

REVIEW

Open Access



Jackknifing for partially linear varying-coefficient errors-in-variables model with missing response at random

Yuye Zou^{1,2} and Chengxin Wu^{3*} 

*Correspondence:

hsuchengxinwu@foxmail.com

³School of Mathematics and Statistics, Huangshan University, Huangshan, 245041, China

Full list of author information is available at the end of the article

Abstract

In this paper, we focus on the response mean of the partially linear varying-coefficient errors-in-variables model with missing response at random. A simulation study is conducted to compare jackknife empirical likelihood method with normal approximation method in terms of coverage probabilities and average interval lengths, and a comparison of the proposed estimators is done based on their biases and mean square errors.

MSC: Primary 62G10; secondary 62G08

Keywords: Asymptotically normal; Errors-in-variables; Jackknife empirical likelihood; Missing at random; Partially linear varying-coefficient model

1 Introduction

In statistical problems, various forms of statistical models are widely sought by many researchers. As a natural compromise between the parametric models and the nonparametric models, semi-parametric regression models allow some predictors to be modeled parametrical and others being modeled non-parametrical, which motivates us to consider the following partially linear varying-coefficient (PLVC) model:

$$Y = X^\top \beta + W^\top \alpha(U) + \varepsilon, \quad (1.1)$$

where Y is a response variable, $(X, W) \in \mathbb{R}^p \times \mathbb{R}^q$ are covariates, $\beta = (\beta_1, \dots, \beta_p)^\top$ is a vector of p -dimensional unknown parameters, $\alpha(\cdot)$ is an unknown q -dimensional vector of the coefficient function, ε is a random error with $E(\varepsilon|X, W, U) = 0$. To avoid the curse of dimensionality, we assume that U is univariate. As one combination of partially linear model and varying-coefficient model, PLVC model has drawn much attention. For example, Kai et al. [9] discussed the variable selection by the composite quantile regression method. He et al. [6] developed an approximate estimator of the functional coefficients by B-spline function and studied the asymptotic properties of the proposed estimators. Shen and Liang [17] considered weighted quantile regression and variable selection un-

© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

der righted-censored data with missing censoring indicators. For more work see Fan and Huang [4], You and Zhou [29], Huang and Zhang [7] among others.

In this paper, we are interested in estimating the mean of response Y , say θ , in model (1.1) when the covariate X is measured with error. We use the surrogate ξ instead of observing X . Hence, we assume the following additive errors-in-variables (EV) model:

$$\xi = X + e, \quad (1.2)$$

where e is the measurement error with zero mean and covariance matrix Σ_e (which is known). The combination of (1.1) with (1.2) is named the partially linear varying-coefficient errors-in-variables (PLVC EV) model, which has been studied by many authors. For example, Fan et al. [2] considered the penalized empirical likelihood and variable selection for high-dimension data. Liu and Liang [12] constructed the asymptotical normality of jackknife estimator for error variance and standard chi-square distribution of jackknife empirical log-likelihood statistic. Fan et al. [3] established penalized profile least squares estimation of parameter and non-parameter in the model. Xu et al. [24] proved the asymptotic properties of the proposed estimators for parameter and coefficient function, and studied asymptotic distribution of empirical log-likelihood ratio function for parameter with missing covariates.

In many practical fields, however, not all response variables can be available for various reasons. For instance, in public opinion poll, non-response is a typical source of missing values. Due to the presence of missing data, the traditional and standard inference procedures cannot be applied directly. A common approach to dealing with missing data is the complete case (CC) analysis, which only uses data with complete observations and is in the loss of information when the missing mechanism is missing at random (MAR). To eliminate this disadvantage, the imputation method is one method of filling in the missing response, which includes linear regression (Yates [27], Wang and Rao [21, 23]), kernel regression imputation (Cheng [1], Wang and Rao [22]), ratio imputation (Rao [16]) and so on. These methods are widely used by many statisticians. Wang et al. [20] proposed an imputation estimator and a number of propensity score weighting estimators, which are consistent and asymptotically normal. Liang [10] extended the idea of Wang et al. [20] to consider partially linear regression model with error-prone covariates. Xue [25] used a weighted linear regression imputation to construct a weighted-corrected empirical likelihood ratio of the response mean so that the ratio has an asymptotic chi-squared distribution. Tang and Zhao [18] proposed imputed empirical likelihood-based estimator for the response mean of the nonlinear regression models.

Throughout this paper, we are interested in inference of the mean of response Y with the missing response at random in the PLVC EV model. Hence, we obtain the following incomplete observations $\{Y, W, U, \delta, \xi\}$, where ξ , W and U are observed, Y may be missing, and $\delta = 0$ if Y is missing, otherwise $\delta = 1$. We assume that Y is MAR, which implies that δ and Y are conditionally independent given X , W and U , that is, $P(\delta = 1|Y, X, W, U) = P(\delta = 1|X, W, U) := P(Z)$ with $Z = (X, W, U)$ and the probability function $P(\cdot)$ represents the heterogeneity in the missingness mechanism. The MAR assumption is common in statistical analysis with missing data and is reasonable in many practical situations; see Little and Rubin [11].

As is well known, the empirical likelihood, introduced by Owen [13, 14], has many advantages over the normal approximation and bootstrap approximation for constructing

the confidence intervals. For example, the empirical likelihood method does not involve the variance estimation because of the complicated variance estimation. Meanwhile, the sharp and orientation of confidence regions based on the empirical likelihood method are determined entirely by the data. However, the estimation based on empirical likelihood method will be computationally difficult and the Wilks theorem does not hold in general. In order to handle the situation where nonlinear statistics are involved, Jing et al. [8] proposed a new approach called jackknife empirical likelihood. Thanks to its advantages, the jackknife empirical likelihood approach has been applied by many researchers. See Gong et al. [5], Peng et al. [15], Yang and Zhao [26], Liu and Liang [12], Yu and Zhao [30] and so on. However, there is a little literature considering the jackknife method for response mean with missing response at random.

In this paper, we are interested in the statistical inference of the mean of response Y in the PLVC EV model with missing response at random, especially the confidence regions of the response mean. In order to avoid the difficulty of calculation and ensure that the Wilks phenomenon is established, we consider the jackknife empirical likelihood method instead of empirical likelihood method. In the spirit of Wang et al. [20], we propose the marginal average estimator, the regression imputation estimator and the augmented inverse probability estimator of the response mean by imputing every missing response variable. At the same time, the corresponding jackknife estimators of the response mean are defined. The estimators are consistent and asymptotical normality under some assumptions. We also establish the asymptotic distribution of the jackknife empirical log-likelihood ratio function and construct the confidence regions. A simulation study is done to evaluate the performance of the proposed methods.

The rest of this paper is organized as follows. In Sect. 2, we give the methodologies and build the estimators. The main results are listed in Sect. 3. A simulation study is conducted in Sect. 4. Our conclusion is drawn in Sect. 5. The proofs of the main results and some lemmas are provided in the [Appendix](#).

2 Methodology

2.1 Estimation

For convenience, we assume that the X_i is directly observable. The estimators of parameter β and coefficient function $\alpha(\cdot)$ can be obtained by profile least squares method as follows. Having multiplied model (1.1) by the observation indicators, then we have

$$\delta_i Y_i = \delta_i X_i^\top \beta + \delta_i W_i^\top \alpha(U_i) + \delta_i \varepsilon_i. \quad (2.1)$$

For given β , we apply the local weighted least squared method to estimate the coefficient function $\{\alpha_j(\cdot), j = 1, \dots, q\}$. For u in a small neighborhood of u_0 , the Taylor expansion for $\alpha_j(u)$ can be written as

$$\alpha_j(u) = \alpha_j(u_0) + \alpha'_j(u_0)(u - u_0) := a_j + b_j(u - u_0).$$

We minimize the following objective function to get $\{(a_j, b_j), j = 1, \dots, q\}$:

$$\sum_{i=1}^n \left\{ Y_i - X_i^\top \beta - \sum_{j=1}^q (a_j + b_j(u - u_0)) W_{ij} \right\}^2 K_{h_n}(U_i - u) \delta_i,$$

where $K_{h_n}(\cdot) = \frac{1}{h_n}K(\cdot/h_n)$ is a kernel function and $0 < h_n \rightarrow 0$ is a bandwidth sequence.

Let $\mathbf{X} = (X_1, X_2, \dots, X_n)^\top$, $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$, $\mathbf{W} = (W_1, W_2, \dots, W_n)^\top$, $\omega_u^\delta = \text{diag}(\delta_1 K_{h_n}(U_1 - u), \dots, \delta_n K_{h_n}(U_n - u))$ and

$$M = \begin{pmatrix} W_1^\top \alpha(U_1) \\ \vdots \\ W_n^\top \alpha(U_n) \end{pmatrix}, \quad D_u = \begin{pmatrix} W_1 & \frac{U_1 - u}{h_n} W_1^\top \\ \vdots & \vdots \\ W_n & \frac{U_n - u}{h_n} W_n^\top \end{pmatrix}.$$

Therefore, when β is known, we can obtain the estimator of $\alpha(\cdot)$ by

$$\tilde{\alpha}(u, \beta) = (I_q \ 0_q) \{D_u^\top \omega_u^\delta D_u\}^{-1} D_u^\top \omega_u^\delta (\mathbf{Y} - \mathbf{X}\beta). \quad (2.2)$$

Substituting (2.2) into (2.1) and eliminating the bias produced by the measurement errors, we give the modified profile least squared estimator of β as follows:

$$\hat{\beta}_n = \left\{ \sum_{i=1}^n \delta_i (\tilde{\xi}_i \tilde{\xi}_i^\top - \Sigma_e) \right\}^{-1} \sum_{i=1}^n \delta_i \tilde{\xi}_i \tilde{Y}_i,$$

where $S_i = (W_i^\top \ 0)(D_u^\top \omega_u^\delta D_u)^{-1} D_u^\top \omega_u^\delta$, $\tilde{Y}_i = Y_i - S_i \mathbf{Y}$, $\tilde{\xi}_i = \xi_i^\top - S_i \xi$ with $\xi = (\xi_1, \dots, \xi_n)^\top$. Hence, one can get the following local linear regression estimator of $\alpha(\cdot)$:

$$\hat{\alpha}_n(u) = (I_q \ 0)(D_u^\top \omega_u^\delta D_u)^{-1} D_u^\top \omega_u^\delta (\mathbf{Y} - \xi \hat{\beta}_n),$$

By Wang et al. [20], we consider the response mean θ by the following general class of estimators:

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\delta_i}{P_n^*(Z_i)} Y_i + \left(1 - \frac{\delta_i}{P_n^*(Z_i)} \right) (\xi_i^\top \hat{\beta}_n + W_i^\top \hat{\alpha}_n(U_i)) \right\}, \quad (2.3)$$

where $P_n^*(z)$ is some sequence of quantities with probability limits $P(z)$. When $P_n^*(z) = \infty$, $\hat{\theta}_n$ reduces to the following marginal average estimator:

$$\hat{\theta}_n^{(1)} = \frac{1}{n} \sum_{i=1}^n \{ \xi_i^\top \hat{\beta}_n + W_i^\top \hat{\alpha}_n(U_i) \}. \quad (2.4)$$

$P_n^*(z) = 1$, $\hat{\theta}_n$ reduces to the following regression imputation estimator:

$$\hat{\theta}_n^{(2)} = \frac{1}{n} \sum_{i=1}^n \{ \delta_i Y_i + (1 - \delta_i) (\xi_i^\top \hat{\beta}_n + W_i^\top \hat{\alpha}_n(U_i)) \}. \quad (2.5)$$

When $P_n^*(z) = \hat{P}_n(z)$, $\hat{\theta}_n$ reduces to the following augmented inverse probability estimator:

$$\hat{\theta}_n^{(3)} = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\delta_i}{\hat{P}_n(Z_i)} Y_i + \left(1 - \frac{\delta_i}{\hat{P}_n(Z_i)} \right) (\xi_i^\top \hat{\beta}_n + W_i^\top \hat{\alpha}_n(U_i)) \right\}, \quad (2.6)$$

where $\hat{P}_n(z) = \frac{\sum_{i=1}^n \delta_i \Omega_{b_n}(z - Z_i)}{\sum_{i=1}^n \Omega_{b_n}(z - Z_i)}$ with kernel function $\Omega_{b_n}(\cdot) = \frac{1}{b_n} \Omega(\cdot/b_n)$ and bandwidth sequence $0 < b_n \rightarrow 0$.

2.2 Jackknife empirical likelihood

In order to avoid the covariance matrix estimation, this subsection proposes the jackknife empirical likelihood method to construct the confidence regions for θ . Let $\hat{\beta}_{n,-i}$ be the estimator of β when the i th observation is deleted, that is,

$$\hat{\beta}_{n,-i} = \left\{ \sum_{j \neq i}^n \delta_j (\tilde{\xi}_j \tilde{\xi}_j^\top - \Sigma_e) \right\}^{-1} \sum_{j \neq i}^n \delta_j \tilde{\xi}_j \tilde{Y}_j.$$

Note that the definitions of $\hat{\theta}_n^{(k)}$ ($k = 1, 2, 3$) in (2.4)–(2.6), one can re-write $\hat{\theta}_n^{(1)} = \frac{1}{n} \times \sum_{i=1}^n \{S_i Y_i + \tilde{\xi}_i^\top \hat{\beta}_n\}$, $\hat{\theta}_n^{(2)} = \frac{1}{n} \sum_{i=1}^n \{\delta_i Y_i + (1 - \delta_i)(S_i Y_i + \tilde{\xi}_i^\top \hat{\beta}_n)\}$ and $\hat{\theta}_n^{(3)} = \frac{1}{n} \sum_{i=1}^n \{\frac{\delta_i}{P_n(Z_i)} Y_i + (1 - \frac{\delta_i}{P_n(Z_i)})(S_i Y_i + \tilde{\xi}_i^\top \hat{\beta}_n)\}$. Let $\hat{\theta}_{n,-i}^{(k)}$ be the estimator of θ when the i th observation is deleted for $k = 1, 2, 3$, which are defined by $\hat{\theta}_{n,-i}^{(1)} = \frac{1}{n-1} \sum_{j \neq i}^n \{S_j Y_j + \tilde{\xi}_j^\top \hat{\beta}_{n,-i}\}$, $\hat{\theta}_{n,-i}^{(2)} = \frac{1}{n-1} \sum_{j \neq i}^n \{\delta_j Y_j + (1 - \delta_j)(S_j Y_j + \tilde{\xi}_j^\top \hat{\beta}_{n,-i})\}$ and $\hat{\theta}_{n,-i}^{(3)} = \frac{1}{n-1} \sum_{j \neq i}^n \{\frac{\delta_j}{P_{n,-i}(Z_j)} Y_j + (1 - \frac{\delta_j}{P_{n,-i}(Z_j)})(S_j Y_j + \tilde{\xi}_j^\top \hat{\beta}_{n,-i})\}$, where $\hat{P}_{n,-i}(\cdot)$ is the estimator of $P(\cdot)$ when the i th observation is deleted, that is, $\hat{P}_{n,-i}(z) = \sum_{j \neq i}^n \delta_j \Omega(\frac{z - Z_j}{b_n}) / \sum_{j \neq i}^n \Omega(\frac{z - Z_j}{b_n})$. Then we have the i th jackknife pseudo samples $\hat{\theta}_{j_i}^{(k)} = n\hat{\theta}_n^{(k)} - (n-1)\hat{\theta}_{n,-i}^{(k)}$ and the jackknife estimators of θ are defined as follows:

$$\hat{\theta}_J^{(k)} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{j_i}^{(k)} = n\hat{\theta}_n^{(k)} - \frac{n-1}{n} \sum_{i=1}^n \hat{\theta}_{n,-i}^{(k)}.$$

Hence, the following jackknife empirical likelihoods of θ are constructed based on the jackknife pseudo-samples:

$$L^{(k)}(\theta) = \sup \left\{ \prod_{i=1}^n (np_i) : p_1 > 0, \dots, p_n > 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \hat{\theta}_{j_i}^{(k)} = \theta \right\}. \quad (2.7)$$

Using the Lagrange multipliers, we get the jackknife empirical log-likelihood ratio functions

$$l^{(k)}(\theta) = 2 \sum_{i=1}^n \log \{1 + \lambda^\top (\hat{\theta}_{j_i}^{(k)} - \theta)\}, \quad (2.8)$$

where λ is the solution to the equation

$$\frac{1}{n} \sum_{i=1}^n \frac{\hat{\theta}_{j_i}^{(k)} - \theta}{1 + \lambda^\top (\hat{\theta}_{j_i}^{(k)} - \theta)} = 0. \quad (2.9)$$

3 Main results

Throughout this paper, let C denote finite positive constants, whose values may change in different scenarios.

- (1) Let $\mu_k = \int u^k K(u) du$ and $\gamma_n = \{\log n / (nh_n)\}^{1/2} + h_n^2$.
- (2) Let $A = E[\delta_1(X_1 - W_1^\top \Gamma^{-1}(U_1)\Phi(U_1))]$, $\Sigma_1 = E[\xi_1 - W_1^\top \Gamma^{-1}(U_1)\Phi(U_1)]$ and $\Sigma_2 = E[(1 - \delta_1)(\xi_1 - W_1^\top \Gamma^{-1}(U_1)\Phi(U_1))]$.
- (3) Let $\Gamma(U_1) = E[\delta_1 W_1 W_1^\top | U_1]$, $\Phi(U_1) = E[\delta_1 X_1 W_1 | U_1]$,

$$\begin{aligned} \Lambda_1 &= E\{X_1^\top \beta + W_1^\top \alpha(U_1) - \theta\}^2 \\ &\quad + E\{e_1^\top \beta + A^{-1} \Sigma_1 \delta_1 (\xi_1^\top - W_1^\top \Gamma^{-1}(U_1)\Phi(U_1))(\varepsilon_1 - e_1^\top) + \Sigma_e \beta\}^2, \end{aligned}$$

$$\begin{aligned}\Lambda_2 &= E\{X_1^\top \beta + W_1^\top \alpha(U_1) - \theta\}^2 \\ &\quad + E\{(1 - \delta_1)e_1^\top \beta + \delta_1 \varepsilon_1 + A^{-1} \Sigma_2 \delta_1 (\xi_1^\top - W_1^\top \Gamma^{-1}(U_1) \Phi(U_1)) (\varepsilon_1 - e_1^\top) \\ &\quad + \Sigma_e \beta\}^2, \\ \Lambda_3 &= E\{X_1^\top \beta + W_1^\top \alpha(U_1) - \theta\}^2 + E\left\{\frac{\delta_1}{P(Z_1)} \varepsilon_1 + e_1^\top \beta\right\}^2.\end{aligned}$$

In order to formulate the main results, we need to impose the following assumptions.

- (A1) The random variable U has bounded support \mathcal{U} and its density function $g(\cdot)$ is Lipschitz continuous and far away from zero. The density function of Z , $f(z)$, is bounded away from zero and has bounded continuous second derivatives.
- (A2) The matrix $\Gamma(U_1)$ is nonsingular for each $U_1 \in \mathcal{U}$. $\Gamma(U_1)$, $E(\delta_1 X_1 X_1^\top | U_1)$ and $\Phi(U_1)$ are Lipschitz continuous.
- (A3) There is one $s > 2$ such that $E(\|X_1\|^{2s} | U_1) < \infty$ a.s., $E(\|W_1\|^{2s} | U_1) < \infty$ a.s., $E(\|\xi_1\|^{2s}) < \infty$ a.s. and $E(\|\varepsilon_1\|^{2s} | X_1, W_1) < \infty$ a.s.
- (A4) The coefficient functions $\{\alpha_j(\cdot), j = 1, 2, \dots, q\}$ have continuous second derivatives.
- (A5) $P(z)$ has bounded partial derivatives up to order 2 almost surely and $\inf_z P(z) > 0$.
- (A6) $K(t)$ is a bounded kernel function of order 2 with bounded support, and has bounded partial derivatives up to order 2 almost surely.
- (A7) $\Omega(\cdot)$ is bounded kernel function of order $r (> 2)$ with bounded support and has a bounded partial derivative.
- (A8) The bandwidths h_n and b_n satisfy $nh_n^8 \rightarrow 0$, $nh_n^2/(\log n)^2 \rightarrow \infty$, $nb_n^{2(p+q+1)}/\log n \rightarrow \infty$ and $nb_n^{2r} \rightarrow 0$.

Remark 3.1 Assumptions (A1)–(A4) are standard conditions, which are commonly used in the literature; see Fan and Huang [4], Liu and Liang [12]. Assumption (A5) is always applied on missing data analysis; see Wang et al. [20]. Assumptions (A6) and (A7) are used in the investigation on some nonparametric kernel estimators. Assumption (A8) implies the relationship between sample size and bandwidths.

We consider the asymptotic normality of the profile least square estimators and jack-knife estimator of the response mean in Theorems 3.1 and 3.2. Also, we give the asymptotic distributions of $l^{(k)}(\theta)$ for $k = 1, 2, 3$ in Theorem 3.3 and construct the confidence regions of θ .

Theorem 3.1 Suppose that Assumptions (A1)–(A8) hold, then for $k = 1, 2, 3$ we have

$$\sqrt{n}(\hat{\theta}_n^{(k)} - \theta) \xrightarrow{\mathcal{D}} N(0, \Lambda_k).$$

Theorem 3.2 Suppose that the assumptions of Theorem 3.1 hold, then, for $k = 1, 2, 3$, we have $\sqrt{n}(\hat{\theta}_j^{(k)} - \theta) = \sqrt{n}(\hat{\theta}_n^{(k)} - \theta) + o_p(1)$. Further, we have $\sqrt{n}(\hat{\theta}_j^{(k)} - \theta) \xrightarrow{\mathcal{D}} N(0, \Lambda_k)$.

Theorem 3.3 Suppose that the assumptions of Theorem 3.1 hold. If θ is the true value, then for $k = 1, 2, 3$ we have

$$l^{(k)}(\theta) \xrightarrow{\mathcal{P}} \chi_1^2,$$

where χ_1^2 is independent standard chi-square random variables with 1 degree of freedom.

Remark 3.2 From Theorem 3.3, it follows immediately that an approximation $1 - \tau$ confidence regions for θ are given by $I_\tau = \{\theta : l^{(k)}(\theta) \leq \chi_1^2(\tau)\}$, where $\chi_1^2(\tau)$ is the upper τ -quantile of the distribution of χ_1^2 . In view of Theorem 3.2, one can construct the confidence regions for θ by estimating the variance Λ_k . The jackknife empirical likelihood method does not relate to an estimation of the asymptotic variance, which makes it more efficient than the normal approximation method. This phenomenon is also exhibited in a simulation study.

4 Simulation study

In this section, we carry out some simulations to demonstrate the finite sample performance of the profile least square estimators and jackknife estimators by comparing their bias and mean square error (MSE). Besides, we compare the jackknife empirical likelihood (JEL) method with normal approximation (NA) in terms of the coverage probability (CP) and average interval length (AL).

The data are generated from the following PLVC EV model:

$$\begin{cases} Y_i = X_{i1}\beta_1 + X_{i2}\beta_2 + W_i\alpha(U_i) + \varepsilon_i, \\ \xi_{i1} = X_{i1} + e_{i1}, \quad \xi_{i2} = X_{i2} + e_{i2}, \quad i = 1, \dots, n, \end{cases} \quad (4.1)$$

where $X_{i1} \sim N(0, 1)$, $X_{i2} \sim N(0, 1)$, $W_i \sim N(0, 1)$, $U_i \sim U(0, 1)$, $\beta_1 = 1$, $\beta_2 = 2$, $\alpha(u) = 2\sin(6\pi u)$ and $\varepsilon_i \sim N(0, 1)$. The measurement error $e_i \sim N(0, \Sigma_e)$. To represent different levels of measurement errors, we take Σ_e as $\Sigma_{e1} = \text{diag}(0.25, 0.25)$ and $\Sigma_{e2} = \text{diag}(0.5, 0.5)$ in the simulations, respectively. The kernel functions $K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1)$ and $\Omega(x, w, u) = \Omega_1(x)\Omega_2(w)\Omega_3(u)$ where $\Omega_1(t) = \Omega_2(t) = \Omega_3(t) = \frac{15}{16}(1 - t^2)^2I(|t| \leq 1)$. The bandwidth $h_n = n^{-1/5}$ and $b_n = n^{-1/3}$.

We generate 500 Monte Carlo random samples of size 50, 100, 150 and 60, 90, 120 based on the following six cases, repeatedly.

- (1) Case 1: $P(x, w, u) = 1/[1 + \exp(-X_{i1} - X_{i2} - W_i - U_i - 5.5)]$;
- (2) Case 2: $P(x, w, u) = 1/[1 + \exp(-X_{i1} - 0.5X_{i2} - W_i - U_i - 2)]$;
- (3) Case 3: $P(x, w, u) = 1/[1 + \exp(-0.75X_{i1} - X_{i2} - W_i - U_i - 1)]$;
- (4) Case 4: $P(x, w, u) = 1/[1 + \exp(-0.5X_{i1} - 0.5X_{i2} - 5W_i - 5.5U_i - 1)]$;
- (5) Case 5: $P(x, w, u) = 1 - 1/[1 + \exp(-0.5X_{i1} - 0.5X_{i2} - 5W_i - 5.5U_i - 1)]$;
- (6) Case 6: $P(x, w, u) = 1 - 1/[1 + \exp(-X_{i1} - X_{i2} - W_i - 2.4U_i - 0.5)]$.

The average missing rate (MR) corresponding to the above six cases are 10%, 20%, 30%, 45%, 55%, 65%, respectively.

In Tables 1–2, we calculate biases and MSEs of $\hat{\theta}_n^{(k)}$ and $\hat{\theta}_j^{(k)}$ for $k = 1, 2, 3$, respectively, to evaluate their finite sample performance. The simulation results indicate the following conclusions. The larger MRs and/or measurement errors produce bigger biases and MSEs. The biases and MSEs decrease as the sample size increases. Both biases and MSEs of $\hat{\theta}_j^{(k)}$ are smaller than those of $\hat{\theta}_n^{(k)}$ under the same settings. In other words, the jackknife estimators $\hat{\theta}_j^{(k)}$ perform better than $\hat{\theta}_n^{(k)}$. Besides, the augment inverse probability estimator $\hat{\theta}_n^{(3)}$ performs best, and $\hat{\theta}_n^{(1)}$ is worst. The corresponding jackknife estimators enjoy the same conclusion.

Table 1 Biases of the estimators $\hat{\theta}_n^{(k)}$ and $\hat{\theta}_j^{(k)}$ with different sample sizes and MRs

Σ_e	MR	n	$\hat{\theta}_n^{(1)}$	$\hat{\theta}_n^{(2)}$	$\hat{\theta}_n^{(3)}$	$\hat{\theta}_j^{(1)}$	$\hat{\theta}_j^{(2)}$	$\hat{\theta}_j^{(3)}$
Σ_{e1}	10%	50	0.0112	0.0040	0.0038	0.0112	0.0039	0.0036
		100	0.0107	0.0016	0.0015	0.0104	0.0013	0.0011
		150	0.0072	0.0004	0.0004	0.0071	0.0003	0.0003
	20%	50	-0.0121	-0.0072	-0.0073	-0.0119	-0.0060	-0.0062
		100	0.0072	0.0053	0.0054	0.0064	0.0049	0.0052
		150	0.0019	0.0023	0.0022	0.0020	0.0023	0.0020
	30%	50	-0.0128	-0.0087	-0.0083	-0.0095	-0.0056	-0.0046
		100	0.0127	0.0075	0.0073	0.0123	0.0067	0.0062
		150	0.0080	0.0026	0.0024	0.0081	0.0024	0.0021
Σ_{e2}	10%	50	0.0134	0.0042	0.0038	0.0132	0.0034	0.0033
		100	-0.0122	0.0029	0.0028	-0.0122	0.0028	0.0026
		150	-0.0054	0.0023	0.0021	-0.0031	0.0017	0.0013
	20%	50	-0.0174	-0.0082	-0.0078	-0.0154	-0.0081	0.0080
		100	-0.0131	0.0089	0.0079	-0.0101	0.0065	0.0065
		150	0.0038	-0.0048	-0.0049	0.0043	-0.0047	-0.0044
	30%	50	-0.0134	-0.0089	-0.0089	-0.0103	-0.0062	-0.0063
		100	0.0130	0.0083	0.0083	0.0129	0.0078	0.0076
		150	0.0086	0.0023	0.0022	0.0085	0.0020	0.0019

Table 2 MSEs of the estimators $\hat{\theta}_n^{(k)}$ and $\hat{\theta}_j^{(k)}$ with different sample sizes and MRs

Σ_e	MR	n	$\hat{\theta}_n^{(1)}$	$\hat{\theta}_n^{(2)}$	$\hat{\theta}_n^{(3)}$	$\hat{\theta}_j^{(1)}$	$\hat{\theta}_j^{(2)}$	$\hat{\theta}_j^{(3)}$
Σ_{e1}	10%	50	0.0734	0.0091	0.0091	0.0910	0.0088	0.0088
		100	0.0376	0.0046	0.0047	0.0463	0.0046	0.0047
		150	0.0257	0.0029	0.0029	0.0284	0.0029	0.0029
	20%	50	0.0761	0.0237	0.0237	0.0919	0.0219	0.0218
		100	0.0408	0.0095	0.0095	0.0479	0.0095	0.0096
		150	0.0276	0.0059	0.0059	0.0309	0.0058	0.0059
	30%	50	0.0787	0.0351	0.0351	0.1040	0.0338	0.0338
		100	0.0422	0.0136	0.0137	0.0491	0.0135	0.0136
		150	0.0272	0.0092	0.0092	0.0301	0.0092	0.0093
Σ_{e2}	10%	50	0.1122	0.0177	0.0178	0.2202	0.0163	0.0163
		100	0.0520	0.0057	0.0057	0.0589	0.0053	0.0053
		150	0.0353	0.0039	0.0039	0.0376	0.0037	0.0037
	20%	50	0.0809	0.0223	0.0225	0.0878	0.0234	0.0237
		100	0.0511	0.0137	0.0137	0.0566	0.0128	0.0128
		150	0.0309	0.0075	0.0075	0.0334	0.0072	0.0072
	30%	50	0.0707	0.0356	0.0355	0.1110	0.0504	0.0504
		100	0.0555	0.0192	0.0192	0.0602	0.0171	0.0171
		150	0.0356	0.0117	0.0117	0.0381	0.0112	0.0112

In Tables 3–4, we give the CPs and ALs for JEL method and NA method on response mean θ . The CPs for JEL method and NA method decrease as MRs, measurement errors decrease and confidence levels increase, and increase as the sample size increases. Besides, the JEL method outperforms the NA method in terms of coverage probability. The CPs for JEL method are larger than those of NA method under the same settings. For both methods, when we have MRs, measurement errors and confidence levels increase, the ALs are getting longer. When the sample size increases, the ALs are getting shorter. The ALs for JEL method are larger than those of NA method under the same settings.

Table 3 CP and AL for JEL method and NA method on θ with confidence level 90%

Σ_e	MR	n	CPJ	CPN	ALJ	ALN
Σ_{e1}	45%	60	0.9640	0.9440	1.6167	1.2151
		90	0.9760	0.9580	1.1201	0.9600
		120	0.9840	0.9620	0.9358	0.8329
	55%	60	0.9600	0.9200	1.8986	1.2511
		90	0.9660	0.9520	1.2045	0.9862
		120	0.9740	0.9560	0.9655	0.8389
	65%	60	0.9520	0.9200	2.4630	1.6925
		90	0.9640	0.9460	1.3061	1.0236
		120	0.9660	0.9440	0.9778	0.8527
Σ_{e2}	45%	60	0.9340	0.9080	3.0637	1.4192
		90	0.9700	0.9240	1.5820	1.0535
		120	0.9760	0.9420	1.1577	0.8839
	55%	60	0.9220	0.8800	4.3876	2.0454
		90	0.9560	0.9220	2.0063	1.1089
		120	0.9720	0.9340	1.2351	0.9158
	65%	60	0.8740	0.8560	6.8074	4.0820
		90	0.9480	0.9180	2.5783	1.3314
		120	0.9740	0.9320	1.3804	0.9752

Table 4 CP and AL for JEL method and NA method on θ with confidence level 95%

Σ_e	MR	n	CPJ	CPN	ALJ	ALN
Σ_{e1}	45%	60	0.9780	0.9680	1.9361	1.4522
		90	0.9880	0.9800	1.3399	1.1473
		120	0.9920	0.9880	1.1155	0.9954
	55%	60	0.9760	0.9540	2.2712	1.4953
		90	0.9840	0.9800	1.4437	1.1787
		120	0.9900	0.9820	1.1528	1.0026
	65%	60	0.9680	0.9640	2.8843	2.0228
		90	0.9800	0.9780	1.5600	1.2234
		120	0.9880	0.9800	1.1668	1.0190
Σ_{e2}	45%	60	0.9560	0.9440	3.5274	1.6961
		90	0.9860	0.9640	1.8857	1.2591
		120	0.9920	0.9720	1.3799	1.0564
	55%	60	0.9440	0.9260	4.9603	2.4445
		90	0.9740	0.9540	2.3540	1.3253
		120	0.9820	0.9640	1.4737	1.0945
	65%	60	0.9320	0.8940	7.4867	4.8785
		90	0.9700	0.9520	2.9689	1.5912
		120	0.9780	0.9620	1.6301	1.1655

5 Conclusion

In this paper, we focus on the response mean of the PLVC EV model with missing response at random. Inspired by Wang et al. [20], we propose the marginal average estimator, the regression imputation estimator and the augmented inverse probability estimator of the response mean to deal with the missing response variable. In order to construct the confidence regions of the response mean, we define the corresponding jackknife estimators and establish the jackknife empirical log-likelihood ratio functions of the response mean. Meanwhile, the consistency and asymptotical normality of the estimators are proved under some assumptions. We also establish the asymptotic chi-square distribution of the jackknife empirical log-likelihood ratio functions and construct the confidence regions for the estimators of the response mean. Finally, one simulation study is conducted to

compare jackknife empirical likelihood method with normal approximation method in terms of coverage probabilities and average interval lengths, and one comparison of the proposed estimators is done based on their biases and mean square errors.

Appendix

To prove Theorems 3.1–3.3, we need the following lemmas.

Lemma A.1 *Suppose that Assumptions (A1)–(A8) hold, then, as $n \rightarrow \infty$, we have*

$$\sup_{u \in \mathcal{U}} \left| \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{U_i - u}{h_n}\right) \left(\frac{U_i - u}{h_n}\right)^\top W_i W_i^\top \right| = g(u) \Gamma(u) \mu_k + O_p(\gamma_n) \quad a.s., \quad (\text{A.1})$$

$$\sup_{u \in \mathcal{U}} \left| \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{U_i - u}{h_n}\right) \left(\frac{U_i - u}{h_n}\right)^\top W_i \varepsilon_i \right| = \left(\frac{\log n}{nh_n} \right)^{1/2} \quad a.s. \quad (\text{A.2})$$

Proof The proof of Lemma A.1 is similar to that of Lemma A.2 in You and Chen [28]. \square

Lemma A.2 *Suppose that Assumptions (A1)–(A8) hold, then we have*

$$\sup_z |\widehat{P}_n(z) - P(z)| = O_p((nb_n^{p+q+1})^{-1/2} + b_n^r).$$

Proof The proof of Lemma A.2 is similar to Eq. (31) in Wang [19]. \square

Lemma A.3 *Suppose that Assumptions (A1)–(A8) hold, then we have*

$$\sqrt{n}(\widehat{\beta}_n - \beta) = A_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \{ \Sigma_e \beta + [\xi_i - \Phi^\top(U_i) \Gamma^{-1}(U_i) W_i] (\varepsilon_i - e_i^\top \beta) \} + o_p(1).$$

Proof Let $A_n = \frac{1}{n} \sum_{i=1}^n \delta_i (\widetilde{\xi}_i \widetilde{\xi}_i^\top - \Sigma_e)$ and $e = (e_1, \dots, e_n)^\top$, we have

$$\begin{aligned} \widehat{\beta}_n - \beta &= A_n^{-1} \frac{1}{n} \sum_{i=1}^n \delta_i \widetilde{\xi}_i^\top (\widetilde{Y}_i - \widetilde{\xi}_i^\top \beta) + A_n^{-1} \frac{1}{n} \sum_{i=1}^n \delta_i \Sigma_e \beta \\ &= A_n^{-1} \frac{1}{n} \sum_{i=1}^n \delta_i \widetilde{\xi}_i^\top (\varepsilon_i - S_i \varepsilon) + A_n^{-1} \frac{1}{n} \sum_{i=1}^n \delta_i \widetilde{\xi}_i^\top (M_i - S_i M) \\ &\quad + \frac{1}{n} \sum_{i=1}^n \delta_i \widetilde{\xi}_i^\top (S_i e - e_i)^\top \beta + A_n^{-1} \frac{1}{n} \sum_{i=1}^n \delta_i \Sigma_e \beta \\ &:= T_1 + T_2 + T_3 + T_4. \end{aligned}$$

From (A.1), (A.3) and Assumption (A5), one simple computation yields $S_i \varepsilon = W_i^\top \times \mathbf{1}_q O_p(\sqrt{\frac{\log n}{nh_n}})$, where $\mathbf{1}_q = (1, \dots, 1_q)$. Then one can get

$$T_1 = A_n^{-1} \frac{1}{n} \sum_{i=1}^n \delta_i (\xi_i - \Phi^\top(U_i) \Gamma^{-1}(U_i) W_i) \varepsilon_i + o_p(n^{-1/2}).$$

Note that $M_i - S_i M = W_i^\top \alpha(U_i) O_p(\gamma_n)$, then $T_2 = o_p(n^{-1/2})$. Based on (A.1), one simple calculation yields

$$S_i X = W_i^\top \Gamma^{-1}(U_i) \Phi(U_i) (1 + O_p(\gamma_n)). \quad (\text{A.3})$$

Since e is independent of (Y, X, W, U) with zero mean, it can be checked that

$$S_i e = W_i^\top \Gamma^{-1}(U_i) E(W_i e_i^\top | U_i) (1 + O_p(\gamma_n)) = 0. \quad (\text{A.4})$$

Hence, simple arguments suggest that

$$T_3 = A_n^{-1} \frac{1}{n} \sum_{i=1}^n \delta_i (\xi_i - \Phi^\top(U_i) \Gamma^{-1}(U_i) W_i) e_i \beta + o_p(n^{-1/2}).$$

Therefore, collecting the results above, Lemma A.3 is proved. \square

Lemma A.4 Suppose that Assumptions (A1)–(A8) hold, then we have

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n W_i^\top (\hat{\alpha}_n(U_i) - \alpha(U_i)) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_i^\top \ 0) (D_u^\top \omega_u^\delta D_u)^{-1} D_u^\top \omega_u^\delta X_i^\top (\beta - \hat{\beta}_n) + o_p(1). \end{aligned}$$

Proof From the definition of $\hat{\alpha}_n(\cdot)$, then one can get

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n W_i^\top (\hat{\alpha}_n(U_i) - \alpha(U_i)) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \{ (W_i^\top \ 0) (D_u^\top \omega_u^\delta D_u)^{-1} D_u^\top \omega_u^\delta (Y - \xi \hat{\beta}_n) - W_i^\top \alpha(U_i) \} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_i^\top \ 0) (D_u^\top \omega_u^\delta D_u)^{-1} D_u^\top \omega_u^\delta X_i^\top (\beta - \hat{\beta}_n) \\ &\quad - \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_i^\top \ 0) (D_u^\top \omega_u^\delta D_u)^{-1} D_u^\top \omega_u^\delta e_i^\top \hat{\beta}_n \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_i^\top \ 0) (D_u^\top \omega_u^\delta D_u)^{-1} D_u^\top \omega_u^\delta \varepsilon_i \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \{ (W_i^\top \ 0) (D_u^\top \omega_u^\delta D_u)^{-1} D_u^\top \omega_u^\delta M_i - M_i \} \\ &:= D_1 + D_2 + D_3 + D_4. \end{aligned}$$

Since $S_i e = 0$, we have $D_2 = o_p(1)$. Due to $S_i \varepsilon = W_i \mathbf{1}_q O_p(\sqrt{\frac{\log n}{nh_n}})$, we have $D_3 = o_p(1)$. Since $S_i M - M_i = W_i^\top \alpha(U_i) O_p(\gamma_n)$, it follows that $D_4 = o_p(1)$. Hence, we have completed the proof of Lemma A.4. \square

Lemma A.5 Suppose that Assumptions (A1)–(A8) hold, then we have

$$\frac{(n-1)^2}{n} \sum_{i=1}^n (\hat{\theta}_n^{(k)} - \hat{\theta}_{n,-i}^{(k)})^2 \xrightarrow{\mathcal{P}} \Lambda_k.$$

Proof (a) We prove Lemma A.5 for $\hat{\theta}_n^{(1)}$. Note that

$$\begin{aligned} \hat{\theta}_n^{(1)} - \hat{\theta}_{n,-i}^{(1)} &= \frac{1}{n-1} [\tilde{\xi}_i^\top (\hat{\beta}_n - \beta) + (W_i^\top \alpha(U_i) + \xi_i^\top \beta - \theta)] + \frac{1}{n-1} \sum_{j=1}^n \tilde{\xi}_j^\top (\hat{\beta}_n - \hat{\beta}_{n,-i}) \\ &\quad - \frac{1}{n-1} \tilde{\xi}_i^\top (\hat{\beta}_n - \hat{\beta}_{n,-i}) - \frac{1}{n-1} (\hat{\theta}_n^{(1)} - \theta) \\ &\quad + \frac{1}{n-1} (S_i M - M_i) + \frac{1}{n-1} S_i (\varepsilon - e^\top \beta) \\ &:= b_{1i} + b_{2i} + b_{3i} + b_{4i} + b_{5i} + b_{6i}. \end{aligned}$$

From the proof of (a) in Theorem 3.1, it is easy to prove

$$\frac{(n-1)^2}{n} \sum_{i=1}^n b_{1i}^2 = \frac{1}{n} \sum_{i=1}^n \{ \tilde{\xi}_i^\top (\hat{\beta}_n - \beta) + (W_i^\top \alpha(U_i) + \xi_i^\top \beta - \theta) \}^2 \xrightarrow{\mathcal{P}} \Lambda_1.$$

Following the fact $\max_{1 \leq i \leq n} \|\hat{\beta}_n - \hat{\beta}_{n,-i}\| = O_p(n^{-1})$ and $\max_{1 \leq i \leq n} |\tilde{\xi}_i| = o(n^{1/(2s)})$, from Assumption (A8), we have $\max_{1 \leq i \leq n} |b_{ki}| = o_p(n^{-1/2})$ for $k = 2, 3$. Theorem 3.1 implies that $\hat{\theta}_n^{(1)} - \theta = O_p(n^{-1/2})$, it can be checked that $\max_{1 \leq i \leq n} |b_{4i}| = o_p(n^{-1/2})$. Simple computation yields $S_i M - M_i = W_i^\top \alpha(U_i) O_p(\gamma_n)$, then we have $\max_{1 \leq i \leq n} |b_{5i}| = o_p(n^{-1/2})$. Note that $S_i \varepsilon = W_i^\top \mathbf{1}_q O_p(\sqrt{\frac{\log n}{n h_n}})$ and $S_i e = 0$, one can get $\max_{1 \leq i \leq n} |b_{6i}| = o_p(n^{-1/2})$. Hence, we have

$$\frac{(n-1)^2}{n} \sum_{i=1}^n (\hat{\theta}_n^{(1)} - \hat{\theta}_{n,-i}^{(1)})^2 \xrightarrow{\mathcal{P}} \Lambda_1.$$

(b) Following the definitions of $\hat{\theta}_n^{(2)}$ and $\hat{\theta}_{n,-i}^{(2)}$, simple computations yield

$$\begin{aligned} \hat{\theta}_n^{(2)} - \hat{\theta}_{n,-i}^{(2)} &= \frac{1}{n-1} [(W_i^\top \alpha(U_i) + \xi_i^\top \beta - \theta) + \delta_i (\varepsilon_i - e_i^\top \beta) + (1 - \delta_i) \tilde{\xi}_i^\top (\hat{\beta}_n - \beta)] \\ &\quad + \frac{1}{n-1} (1 - \delta_i) (S_i M - M_i - S_i \varepsilon) + \frac{1}{n-1} \delta_i \tilde{\xi}_j^\top (\hat{\beta}_{n,-i} - \hat{\beta}_n) \\ &\quad - \frac{1}{n-1} (\hat{\theta}_n^{(1)} - \theta) + \frac{1}{n-1} \sum_{j=1}^n (1 - \delta_j) \tilde{\xi}_j^\top (\hat{\beta}_{n,-i} - \hat{\beta}_n) \\ &\quad - \frac{1}{n(n-1)} \sum_{j=1}^n \delta_j (\tilde{Y}_j - \tilde{\xi}_j^\top \beta) + \frac{1}{n(n-1)} \sum_{j=1}^n \delta_j \tilde{\xi}_j^\top (\hat{\beta}_n - \beta) \\ &:= l_{1i} + l_{2i} + l_{3i} + l_{4i} + l_{5i} + l_{6i} + l_{7i}. \end{aligned}$$

According to the proof of (b) in Theorem 3.1, then one can get

$$\frac{(n-1)^2}{n} \sum_{i=1}^n l_{1i}^2 = \frac{1}{n} \sum_{i=1}^n \left\{ (W_i^\top \alpha(U_i) + \xi_i \beta - \theta) + \delta_i (\varepsilon_i - e_i^\top \beta) + (1 - \delta_i) \tilde{\xi}_i^\top (\hat{\beta}_n - \beta) \right\}^2 \\ \xrightarrow{\mathcal{P}} \Lambda_2.$$

By similar arguments to that of (a), we have $\max_{1 \leq i \leq n} |l_{ki}| = o_p(n^{-1/2})$ for $k = 2, \dots, 7$. Hence, we have

$$\frac{(n-1)^2}{n} \sum_{i=1}^n (\hat{\theta}_n^{(2)} - \hat{\theta}_{n,-i}^{(2)})^2 \xrightarrow{\mathcal{P}} \Lambda_2.$$

(c) Note that

$$\begin{aligned} \hat{\theta}_n^{(3)} - \hat{\theta}_{n,-i}^{(3)} &= \frac{1}{n-1} \left[(W_i^\top \alpha(U_i) + \xi_i^\top \beta - \theta) + \frac{\delta_i}{P(Z_i)} (\varepsilon_i - e_i^\top \beta) \right] \\ &\quad + \frac{1}{n-1} \frac{P(Z_i) - \hat{P}_n(Z_i)}{P(Z_i) \hat{P}_n(Z_i)} \delta_i (\varepsilon_i - e_i^\top \beta) \\ &\quad + \frac{1}{n-1} \left(1 - \frac{\delta_i}{\hat{P}_n(Z_i)} \right) (S_i M - M_i - S_i \varepsilon) + \frac{1}{n-1} \left(1 - \frac{\delta_i}{\hat{P}_n(Z_i)} \right) \tilde{\xi}_i^\top (\hat{\beta}_n - \beta) \\ &\quad + \frac{1}{n-1} \frac{\delta_i}{\hat{P}_n(Z_i)} \tilde{\xi}_j^\top (\hat{\beta}_{n,-i} - \hat{\beta}_n) + \frac{1}{n-1} \frac{\hat{P}_{n,-i}(Z_i) - \hat{P}_n(Z_i)}{\hat{P}_n(Z_i) \hat{P}_{n,-i}(Z_i)} \delta_i (\tilde{Y}_i - \tilde{\xi}_i^\top \hat{\beta}_{n,-i}) \\ &\quad - \frac{1}{n-1} (\hat{\theta}_n^{(1)} - \theta) + \frac{1}{n-1} \sum_{j=1}^n \left(1 - \frac{\delta_j}{\hat{P}_n(Z_j)} \right) \tilde{\xi}_j^\top (\hat{\beta}_{n,-i} - \hat{\beta}_n) \\ &:= m_{1i} + m_{2i} + m_{3i} + m_{4i} + m_{5i} + m_{6i} + m_{7i} + m_{8i}. \end{aligned}$$

Standard calculations yield

$$\frac{(n-1)^2}{n} \sum_{i=1}^n m_{1i}^2 = \frac{1}{n} \sum_{i=1}^n \left\{ (W_i^\top \alpha(U_i) + \xi_i \beta - \theta) + \frac{\delta_i}{P(Z_i)} (\varepsilon_i - e_i^\top \beta) \right\}^2 \xrightarrow{\mathcal{P}} \Lambda_3.$$

From Lemma A.2 and the fact $\hat{P}_{n,-i}(z) - \hat{P}_n(z) = O_p(n^{-1})$, by similar arguments to that of (a), we have $\max_{1 \leq i \leq n} |m_{ki}| = o_p(n^{-1/2})$ for $k = 2, \dots, 8$. Hence, we have

$$\frac{(n-1)^2}{n} \sum_{i=1}^n (\hat{\theta}_n^{(3)} - \hat{\theta}_{n,-i}^{(3)})^2 \xrightarrow{\mathcal{P}} \Lambda_3. \quad \square$$

Proof of Theorem 3.1 (a) We first prove Theorem 3.1 for $\hat{\theta}_n^{(1)}$. Recalling the definition of $\hat{\theta}_n^{(1)}$ in (2.4), then one can write

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n^{(1)} - \theta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi_i^\top \beta + W_i^\top \alpha(U_i) - \theta) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i^\top (\hat{\beta}_n - \beta) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n W_i^\top (\hat{\alpha}_n(U_i) - \alpha(U_i)) \\ &:= A_1 + A_2 + A_3. \end{aligned} \quad (\text{A.5})$$

Following Lemma A.4 in the Appendix, it can be checked that

$$A_3 = \frac{1}{\sqrt{n}} \sum_{i=1}^n (W_i^\top \ 0) \{D_u^\top \omega_u^\delta D_u\}^{-1} D_u^\top \omega_u^\delta X_i^\top (\beta - \hat{\beta}_n) + o_p(1).$$

Combining A_2 with A_3 , and from Lemma A.3, one can get

$$\begin{aligned} A_2 + A_3 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi_i^\top - (W_i^\top \ 0) \{D_u^\top \omega_u^\delta D_u\}^{-1} D_u^\top \omega_u^\delta X_i^\top) (\hat{\beta}_n - \beta) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi_i^\top - W_i^\top \Gamma^{-1}(U_i) \Phi(U_i)) (\hat{\beta}_n - \beta) + o_p(1) \\ &= A_n^{-1} \Sigma_1 \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i (\{\xi_i^\top - W_i^\top \Gamma^{-1}(U_i) \Phi(U_i)\} (\varepsilon_i - e_i^\top \beta) + \Sigma_e \beta) \\ &\quad + o_p(1). \end{aligned} \tag{A.6}$$

From (A.5) and (A.6), we have

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n^{(1)} - \theta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi_i^\top \beta + W_i^\top \alpha(U_i) - \theta) \\ &\quad + A_n^{-1} \Sigma_1 \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i (\{\xi_i^\top - W_i^\top \Gamma^{-1}(U_i) \Phi(U_i)\} (\varepsilon_i - e_i^\top \beta) + \Sigma_e \beta) \\ &\quad + o_p(1). \end{aligned}$$

By the central limit theorem, the proof of Theorem 3.1 for $\hat{\theta}_n^{(1)}$ is finished.

(b) We prove Theorem 3.1 for $\hat{\theta}_n^{(2)}$. In view of the definition of $\hat{\theta}_n^{(2)}$ in (2.4), by similar arguments to that of $\hat{\theta}_n^{(1)}$ in (a), then

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n^{(2)} - \theta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (\delta_i \{Y_i - \xi_i^\top \hat{\beta}_n - W_i^\top \hat{\alpha}_n(U_i)\} + \{\xi_i^\top \hat{\beta}_n + W_i^\top \hat{\alpha}_n(U_i) - \theta\}) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \varepsilon_i - \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i e_i^\top \beta + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi_i^\top \beta + W_i^\top \alpha(U_i) - \theta) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n (1 - \delta_i) \xi_i^\top (\hat{\beta}_n - \beta) + \frac{1}{\sqrt{n}} \sum_{i=1}^n (1 - \delta_i) W_i^\top (\hat{\alpha}_n(U_i) - \alpha(U_i)) \\ &:= B_1 + B_2 + B_3 + B_4 + B_5. \end{aligned} \tag{A.7}$$

For B_5 , applying the same proof as of Lemma A.4, it is easy to prove

$$B_5 = \frac{1}{\sqrt{n}} \sum_{i=1}^n (1 - \delta_i) (W_i^\top \ 0) \{D_u^\top \omega_u^\delta D_u\}^{-1} D_u^\top \omega_u^\delta X_i^\top (\beta - \hat{\beta}_n) + o_p(1),$$

by which together with B_4 , and following the verification of Lemma A.3, then we have

$$\begin{aligned} B_4 + B_5 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (1 - \delta_i) (\xi_i^\top - (W_i^\top \ 0) \{D_u^\top \omega_u^\delta D_u\}^{-1} D_u^\top \omega_u^\delta X_i^\top) (\hat{\beta}_n - \beta) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (1 - \delta_i) \{ \xi_i^\top - W_i^\top \Gamma(U_i)^{-1} \Phi(U_i) \} (\hat{\beta}_n - \beta) + o_p(1) \\ &= A_n^{-1} \Sigma_2 \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \{ \{ \xi_i^\top - W_i^\top \Gamma^{-1}(U_i) \Phi(U_i) \} (\varepsilon_i - e_i^\top \beta) + \Sigma_e \beta \} \\ &\quad + o_p(1). \end{aligned} \quad (\text{A.8})$$

Based on (A.7) and (A.8), it follows that

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n^{(2)} - \theta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \varepsilon_i - \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i e_i^\top \beta + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi_i^\top \beta + W_i^\top \alpha(U_i) - \theta) \\ &\quad + A_n^{-1} \Sigma_2 \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \{ \{ \xi_i^\top - W_i^\top \Gamma^{-1}(U_i) \Phi(U_i) \} (\varepsilon_i - e_i^\top \beta) + \Sigma_e \beta \} \\ &\quad + o_p(1). \end{aligned}$$

By the central limit theorem, the proof of Theorem 3.1 for $\hat{\theta}_n^{(2)}$ is completed.

(c) We prove Theorem 3.1 for $\hat{\theta}_n^{(3)}$. According to the definition of $\hat{\theta}_n^{(3)}$, from Lemma A.2, we have

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n^{(3)} - \theta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\frac{\delta_i}{\hat{P}_n(Z_i)} \{ Y_i - \xi_i^\top \hat{\beta}_n - W_i^\top \hat{\alpha}_n(U_i) \} + \{ \xi_i^\top \hat{\beta}_n + W_i^\top \hat{\alpha}_n(U_i) - \theta \} \right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{\hat{P}_n(Z_i)} \varepsilon_i + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(1 - \frac{\delta_i}{\hat{P}_n(Z_i)} \right) e_i^\top \beta \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi_i^\top \beta + W_i^\top \alpha(U_i) - \theta) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(1 - \frac{\delta_i}{\hat{P}_n(Z_i)} \right) \xi_i^\top (\hat{\beta}_n - \beta) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(1 - \frac{\delta_i}{\hat{P}_n(Z_i)} \right) W_i^\top (\hat{\alpha}_n(U_i) - \alpha(U_i)) \\ &:= D_1 + D_2 + D_3 + D_4 + D_5. \end{aligned} \quad (\text{A.9})$$

For D_1 , we replace $\hat{P}_n(Z_i)$ with its true value $P(Z_i)$, then

$$\begin{aligned} D_1 &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{P(Z_i)} \varepsilon_i + \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{P(Z_i) - \hat{P}_n(Z_i)}{P^2(Z_i)} \delta_i \varepsilon_i + \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{(P(Z_i) - \hat{P}_n(Z_i))^2}{P^2(Z_i) \hat{P}_n(Z_i)} \delta_i \varepsilon_i \\ &:= D_{11} + D_{12} + D_{13}. \end{aligned}$$

Recalling the definition of $\widehat{P}_n(Z_i)$ in Sect. 2 and Assumption (A8), we have

$$\begin{aligned} D_{12} &= \frac{1}{n\sqrt{n}b_n^{p+q+1}} \sum_{i=1}^n \sum_{j=1}^n \frac{[P(Z_j) - \delta_j]}{f(Z_i)P^2(Z_i)} \delta_i \varepsilon_i \Omega\left(\frac{Z_i - Z_j}{b_n}\right) \\ &\quad + \frac{1}{n\sqrt{n}b_n^{p+q+1}} \sum_{i=1}^n \sum_{j=1}^n \frac{[P(Z_i) - P(Z_j)]}{f(Z_i)P^2(Z_i)} \delta_i \varepsilon_i \Omega\left(\frac{Z_i - Z_j}{b_n}\right) + o_p(1) \\ &:= D_{121} + D_{122} + o_p(1). \end{aligned}$$

Under Assumptions (A1), (A5) and (A8), standard computation yields

$$\begin{aligned} E[D_{121}]^2 &= \frac{1}{n^3 b_n^{2(p+q+1)}} \sum_{i=1}^n \sum_{j=1}^n E\left\{ \frac{[P(Z_j) - \delta_j]}{f(Z_i)P^2(Z_i)} \delta_i \varepsilon_i \Omega\left(\frac{Z_i - Z_j}{b_n}\right) \right\}^2 \\ &\leq \frac{1}{n^3 b_n^{2(p+q+1)}} \sum_{i=1}^n \sum_{j=1}^n E\left\{ \Omega^2\left(\frac{Z_i - Z_j}{b_n}\right) \right\} \leq \frac{C}{nb_n^{p+q+1}}. \end{aligned}$$

From Assumption (A8), we have $D_{121} = o_p(1)$. Similarly, $D_{122} = o_p(1)$. Lemma A.2 implies $D_{13} = o_p(1)$. Hence

$$D_1 = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{P(Z_i)} \varepsilon_i + o_p(1). \quad (\text{A.10})$$

Analogous to the arguments of D_1 , it is easy to prove

$$D_2 = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(1 - \frac{\delta_i}{P(Z_i)}\right) e_i^\top \beta + o_p(1). \quad (\text{A.11})$$

From Lemmas A.3 and A.4, and the missing mechanism, one simple computation yields $D_4 = o_p(1)$ and $D_5 = o_p(1)$. Hence, collecting the results above, (A.9)–(A.11), one can get

$$\begin{aligned} \sqrt{n}(\widehat{\theta}_n^{(3)} - \theta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{P(Z_i)} \varepsilon_i + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(1 - \frac{\delta_i}{P(Z_i)}\right) e_i^\top \beta \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n (\xi_i^\top \beta + W_i^\top \alpha(U_i) - \theta) + o_p(1). \end{aligned}$$

By the central limit theorem, the proof of Theorem 3.1 for $\widehat{\theta}_n^{(3)}$ is finished. \square

Proof of Theorem 3.2 (a) We first prove Theorem 3.2 for $\widehat{\theta}_j^{(1)}$. In order to verify $\sqrt{n}(\widehat{\theta}_j^{(1)} - \theta) = \sqrt{n}(\widehat{\theta}_n^{(1)} - \theta) + o_p(1)$, it suffices prove that $\widehat{\theta}_j^{(1)} = \widehat{\theta}_n^{(1)} + o_p(n^{-1/2})$. Recalling the definition of $\widehat{\theta}_j^{(1)}$ given in Sect. 2, then one can re-write

$$\widehat{\theta}_j^{(1)} = \widehat{\theta}_n^{(1)} + \frac{n-1}{n} \sum_{i=1}^n (\widehat{\theta}_n^{(1)} - \widehat{\theta}_{n-i}^{(1)}).$$

Therefore, to obtain the desired results, we just need to prove

$$\sqrt{n} \sum_{i=1}^n (\hat{\theta}_n^{(1)} - \hat{\theta}_{n-i}^{(1)}) = o_p(1). \quad (\text{A.12})$$

Following the definitions of $\hat{\theta}_n^{(1)}$ and $\hat{\theta}_{n-i}^{(1)}$, then we have

$$\begin{aligned} \sqrt{n} \sum_{i=1}^n (\hat{\theta}_n^{(1)} - \hat{\theta}_{n-i}^{(1)}) &= \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \sum_{j=1}^n \tilde{\xi}_j^\top (\hat{\beta}_n - \hat{\beta}_{n-i}) - \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \tilde{\xi}_i^\top (\hat{\beta}_n - \hat{\beta}_{n-i}) \\ &:= Q_1 + Q_2. \end{aligned} \quad (\text{A.13})$$

From the proof of (6.16) in Liu and Liang [12], it can be checked that $\sqrt{n} \sum_{i=1}^n (\hat{\beta}_n - \hat{\beta}_{n-i}) = o_p(1)$. According to Lemma A.5, one can write

$$Q_1 = \frac{1}{n-1} \sum_{j=1}^n \tilde{\xi}_j \cdot \sqrt{n} \sum_{i=1}^n (\hat{\beta}_n - \hat{\beta}_{n-i}) = O_p(n^{-1/2}) O_p(1) o_p(1) = o_p(1). \quad (\text{A.14})$$

Following Lemma 6.11 in Liu and Liang [12], it follows that $\|\hat{\beta}_n - \hat{\beta}_{n-i}\| = O_p(n^{-1})$. We combine this with the fact $\max_{1 \leq i \leq n} \|\tilde{\xi}_i\| = o(n^{1/(2s)})$ for $s > 2$. Hence, under Assumption (A8), we have

$$|Q_2| \leq \sqrt{n} \max \|\tilde{\xi}_i\| \cdot \|\hat{\beta}_n - \hat{\beta}_{n-i}\| = o(n^{1/(2s)}) O_p(n^{-1/2}) = o_p(n^{\frac{1}{2s}-\frac{1}{2}}) = o_p(1). \quad (\text{A.15})$$

Hence, from (A.13)–(A.15), it is easy to prove $\sqrt{n} \sum_{i=1}^n (\hat{\theta}_n^{(1)} - \hat{\theta}_{n-i}^{(1)}) = o_p(1)$.

(b) The definition of $\hat{\theta}_n^{(2)}$ can be re-written as $\hat{\theta}_n^{(2)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i (\tilde{Y}_i - \tilde{\xi}_i^\top \hat{\beta}_n) + \hat{\theta}_n^{(1)}$. Hence, it is easy to prove

$$\begin{aligned} \sqrt{n} \sum_{i=1}^n (\hat{\theta}_n^{(2)} - \hat{\theta}_{n-i}^{(2)}) &= \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \sum_{j=1}^n \delta_j \tilde{\xi}_j^\top (\hat{\beta}_n - \hat{\beta}_{n-i}) + \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \delta_i \tilde{\xi}_i^\top (\hat{\beta}_n - \hat{\beta}_{n-i}) \\ &\quad + \sqrt{n} \sum_{i=1}^n (\hat{\theta}_n^{(1)} - \hat{\theta}_{n-i}^{(1)}) \\ &:= J_1 + J_2 + J_3. \end{aligned}$$

By a similar argument to that of (a), $J_1 = o_p(1)$ and $J_2 = o_p(1)$ can be proved easily. From the result (A.12), then we have $J_3 = o_p(1)$. Hence, $\sqrt{n} \sum_{i=1}^n (\hat{\theta}_n^{(2)} - \hat{\theta}_{n-i}^{(2)}) = o_p(1)$ can be proved.

(c) Following the definition of $\hat{\theta}_n^{(3)}$, we find $\hat{\theta}_n^{(3)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{\hat{P}_n(Z_i)} (\tilde{Y}_i - \tilde{\xi}_i^\top \hat{\beta}_n) + \hat{\theta}_n^{(1)}$. Hence, simple calculation yields

$$\begin{aligned} \sqrt{n} \sum_{i=1}^n (\hat{\theta}_n^{(3)} - \hat{\theta}_{n-i}^{(3)}) &= \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\hat{P}_n(Z_j) - \hat{P}_{n-i}(Z_j)}{\hat{P}_n(Z_j) \hat{P}_{n-i}(Z_j)} \delta_j (\tilde{Y}_j - \tilde{\xi}_j^\top \hat{\beta}_n) \\ &\quad - \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\delta_j}{\hat{P}_n(Z_j)} \tilde{\xi}_j^\top (\hat{\beta}_n - \hat{\beta}_{n-i}) \\ &\quad + \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\hat{P}_n(Z_j) - \hat{P}_{n-i}(Z_j)}{\hat{P}_n(Z_j) \hat{P}_{n-i}(Z_j)} \delta_j \tilde{\xi}_j^\top (\hat{\beta}_n - \hat{\beta}_{n-i}) \end{aligned}$$

$$\begin{aligned}
& + \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \frac{\delta_i}{\widehat{P}_n(Z_i)} (\widetilde{Y}_i - \widetilde{\xi}_i^\top \widehat{\beta}_n) \\
& + \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \frac{\widehat{P}_{n,-i}(Z_i) - \widehat{P}_n(Z_i)}{\widehat{P}_n(Z_i) \widehat{P}_{n,-i}(Z_i)} \delta_i (\widetilde{Y}_i - \widetilde{\xi}_i^\top \widehat{\beta}_n) \\
& + \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \frac{\widehat{P}_{n,-i}(Z_i) - \widehat{P}_n(Z_i)}{\widehat{P}_n(Z_i) \widehat{P}_{n,-i}(Z_i)} \delta_i \widetilde{\xi}_i^\top (\widehat{\beta}_n - \widehat{\beta}_{n,-i}) \\
& + \sqrt{n} \sum_{i=1}^n (\widehat{\theta}_n^{(1)} - \widehat{\theta}_{n,-i}^{(1)}) \\
& := L_1 + L_2 + L_3 + L_4 + L_5 + L_6 + L_7.
\end{aligned}$$

Note that $\widetilde{Y}_j - \widetilde{\xi}_j^\top \widehat{\beta}_n = \varepsilon_j - e_j^\top \beta + \widetilde{\xi}_j^\top (\beta - \widehat{\beta}_n) + W_j^\top (\alpha(U_j) - \widehat{\alpha}_n(U_j))$, then one can get

$$\begin{aligned}
L_1 &= \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\widehat{P}_n(Z_j) - \widehat{P}_{n,-i}(Z_j)}{\widehat{P}_n(Z_j) \widehat{P}_{n,-i}(Z_j)} \delta_j (\varepsilon_j - e_j^\top \beta) \\
&+ \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\widehat{P}_n(Z_j) - \widehat{P}_{n,-i}(Z_j)}{\widehat{P}_n(Z_j) \widehat{P}_{n,-i}(Z_j)} \delta_j \widetilde{\xi}_j^\top (\beta - \widehat{\beta}_n) \\
&+ \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\widehat{P}_n(Z_j) - \widehat{P}_{n,-i}(Z_j)}{\widehat{P}_n(Z_j) \widehat{P}_{n,-i}(Z_j)} \delta_j W_j^\top (\alpha(U_j) - \widehat{\alpha}_n(U_j)) \\
&:= L_{11} + L_{12} + L_{13}.
\end{aligned}$$

Let $a_{ni} = \frac{\Omega(\frac{z-Z_i}{b_n})}{\sum_{j=1}^n \Omega(\frac{z-Z_j}{b_n})}$, from Assumption (A7), it is easy to prove $a_{ni}(z) = O_p(n^{-1})$. Simple computation yields

$$\begin{aligned}
\widehat{P}_{n,-i}(z) &= \left(\widehat{P}_n(z) - \frac{\delta_i \Omega(\frac{z-Z_i}{b_n})}{\sum_{j=1}^n \Omega(\frac{z-Z_j}{b_n})} \right) \times \left(1 + \frac{\Omega(\frac{z-Z_j}{b_n})}{\sum_{j=1}^n \Omega(\frac{z-Z_j}{b_n})} \right) + O_p(n^{-2}) \\
&= (\widehat{P}_n(z) - \delta_i a_{ni}(z)) (1 + a_{ni}(z)) + O_p(n^{-2}).
\end{aligned}$$

Hence, applying the equation above, it follows that

$$\widehat{P}_{n,-i}(z) - \widehat{P}_n(z) = (\widehat{P}_n(z) - \delta_i) a_{ni}(z) - \delta_i a_{ni}^2(z) + O_p(n^{-2}), \quad (\text{A.16})$$

which indicates $\widehat{P}_{n,-i}(z) = \widehat{P}_n(z) + O_p(n^{-1})$. Together with Lemma A.2, one can compute

$$\begin{aligned}
L_{11} &= \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\widehat{P}_n(Z_j) - \widehat{P}_{n,-i}(Z_j)}{P^2(Z_j)} \delta_j (\varepsilon_j - e_j^\top \beta) + o_p(1) \\
&= \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\delta_j (\varepsilon_j - e_j^\top \beta)}{P^2(Z_j)} (\widehat{P}_n(Z_j) - \delta_i) a_{ni}(Z_j) + o_p(1) \\
&= \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\delta_j (\varepsilon_j - e_j^\top \beta)}{P^2(Z_j)} (P(Z_i) - \delta_i) \frac{\Omega(\frac{Z_j - Z_i}{b_n})}{nb_n^{p+q+1} f(Z_j)}
\end{aligned}$$

$$\begin{aligned}
& + \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\delta_j(\varepsilon_j - e_j^\top \beta)}{P^2(Z_j)} (P(Z_j) - P(Z_i)) \frac{\Omega\left(\frac{Z_j - Z_i}{b_n}\right)}{nb_n^{p+q+1}f(Z_j)} \\
& + \frac{\sqrt{n}}{n-1} \sum_{i=1}^n \sum_{j=1}^n \frac{\delta_j(\varepsilon_j - e_j^\top \beta)}{P^2(Z_j)} (\widehat{P}_n(Z_j) - P(Z_j)) \frac{\Omega\left(\frac{Z_j - Z_i}{b_n}\right)}{nb_n^{p+q+1}f(Z_j)} + o_p(1) \\
& = L_{111} + L_{112} + L_{113} + o_p(1).
\end{aligned}$$

Under Assumptions (A1), (A5) and (A7), $E(\varepsilon_j) = 0$ and $E(\varepsilon_j|Z_j) = 0$, then it is easy to verify

$$\begin{aligned}
EL_{111}^2 &= E \left\{ \frac{\sqrt{n}}{n(n-1)b_n^{p+q+1}} \sum_{i=1}^n \sum_{j=1}^n \frac{\delta_j(\varepsilon_j - e_j^\top \beta)}{P^2(Z_j)f(Z_j)} (P(Z_i) - \delta_i) \Omega\left(\frac{Z_j - Z_i}{b_n}\right) \right\}^2 \\
&= \frac{n}{n(n-1)^2 b_n^{2(p+q+1)}} \sum_{i=1}^n \sum_{j=1}^n E \left\{ \frac{\delta_j(\varepsilon_j - e_j^\top \beta)}{P^2(Z_j)f(Z_j)} (P(Z_i) - \delta_i) \Omega\left(\frac{Z_j - Z_i}{b_n}\right) \right\}^2 \\
&\leq \frac{C}{n^3 b_n^{2(p+q+1)}} \sum_{i=1}^n \sum_{j=1}^n E \left\{ \Omega^2\left(\frac{Z_j - Z_i}{b_n}\right) \right\} \leq C(n b_n^{p+q+1})^{-1}.
\end{aligned}$$

For any random variable X , we have $X = EX + O_p(\sqrt{\text{Var}(X)})$. Then from Assumption (A8), we have $L_{111} = o_p(1)$. Similarly, we have $L_{112} = o_p(1)$ and $L_{113} = o_p(1)$. Analogous to the proof of L_{11} , and from Lemmas A.3 and A.4, it is easy to prove $L_{12} = o_p(1)$ and $L_{13} = o_p(1)$. Hence, we have $L_1 = o_p(1)$. Similarly, $L_4 = o_p(1)$ and $L_5 = o_p(1)$. Note that $\|\widehat{\beta}_n - \widehat{\beta}_{n-i}\| = O_p(n^{-1})$ and $\max_{1 \leq i \leq n} \|\widetilde{\xi}_i\| = o(n^{1/2s})$, which indicate $L_i = o_p(1)$ for $i = 2, 3, 6$. From (a), we have $L_7 = o_p(1)$. Therefore, collecting the results above, one can get $\sqrt{n} \sum_{i=1}^n (\widehat{\theta}_n^{(3)} - \widehat{\theta}_{n-i}^{(3)}) = o_p(1)$. \square

Proof of Theorem 3.3 Let $\eta^{(k)}(\lambda) = \frac{1}{n} \sum_{i=1}^n \frac{\widehat{\theta}_{f_i}^{(k)} - \theta}{1 + \lambda(\widehat{\theta}_{f_i}^{(k)} - \theta)}$ for $k = 1, 2, 3$. It is easy to prove

$$0 = |\eta^{(k)}(\lambda)| = \left| \frac{1}{n} \sum_{i=1}^n (\widehat{\theta}_{f_i}^{(k)} - \theta) - \frac{\lambda}{n} \sum_{i=1}^n \frac{(\widehat{\theta}_{f_i}^{(k)} - \theta)^2}{1 + \lambda(\widehat{\theta}_{f_i}^{(k)} - \theta)} \right| \geq \frac{|\lambda| S_f^{(k)}}{1 + |\lambda| R_n^{(k)}} - \left| \frac{1}{n} \sum_{i=1}^n \widehat{\theta}_{f_i}^{(k)} - \theta \right|,$$

where $S_f^{(k)} = \frac{1}{n} \sum_{i=1}^n (\widehat{\theta}_{f_i}^{(k)} - \theta)^2$ and $R_n^{(k)} = \max_{1 \leq i \leq n} |\widehat{\theta}_{f_i}^{(k)} - \theta|$. Next, we just need to verify

$$S_f^{(k)} = \frac{1}{n} \sum_{i=1}^n (\widehat{\theta}_{f_i}^{(k)} - \theta)^2 \xrightarrow{\mathcal{P}} \Lambda_k, \quad (\text{A.17})$$

$$R_n^{(k)} = \max_{1 \leq i \leq n} |\widehat{\theta}_{f_i}^{(k)} - \theta| = o_p(\sqrt{n}). \quad (\text{A.18})$$

Theorem 3.2 implies that $S_f^{(k)} = \frac{1}{n} \sum_{i=1}^n (\widehat{\theta}_{f_i}^{(k)})^2 - \theta^2 + o_p(1)$. Since $\sqrt{n} \sum_{i=1}^n (\widehat{\theta}_n^{(k)} - \widehat{\theta}_{n-i}^{(k)}) = o_p(1)$ and $\widehat{\theta}_{f_i}^{(k)} = n\widehat{\theta}_n^{(k)} - (n-1)\widehat{\theta}_{n-i}^{(k)}$, we have

$$\frac{1}{n} \sum_{i=1}^n (\widehat{\theta}_{f_i}^{(k)})^2 = \widehat{\theta}_n^2 + \frac{(n-1)^2}{n} \sum_{i=1}^n (\widehat{\theta}_n^{(k)} - \widehat{\theta}_{n-i}^{(k)})^2 + o_p(1).$$

Lemma A.5 suggests that $S_J^{(k)} \xrightarrow{\mathcal{P}} \Lambda_k$. Similar to Owen [13], we derive that $\|\lambda\| = O_p(n^{-1/2})$. For convenience, let $\zeta_i^{(k)} = \lambda(\widehat{\theta}_{j_i}^{(k)} - \theta)$, then

$$\max_{1 \leq i \leq n} \|\zeta_i^{(k)}\| \leq \|\lambda\| \cdot \max_{1 \leq i \leq n} |\widehat{\theta}_{j_i}^{(k)} - \theta| = O_p(n^{-1}) o_p(\sqrt{n}) = o_p(n^{-1/2}). \quad (\text{A.19})$$

Note that

$$\begin{aligned} 0 &= \eta^{(k)}(\lambda) = \frac{1}{n} \sum_{i=1}^n (\widehat{\theta}_{j_i}^{(k)} - \theta) \cdot \frac{1}{1 + \zeta_i^{(k)}} \\ &= \frac{1}{n} \sum_{i=1}^n (\widehat{\theta}_{j_i}^{(k)} - \theta) - \lambda S_J^{(k)} + \frac{1}{n} \sum_{i=1}^n (\widehat{\theta}_{j_i}^{(k)} - \theta) \cdot \frac{(\zeta_i^{(k)})^2}{1 + \zeta_i^{(k)}}. \end{aligned}$$

Applying (A.17) and (A.18), it is easy to derive that $\frac{1}{n} \sum_{i=1}^n (\widehat{\theta}_{j_i}^{(k)} - \theta) \cdot \frac{(\zeta_i^{(k)})^2}{1 + \zeta_i^{(k)}} = o_p(n^{-1/2})$. Thus we have

$$\lambda = (S_J^{(k)})^{-1} \frac{1}{n} \sum_{i=1}^n (\widehat{\theta}_{j_i}^{(k)} - \theta) + o_p(n^{-1/2}).$$

Let $\rho_i^{(k)} = \sum_{l=3}^{\infty} \frac{(-1)^{l-1}}{l} (\zeta_i^{(k)})^l = O((\zeta_i^{(k)})^3)$, then from (A.19), one can get $|\sum_{i=1}^n \rho_i^{(k)}| \leq C \sum_{i=1}^n (\zeta_i^{(k)})^3 = o_p(n^{-1/2})$. By Taylor expansion, we have

$$\begin{aligned} l^{(k)}(\theta) &= 2 \sum_{i=1}^n \log(1 + \lambda(\widehat{\theta}_{j_i}^{(k)} - \theta)) = 2 \sum_{i=1}^n \zeta_i - \sum_{i=1}^n \zeta_i^2 + 2 \sum_{i=1}^n \rho_i^{(k)} \\ &= 2\lambda n(\widehat{\theta}_j^{(k)} - \theta) - n\lambda^2 S_J^{(k)} + o_p(1) = n(S_J^{(k)})^{-1} (\widehat{\theta}_j^{(k)} - \theta)^2 + o_p(1). \end{aligned}$$

Finally, combining (A.17) with Theorem 3.2, the proof of Theorem 3.3 is finished. \square

Acknowledgements

The authors thank the editor, the associate editor and two anonymous referees for their constructive comments and suggestions.

Funding

This work was supported by the China Postdoctoral Science Foundation (2019M651422), the National Natural Science Foundation of China (71701127) and the Key Project of National Social and Scientific Fund Program (18ZDA052).

Availability of data and materials

This paper did not use real data. Simulation data was produced by Matlab software.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

YZ gave the framework of the article and completed the theoretical proof of the article. CW completed the simulation analysis. All authors read and approved the final manuscript.

Author details

¹School of Economics and Management, Shanghai Maritime University, Shanghai, 201306, China. ²Key Laboratory of Advanced Theory and Application in Statistics and Data Science—MOE, School of Statistics, East China Normal University, Shanghai, 200062, China. ³School of Mathematics and Statistics, Huangshan University, Huangshan, 245041, China.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 1 June 2020 Accepted: 11 September 2020 Published online: 21 September 2020

References

1. Cheng, P.E.: Nonparametric estimation of mean functionals with data missing at random. *J. Am. Stat. Assoc.* **89**, 81–87 (1994)
2. Fan, G.L., Liang, H.Y., Shen, Y.: Penalized empirical likelihood for high-dimensional partially linear varying coefficient model with measurement errors. *J. Multivar. Anal.* **147**, 183–201 (2016)
3. Fan, G.L., Liang, H.Y., Zhu, L.X.: Penalized profile least squares-based statistical inference for varying coefficient partially linear errors-in-variables models. *Sci. China Math.* **61**, 1677–1694 (2018)
4. Fan, J., Huang, T.: Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* **11**(6), 1031–1057 (2005)
5. Gong, Y., Peng, L., Qi, Y.: Smoothed jackknife empirical likelihood method for ROC curve. *J. Multivar. Anal.* **101**, 1520–1531 (2010)
6. He, X.X., Feng, X.T., Zhao, X.: Semiparametric partially linear varying coefficient models with panel count data. *Lifetime Data Anal.* **23**(3), 439–466 (2017)
7. Huang, Z., Zhang, R.: Empirical likelihood for nonparametric parts in semiparametric varying-coefficient partially linear models. *Stat. Probab. Lett.* **79**, 1798–1808 (2009)
8. Jing, B.Y., Yuan, J., Zhou, W.: Jackknife empirical likelihood. *J. Am. Stat. Assoc.* **104**, 1224–1232 (2009)
9. Kai, B., Li, R., Zou, H.: New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Ann. Stat.* **39**, 305–332 (2011)
10. Liang, H.: Partially linear models with missing response variables and error-prone covariates. *Biometrika* **94**(1), 185–198 (2007)
11. Little, R.J.A., Rubin, D.B.: *Statistical Analysis with Missing Data*. Wiley, New York (1987)
12. Liu, A.A., Liang, H.Y.: Jackknife empirical likelihood of error variance in partially linear varying-coefficient errors-in-variables models. *Stat. Pap.* **58**(1), 1–28 (2017)
13. Owen, A.B.: Empirical likelihood ratio confidence intervals for a single function. *Biometrika* **75**, 237–249 (1988)
14. Owen, A.B.: Empirical likelihood ratio confidence regions. *Ann. Stat.* **18**(1), 90–120 (1990)
15. Peng, L., Qi, Y., Van Keilegom, I.: Jackknife empirical likelihood method for copulas. *Test* **21**, 74–92 (2012)
16. Rao, J.N.K.: On variance estimation with imputed survey data (with discussion). *J. Am. Stat. Assoc.* **91**, 499–520 (1996)
17. Shen, Y., Liang, H.Y.: Quantile regression for partially linear varying-coefficient model with censoring indicators missing at random. *Comput. Stat. Data Anal.* **117**, 1–18 (2017)
18. Tang, N.S., Zhao, P.Y.: Empirical likelihood-based inference in nonlinear regression models with missing responses at random. *Statistics* **47**(4–6), 1141–1159 (2013)
19. Wang, Q.: Statistical estimation in partial linear models with covariate data missing at random. *Ann. Inst. Stat. Math.* **61**(1), 47–84 (2009)
20. Wang, Q., Linton, O., Härdle, W.: Semiparametric regression analysis with missing response at random. *J. Am. Stat. Assoc.* **99**(466), 334–345 (2004)
21. Wang, Q.H., Rao, J.N.K.: Empirical likelihood for linear regression models under imputation for missing responses. *Can. J. Stat.* **29**, 597–608 (2001)
22. Wang, Q.H., Rao, J.N.K.: Empirical likelihood-based inference under imputation with missing response. *Ann. Stat.* **30**, 896–924 (2002)
23. Wang, Q.H., Rao, J.N.K.: Empirical likelihood-based inference in linear models with missing data. *Scand. J. Stat.* **29**, 563–576 (2002)
24. Xu, H.X., Fan, G.L., Wu, C.X., Chen, Z.L.: Statistical inference for varying-coefficient partially linear errors-in-variables models with missing data. *Commun. Stat., Theory Methods* **48**(22), 5621–5636 (2019)
25. Xue, L.: Empirical likelihood for linear models with missing responses. *J. Multivar. Anal.* **100**(7), 1353–1366 (2009)
26. Yang, H., Zhao, Y.: Smoothed jackknife empirical likelihood inference for ROC curves with missing data. *J. Multivar. Anal.* **140**, 123–138 (2015)
27. Yates, F.: The analysis of replicated experiments when the field results are incomplete. *Emp. J. Exp. Agric.* **1**, 129–142 (1933)
28. You, J.H., Chen, G.: Estimation of a semiparametric varying-coefficient partially linear errors-in-variables model. *J. Multivar. Anal.* **97**(2), 324–341 (2006)
29. You, J.H., Zhou, Y.: Empirical likelihood for semiparametric varying-coefficient partially linear regression models. *Stat. Probab. Lett.* **76**(4), 412–422 (2006)
30. Yu, X., Zhao, Y.: Jackknife empirical likelihood inference for the accelerated failure time model. *Test* **28**, 269–288 (2019)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)