sum_{j=1}^{n}W_{nj}(t_i)(1-\delta_j)[g(t_i)-\hat{g}_n^c(t_i)]+\sum_{i=1}^{n}\mu_i\sum_{j=1}^{n}W_{nj}(t_i)(1-\delta_j)[g(t_i)-\hat{g}_n^c(t_i)]$$
$$\sum_{i=1}^{n}\tilde{\xi}_i\tilde{g}_i+\sum_{i=1}^{n}\mu_i\tilde{g}_i-\sum_{i=1}^{n}\sum_{k=1}^{n}W_{nk}(t_i)\mu_k\delta_i(\epsilon_i-\mu_i\beta)$$
$$+\sum_{i=1}^{n}\sum_{k=1}^{n}W_{nk}(t_i)\mu_k(1-\delta_i)[g(t_i)-\hat{g}_n^c(t_i)]+\sum_{i=1}^{n}\sum_{k=1}^{n}W_{nk}(t_i)\mu_k\sum_{j=1}^{n}W_{nj}(t_i)\delta_j(\epsilon_i-\mu_i\beta)$$
$$-\sum_{i=1}^{n}\sum_{k=1}^{n}W_{nk}(t_i)\mu_k\sum_{j=1}^{n}W_{nj}(t_i)(1-\delta_j)[g(t_i)-\hat{g}_n^c(t_i)]-\sum_{i=1}^{n}\sum_{k=1}^{n}W_{nk}(t_i)\mu_k\tilde{g}_i$$

$$+\sum_{i=1}^{n}(1-\delta_i)\tilde{\xi}_i\sum_{j=1}^{n}W_{nj}(t_i)\xi_j(\hat{\beta}_c-\beta)+\sum_{i=1}^{n}(1-\delta_i)\tilde{\xi}_i\mu_i(\hat{\beta}_c-\beta)$$
$$+\sum_{i=1}^{n}(1-\delta_i)\xi_i\mu_i(\hat{\beta}_c-\beta)+\sum_{i=1}^{n}(1-\delta_i)(\mu_i-\Xi_\mu^2)(\hat{\beta}_c-\beta)$$
$$-\sum_{i=1}^{n}\tilde{\xi}_i\sum_{j=1}^{n}W_{nj}(t_i)(1-\delta_j)\xi_j(\hat{\beta}_c-\beta)-\sum_{i=1}^{n}\mu_i\sum_{j=1}^{n}W_{nj}(t_i)(1-\delta_j)\xi_j(\hat{\beta}_c-\beta)$$
$$-\sum_{i=1}^{n}\tilde{\xi}_i\sum_{j=1}^{n}W_{nj}(t_i)(1-\delta_j)\mu_j(\hat{\beta}_c-\beta)-\sum_{i=1}^{n}\mu_i\sum_{j=1}^{n}W_{nj}(t_i)(1-\delta_j)\mu_j(\hat{\beta}_c-\beta)$$
$$-\sum_{i=1}^{n}\sum_{k=1}^{n}W_{nk}(t_i)\mu_k(1-\delta_i)\xi_i(\hat{\beta}_c-\beta)-\sum_{i=1}^{n}\sum_{k=1;k\neq i}^{n}W_{nk}(t_i)\mu_k(1-\delta_i)\mu_i(\hat{\beta}_c-\beta)$$
$$-\sum_{i=1}^{n}W_{ni}(t_i)(1-\delta_j)(\mu_i^2-\Xi_\mu^2)(\hat{\beta}_c-\beta)+\sum_{i=1}^{n}\sum_{k=1}^{n}W_{nk}(t_i)\mu_k\sum_{j=1}^{n}W_{nj}(t_i)(1-\delta_j)\xi_j(\hat{\beta}_c-\beta)$$
$$+\sum_{i=1}^{n}\sum_{k=1}^{n}W_{nk}(t_i)\mu_k\sum_{j=1}^{n}W_{nj}(t_i)(1-\delta_j)\mu_j(\hat{\beta}_c-\beta)$$
$$:=\sum_{l=1}^{15}B_{ln}+\sum_{l=16}^{28}B_{ln}\cdot(\hat{\beta}_c-\beta).$$

Similar to the proof of Theorem 3.1, one can deduce that $S_{2n}^2/R_n^2\xrightarrow{a.s.}1$ and $\hat{\beta}_c-\beta=O_p(n^{-\frac{1}{2}})$. According to the above, to prove $R_n^2(\hat{\beta}_I-\beta)/\Sigma_{1n}\overset{\mathcal{D}}{\to}N(0,1)$, we only need to verify that $B_{ln}=o_p(n^{1/2})$ for $l=2,3,\cdots,15$, $B_{kn}=o_p(n)$ for $k=16,17,\cdots,28$ and $B_{1n}/\Sigma_{1n}\overset{\mathcal{D}}{\to}N(0,1)$.

Step 1. We verify that $B_{ln}=o_p(n^{1/2})$ for $l=2,3,\cdots,15$ and $B_{kn}=o_p(1)$ for $k=16,17,\cdots,28$. From the proof of Theorem 3.1 , we can get that $B_{2n}=o_p(n^{1/2})$. According to the Lemmas 4.1-4.4, we deduce that,

$$|B_{3n}|=|\sum_{i=1}^{n}(1-\delta_i)[\tilde{h}_i+v_i-\sum_{j=1}^{n}W_{nj}(t_i)v_j][g(t_i)-\hat{g}_n^c(t_i)]|$$
$$\leq C\cdot|\sum_{i=1}^{n}(1-\delta_i)\tilde{h}_i[g(t_i)-\hat{g}_n^c(t_i)]|+C\cdot|\sum_{i=1}^{n}(1-\delta_i)v_i[g(t_i)-\hat{g}_n^c(t_i)]|$$
$$-C\cdot|\sum_{i=1}^{n}[g(t_i)-\hat{g}_n^c(t_i)](1-\delta_i)\sum_{j=1}^{n}W_{nj}(t_i)v_j|=o(n^{\frac{1}{2}})\ \text{a.}s.$$
$$|B_{4n}|\ \leq C\cdot|\sum_{i=1}^{n}(1-\delta_i)\mu_i[g(t_i)-\hat{g}_n^c(t_i)]|=O(n^{\frac{1}{4}}\log n)\ \text{a.}s.$$
$$E(B_{5n})^2\ \leq C\cdot E\{\sum_{j=1}^{n}[\sum_{i=1}^{n}\tilde{\xi}_iW_{nj}(t_i)\delta_j](\epsilon_i-\mu_j\beta)\}^2$$
$$\leq C\cdot\sum_{j=1}^{n}[\sum_{i=1}^{n}\tilde{\xi}_iW_{nj}(t_i)\delta_j]^2E(\epsilon_j-\mu_j\beta)^2=O(n^{\frac{1}{2}}\log n)$$

$$|B_{8n}| \le C|\sum_{i=1}^{n}\mu_i\sum_{j=1}^{n}W_{nj}(t_i)\delta_j[g(t_j)-\hat{g}_n^c(t_j)]|$$

$$=C|\sum_{i=1}^{n}\mu_i W_{nj}(t_i)|\cdot\sum_{j=1}^{n}|g(t_j)-\hat{g}_n^c(t_j)|=o(n^{\frac{1}{2}})\text{ a.}s.$$

$$|B_{9n}| \le C|\sum_{i=1}^{n}\tilde{\xi}_i\tilde{g}_i|=Cn^{-\frac{1}{2}}|\sum_{i=1}^{n}(\tilde{h}_i+v_i-\sum_{j=1}^{n}W_{nj}(t_i)\tilde{g}_i)|=o(n^{\frac{1}{2}})\text{ a.}s.$$

$$E(B_{10n})^2 \le C\cdot E(\sum_{i=1}^{n}\mu_i\tilde{g}_i)^2=C\cdot\sum_{i=1}^{n}E\mu_i^2\tilde{g}_i^2=o(n^{\frac{1}{2}})$$

$$E(B_{11n})^2 \le C\cdot E[\sum_{i=1}^{n}\sum_{k=1}^{n}W_{nk}(t_i)\mu_k\delta_i(\epsilon_i-\mu_i\beta)]^2$$

$$=C\cdot E[\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}W_{nk_1}(t_{i_1})W_{nk_2}(t_{i_2})\mu_{k_1}\mu_{k_2}\delta_{i_1}\delta_{i_2}(\epsilon_{i_1}-\mu_{i_1}\beta)(\epsilon_{i_2}-\mu_{i_2}\beta)]$$

$$\le C\sum_{i=1}^{n}\delta_i^2E(\epsilon_i-\mu_i\beta)^2\sum_{k=1}^{n}W_{n_k}^2(t_i)E(\mu_k)^2$$

$$+C\sum_{i_1=1}^{n}W_{n_{i_1}}(t_i)\delta_{i_1}E(\mu_{i_1})E(\epsilon_{i_1}-\mu_{i_1}\beta)\cdot\sum_{i_2=1}^{n}W_{n_{i_2}}(t_{i_2})\delta_{i_2}E(\mu_{i_2})E(\epsilon_{i_2}-\mu_{i_2}\beta)$$

$$=o(n^{\frac{1}{2}}\log^{-1}n)$$

$$E(B_{15n})^2 \le C\cdot E[\sum_{i=1}^{n}\sum_{k=1}^{n}W_{nk}(t_i)\mu_k\tilde{g}_i]^2\le C\cdot\sum_{k=1}^{n}[\sum_{i=1}^{n}W_{nk}(t_i)\tilde{g}_i]^2E(\mu_k)^2$$

$$\le C\cdot\sum_{k=1}^{n}(O(1)\cdot\tilde{g}_i^2)=o(n^{\frac{1}{2}})$$

In the same way, we can verify that $B_{ln}=o_p(n^{1/2})$ for $l=6,7,\cdots,16$ and $B_{kn}=o_p(1)$ for $k=17,\cdots,28$.

Step 2. We verify that $B_{1n}/\Sigma_{1n}\overset{\mathcal{D}}{\to}N(0,1)$. Let $B_{1n}=\sum_{i=1}^{n}\gamma_{in}$, where $\gamma_{in}=\delta_i[(\tilde{\xi}_i+D_{in}\tilde{\xi}_i^c)(\epsilon_i-\mu_i\beta)+(1+D_{in})(\mu_i\epsilon_i-(\mu_i^2-\Xi_\mu^2)\beta)]$. As a result, γ_{in} is a dependent and random variables sequence with $E\gamma_{in}=0$ and $Var(\sum_{i=1}^{n}\gamma_{in})=\Sigma_{1n}^2$. Meanwhile, we can deduce that $|\sum_{i=1}^{n}\tilde{\xi}_i|=o(n^{3/4})$, $|\sum_{i=1}^{n}v_i|=O(n^{1/2}\log n)$, $|\sum_{j=1}^{n}W_{nj}(t_i)\xi_j|=O(1)$ and $D_{in}=o(1)$ by the A(0)-A(4). From the condition that Σ_{1n}^2, Lemma 4.3, taking $r=\min\{r_1,r_2/2\}>2$ in (A0) and arbitrary $\gamma>0$, as $n\to\infty$, one can verify that

$$\frac{1}{n}\cdot\sum_{i=1}^{n}E[\gamma_{in}^2\cdot I(|\gamma_{in}|>\gamma\cdot n^{\frac{1}{2}})] \le\frac{C}{n}\cdot\sum_{i=1}^{n}E|\gamma_{in}|^r\cdot I(|\gamma_{in}|>\gamma\cdot n^{\frac{1}{2}})(\gamma\cdot n^{\frac{1}{2}})^{-(r-2)}$$

$$\le\frac{C}{n}\cdot\sum_{i=1}^{n}[E|\delta_i(\tilde{\xi}_i+D_{in}\tilde{\xi}_i^c)(\epsilon_i-\mu_i\beta)|^r+E|\delta_i\mu_i\epsilon_i|^r+E|\delta_i(\mu_i^2-\Xi_\mu^2)|^r\beta^r](\gamma\cdot n^{\frac{1}{2}})^{-(r-2)}$$

$$\leq \frac{C}{n}\cdot\sum_{i=1}^{n}[E|\delta_i\tilde{\xi}_i\epsilon_i|^r+E|\delta_i\tilde{\xi}_i\mu_i\beta|^r+E|\delta_i\tilde{\xi}_i^c\epsilon_i|^r+E|\delta_i\tilde{\xi}_i^c\mu_i\beta|^r+E|\delta_i\mu_i\epsilon_i|^r$$

$$+E|\delta_i(\mu_i^2-\Xi_\mu^2)|^r\beta^r](\gamma\cdot n^{\frac{1}{2}})^{-(r-2)}$$

$$\leq \frac{C_1}{n}\sum_{i=1}^{n}(\tilde{\xi}_i)^2\max_{1\leq i\leq n}|\tilde{\xi}_i|^{r-2}n^{-\frac{r-2}{2}}+\frac{C_2}{n}\sum_{i=1}^{n}(\tilde{\xi}_i^c)^2\max_{1\leq i\leq n}|\tilde{\xi}_i^c|^{r-2}n^{-\frac{r-2}{2}}+C_3n^{-\frac{r-2}{2}}=o(1).$$

The $Lindeberg\ Condition$ is workable. Therefore, one can get $B_{1n}/\Sigma_{1n}\xrightarrow{\mathcal{D}}N(0,1)$. Thus, the proof of Theorem 3.2 is completed.

Acknowledgments

The authors greatly appreciate the constructive comments and suggestions of the Editor and referee. This research was supported by the National Natural Science Foundation of China (11701368).

References

[1] Engle, R. F., Granger, C. W. J., Rice, J. and Weiss, A.: Semiparametric estimation of the relation between weather and electricity sales. *Journal of the American Statistical Association.* 81(1986), 310-320.
[2] N. Wang, R.J. Carroll, X. Lin: Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association.* 100 (2005) 147-157.
[3] Cui, H.J., Chen, S.X.: Empirical likelihood confidence region for parameter in the errors-in-variable models. *Journal of Multivariate Analysis* 84 (2003) 101-115.
[4] Ahmad, I., Leelahanon, S., Li, Q.: Efficient estimation of a semiparametric partially linear varying coefficient model. *The Analysis of Statistics* 33 (2005) 258-283.
[5] Healy, M.J.R. and Westmacott, M.: Missing values in experiments analysis on automatic computers. *Appl. Statist.* 5 (1956), 203-206.
[6] Cheng, P.E.: Nonparametric estimation of mean functionals with data missing at random. *J. Amer. Statist. Assoc.* 89 (1994), 81-87.
[7] Wang, Q. and Sun, Z.: Estimation in partially linear models with missing responses at random. *J. Multivariate Anal.* 98 (2007) 1470-1493.
[8] Härdle, W., Liang, H. and Gao, J. T.: Partially Linear Models. *Physica-Verlag, Heidelberg* 2000.
[9] Gao, J. T., Chen, X. R. and Zhao, L. C.: Asymptotic normality of a class of estimators in partial linear models. *Acta. Math. Sinica* 37(2) (1994) 256-268.
[10] Chen, H.: Convergence rates for parametric components in a partly linear model. *Ann. Statist.* 16 (1988) 136-146.
[11] Liang, H., Härdle, W. and Carrol, R. J.: Estimation in a semiparametric partially linear errosr-in-variables model. *The Annals of Statistics* 27(5) (1999) 1519-1935.
[12] Baek, J. I. and Liang, H. Y.: Asymptotic of estimators in semi-parametric model under NA samples. *Journal of Statist. Plann. Infer.* 136 (2006) 3362-3382.