Metrika (2012) 75:621–653
DOI 10.1007/s00184-011-0345-9

# Semiparametric estimation of logistic regression model with missing covariates and outcome

**Shen-Ming Lee · Chin-Shang Li · Shu-Hui Hsieh · Li-Hui Huang**

Received: 16 April 2010 / Published online: 5 February 2011


**Abstract** We consider a semiparametric method to estimate logistic regression models with missing both covariates and an outcome variable, and propose two new estimators. The first, which is based solely on the validation set, is an extension of the validation likelihood estimator of Breslow and Cain (Biometrika 75:11–20, 1988). The second is a joint conditional likelihood estimator based on the validation and non-validation data sets. Both estimators are semiparametric as they do not require any model assumptions regarding the missing data mechanism nor the specification of the conditional distribution of the missing covariates given the observed covariates. The asymptotic distribution theory is developed under the assumption that all covariate variables are categorical. The finite-sample properties of the proposed estimators are investigated through simulation studies showing that the joint conditional likelihood estimator is the most efficient. A cable TV survey data set from Taiwan is used to illustrate the practical use of the proposed methodology.

**Keywords** Missing value · Logistic regression model · Missing outcome · Missing covariates

S.-M. Lee · S.-H. Hsieh
Department of Statistics, Feng Chia University, Taichung, Taiwan

C.-S. Li (✉)
Department of Public Health Sciences, Division of Biostatistics, University of California, Davis, CA, USA
e-mail: cssli@ucdavis.edu

L.-H. Huang
Department of International Trade, Overseas Chinese University, Taichung, Taiwan

## 1 Introduction

Missing values inevitably occur in medical and other scientific studies due to several reasons (e.g., study design or conditions). A common approach to dealing with missing values is the complete-case analysis that simply ignores the observations with missing values. Such method may cause a loss of efficiency and a biased estimation when the occurrence of the missing values is seldom completely at random (MCAR). The same observation is made for other types of missingness such as missing at random (MAR) or not missing at random (NMAR). See Rubin (1976) for further insights into the types of missing data mechanism. It is then important to develop techniques that can account for the missingness in these situations.

Analysis of medical studies typically includes estimation of the association between an outcome variable and a set of covariates or risk factors. Sometimes, it is possible to obtain a crude outcome measure on a large sample, but the true outcome can be ascertained only for a subsample. For example, Pepe et al. (1994) described a setting where patients who received allogeneic bone marrow transplant for aplastic anemia may develop graft versus host disease (GVHD). Measuring chronic GVHD requires longitudinal follow-up, but acute GVHD can be readily ascertained when patients are still being treated at the hospital. Because of high cost, it is sometimes difficult to make a diagnosis of the chronic disease outcome. Variables like acute GVHD are often called surrogate outcomes. Several methods have been proposed to handle the problem resulting in the substitution of the surrogate outcome for the true outcome to make scientific inference (e.g., Pepe 1992; Pepe et al. 1994; Chu and Halloran 2004; Chen and Breslow 2004). In a different situation, Albert et al. (1997) and Bollinger and David (1997) gave examples showing that the binary true outcome of interest may also be misclassified.

Logistic regression is the most popular form of binary regression; see, e.g., Cox (1970) and Pregibon (1981). The model assumes that the logarithm of the odds of the outcome is a linear function of covariates. The topic of a logistic regression with missing covariates has been recently studied by Reilly and Pepe (1995), and Wang et al. (2002), among others. Both Pepe (1992) and Cheng and Hsueh (1999) discussed bias corrections in the estimation of parameters of a logistic regression model when the binary outcome is subject to missing values and misclassification. Cheng and Hsueh (2003) proposed estimation methods for a logistic regression model when the binary outcome and covariate values are both subject to measurement errors. Note, they assumed the complete data set consists of a primary sample plus a smaller validation subsample, which is obtained by double sampling scheme. We analyze missing data for both covariates and an outcome variable. However, when the surrogate covariates and outcome are always observed, it may be more accurate to characterize it as a measurement error problem. Many researchers have proposed methods to adjust for error in measurement. Two related methods have been referred in the literature to as regression calibration. Rosner et al. (1989) and Ronser et al. (1990) proposed to correct the regression coefficient estimates when covariates are measured with error, utilizing a separate validation study where the relation between true covariates value and observed surrogate can be estimated using linear regression; and Carroll and Stefanski (1990) proposed a general model for quasi-likelihood estimation methods when covariates are measured with error.

Assuming that all covariate variables are categorical, we use two semiparametric methods to estimate logistic regression models that are missing both covariates and an outcome variable, and where the missing data possibly depends on the observed data. The first method is an extension of validation likelihood approach of Breslow and Cain (1988). The second is a joint conditional likelihood based-method that uses the validation and non-validation data sets. For both methods, it is unnecessary to make any model assumptions for the probability of missingness and specification of the conditional distribution of the missing covariates given the observed covariates. To the best of our knowledge, no one has proposed semiparametric methods for estimation of a logistic regression model that is missing both covariates and an outcome variable and in which the type of missingness is MAR. We organize the paper as follows. In Sect. 2, we describe the proposed estimators. Section 3 studies the asymptotic properties and relative efficiencies of these estimators. In Sect. 4, we review some existing estimates. In Sect. 5, we conduct a simulation study to investigate their finite-sample performance. In Sect. 6, the proposed methodology and other existing methodology are applied to a cable TV survey data set from Taiwan. Finally, some concluding remarks are provided in Sect. 7.

## 2 The proposed estimators

Let $Y$ be a binary outcome and $X$ a vector of covariates, which may be missing on some subjects. Assume that $Y^0$ and $W$ are surrogate variables for $Y$ and $X$, respectively. Suppose that the binary surrogate outcome $Y^0$, the surrogate variable $W$, and a covariate vector $Z$ are always observed. Then, we consider the following logistic regression model:

$$P\left(Y=1|X,Z\right)=H\left(\beta_0+\beta_1^{\mathrm{T}}X+\beta_2^{\mathrm{T}}Z\right),$$

where $H(u)=[1+\exp(-u)]^{-1}$. The term of auxiliary data refers to data not in the regression model, but thought to be informative about the $Y$ and $X$, which are of interest.

For subject $i, i=1,2,\ldots,n$, let $\delta_i$ indicate whether $Y_i$ and $X_i$ are observed ($\delta_i=1$) or not ($\delta_i=0$). The validation data set ($\delta_i=1$) consists of $\left(Y_i, Y_i^0, X_i, Z_i, W_i\right)$, and the non-validation data set ($\delta_i=0$) consists of $\left(Y_i^0, Z_i, W_i\right)$. To deal with the missingness, we assume the selection probability of observing $Y_i$ and $X_i$ as follows:

$$P\left(\delta_i=1|Y_i,Y_i^0,X_i,Z_i,W_i\right)=\pi\left(Y_i^0,Z_i,W_i\right)=\pi\left(Y_i^0,V_i\right), \qquad (1)$$

where $V_i=\left(Z_i^T, W_i^T\right)^T$. Under assumption (1), $Y_i$ and $X_i$ are MAR as described in Rubin (1976). Here $\pi\left(Y_i^0, V_i\right)$ is an unknown nuisance parameter although it may be prespecified at design stage in some other applications.

Under the assumption that all covariate variables are categorical, let $v_1, v_2, \ldots, v_g$ denote the distinct values of the $V_i$'s, the nonparametric estimator of $\pi\left(y^0, v\right)$ is then given by

$$\widehat{\pi}\left(y^0, v\right) = \frac{\sum_{i=1}^{n} \delta_i I\left(Y_i^0 = y^0, V_i = v\right)}{\sum_{i=1}^{n} I\left(Y_i^0 = y^0, V_i = v\right)}, \tag{2}$$

where $v \in \left(v_1, v_2, \ldots, v_g\right)$ and $y^0 = 0, 1$.

## 2.1 Validation likelihood estimator

When $Y$ is binary and observable, Breslow and Cain (1988) proposed a semiparametric estimator of $\boldsymbol{\beta} = \left(\beta_0, \beta_1^{\mathrm{T}}, \beta_2^{\mathrm{T}}\right)^{\mathrm{T}}$ based on conditional likelihood of $Y$ given $X$, $Z$, and $\delta = 1$. When $X$ and $Y$ are MAR, it can be shown that

$$\begin{aligned}
&P\left(Y_i = 1 | X_i, Z_i, \delta_i = 1\right)\\
&= \frac{P\left(Y_i = 1, \delta_i = 1 | X_i, Z_i\right)}{P\left(\delta_i = 1 | X_i, Z_i\right)}\\
&= \frac{P\left(\delta_i = 1 | Y_i = 1, X_i, Z_i\right) P\left(Y_i = 1 | X_i, Z_i\right)}{P\left(\delta_i = 1 | Y_i = 1, X_i, Z_i\right) P\left(Y_i = 1 | X_i, Z_i\right) + P\left(\delta_i = 1 | Y_i = 0, X_i, Z_i\right) P\left(Y_i = 0 | X_i, Z_i\right)}\\
&= \frac{1}{1 + \frac{P(\delta_i = 1 | Y_i = 0, X_i, V_i) P(Y_i = 0 | X_i, V_i)}{P(\delta_i = 1 | Y_i = 1, X_i, V_i) P(Y_i = 1 | X_i, V_i)}}\\
&= \frac{1}{1 + e^{-\left(\beta_0 + \beta_1^{\mathrm{T}} X_i + \beta_2^{\mathrm{T}} Z_i\right) - \ln \frac{P(\delta_i = 1 | Y_i = 1, X_i, V_i)}{P(\delta_i = 1 | Y_i = 0, X_i, V_i)}}}\\
&= H\left[\beta_0 + \beta_1^{\mathrm{T}} X_i + \beta_2^{\mathrm{T}} Z_i + \ln \frac{P\left(\delta_i = 1 | Y_i = 1, X_i, V_i\right)}{P\left(\delta_i = 1 | Y_i = 0, X_i, V_i\right)}\right]\\
&= H\left[\beta_0 + \beta_1^{\mathrm{T}} X_i + \beta_2^{\mathrm{T}} Z_i + \ln \frac{\pi(1, V_i)[1 - \theta(X_i, V_i)] + \pi(0, V_i)\theta(X_i, V_i)}{\pi(1, V_i)\phi(X_i, V_i) + \pi(0, V_i)[1 - \phi(X_i, V_i)]}\right]\\
&\equiv H_+(X_i, V_i; \boldsymbol{\beta}),
\end{aligned} \tag{3}$$

where

$$\begin{aligned}
P\left(\delta_i = 1 | Y_i = 1, X_i, V_i\right) &= \sum_{y^0=0}^{1} P\left(\delta_i = 1 | Y_i^0 = y^0, Y_i = 1, X_i, V_i\right)\\
&\quad \times \; P\left(Y_i^0 = y^0 | Y_i = 1, X_i, V_i\right)\\
&= \pi(1, V_i)\left[1 - \theta\left(X_i, V_i\right)\right] + \pi(0, V_i)\theta(X_i, V_i)
\end{aligned}$$

for $\theta(x, v) = P\left(Y_i^0 = 0 | Y_i = 1, X_i = x, V_i = v\right)$, and

$$P\left(\delta_i = 1 | Y_i = 0, X_i, V_i\right) = \pi(1, V_i)\phi\left(X_i, V_i\right) + \pi(0, V_i)\left[1 - \phi(X_i, V_i)\right]$$

for $\phi(x, v) = P\left(Y_i^0 = 1 | Y_i = 0, X_i = x, V_i = v\right)$. It is noted that $\theta(x, v)$ and $\phi(x, v)$ are defined by misclassification probability functions. Based on the distribution of

$Y$ given $(X, Z, \delta = 1)$, we consider the validation likelihood $L(Y|X, Z, \delta = 1)$ by taking the production of the probability of $P(Y = 1|X, Z, \delta = 1)$, where $L(\cdot)$ denotes the likelihood function of a specified random variable. To estimate $\boldsymbol{\beta}$, we employ the following estimating function:

$$U_{1n}(\boldsymbol{\beta}) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \mathcal{X}_i \left[ Y_i - H_+(X_i, V_i; \boldsymbol{\beta}) \right] \tag{4}$$

for $\mathcal{X}_i = \left(1, X_i^{\mathrm{T}}, Z_i^{\mathrm{T}}\right)^{\mathrm{T}}$. The validation likelihood estimator of $\boldsymbol{\beta}$ is the solution to $U_{1n}(\boldsymbol{\beta}) = 0$. It can be shown with some algebra that $U_{1n}(\boldsymbol{\beta})$ is an unbiased estimating function for $\boldsymbol{\beta}$ when $\pi(Y_i^0, V_i)$, $\theta(X_i, V_i)$, and $\phi(X_i, V_i)$ are known. However, when these quantities are unknown, they need to be estimated first before solving the equation $U_{1n}(\boldsymbol{\beta}) = 0$ for the parameter vector $\boldsymbol{\beta}$. Suppose that $X$ has $m$ categories. For $X \in (x_1, x_2, \ldots, x_m)$, $V \in \left(v_1, v_2, \ldots, v_g\right)$, and $Y^0 \in (0, 1)$, let $A\left(x, y^0, v\right) = P\left(Y^0 = y^0|\delta = 1, Y = 0, X = x, V = v\right)$ and $B\left(x, y^0, v\right) = P\left(Y^0 = y^0|\delta = 1, Y = 1, X = x, V = v\right)$. One can then express $\theta(x, v)$ as follows:

$$\begin{aligned}
\theta(x, v) &= P\left(Y_i^0 = 0|Y_i = 1, X_i = x, V_i = v\right) \\
&= \frac{P\left(Y_i^0 = 0, Y_i = 1, X_i = x, V_i = v\right)}{P\left(Y_i = 1, X_i = x, V_i = v\right)} \\
&= \frac{\pi(1, v)\pi(0, v)P\left(Y_i^0 = 0, Y_i = 1, X_i = x, V_i = v\right)}{\pi(1, v)\pi(0, v)P\left(Y_i = 1, X_i = x, V_i = v\right)} \\
&= \frac{\pi(1, v)B(x, 0, v)}{\pi(1, v)B(x, 0, v) + \pi(0, v)B(x, 1, v)},
\end{aligned} \tag{5}$$

where the following fact has been used

$$\begin{aligned}
&\pi(1, v)\pi(0, v)P\left(Y_i = 1, X_i = x, V_i = v\right) \\
&\quad = \pi(1, v)\pi(0, v)\Big[P\left(Y_i^0 = 0, Y_i = 1, X_i = x, V_i = v\right) \\
&\qquad + P\left(Y_i^0 = 1, Y_i = 1, X_i = x, V_i = v\right)\Big] \\
&\quad = \pi(1, v)\pi(0, v)P\left(Y_i^0 = 0, Y_i = 1, X_i = x, V_i = v\right) \\
&\qquad + \pi(1, v)\pi(0, v)P\left(Y_i^0 = 1, Y_i = 1, X_i = x, V_i = v\right) \\
&\quad = \Big[\pi(1, v)P\left(Y_i^0 = 0|\delta_i = 1, Y_i = 1, X_i = x, V_i = v\right) \\
&\qquad + \pi(0, v)P\left(Y_i^0 = 1|\delta_i = 1, Y_i = 1, X_i = x, V_i = v\right)\Big] \\
&\qquad \times P\left(\delta_i = 1, Y_i = 1, X_i = x, V_i = v\right) \\
&\quad = \left[\pi(1, v)B(x, 0, v) + \pi(0, v)B(x, 1, v)\right] P\left(\delta_i = 1, Y_i = 1, X_i = x, V_i = v\right).
\end{aligned}$$

Similarly, $\phi(x, v)$ can be expressed as

$$\phi(x, v) = P\left(Y_i^0 = 1 | Y_i = 0, X_i = x, V_i = v\right) = \frac{\pi(0, v)A(x, 1, v)}{\pi(0, v)A(x, 1, v) + \pi(1, v)A(x, 0, v)}. \tag{6}$$

Because the probabilities $A\left(x, y^0, v\right)$ and $B\left(x, y^0, v\right)$ can be estimated by

$$\widehat{A}\left(x, y^0, v\right) = \frac{\sum_{i=1}^n \delta_i I\left(Y_i^0 = y^0, Y_i = 0, X_i = x, V_i = v\right)}{\sum_{i=1}^n \delta_i I\left(Y_i = 0, X_i = x, V_i = v\right)}, \tag{7}$$

and

$$\widehat{B}\left(x, y^0, v\right) = \frac{\sum_{i=1}^n \delta_i I\left(Y_i^0 = y^0, Y_i = 1, X_i = x, V_i = v\right)}{\sum_{i=1}^n \delta_i I\left(Y_i = 1, X_i = x, V_i = v\right)}, \tag{8}$$

we can estimate $\phi(x, v)$ and $\theta(x, v)$ by plugging (2), (7), and (8) into (6) and (5), respectively. Therefore, the resulting estimated score function is given by

$$\widehat{U}_{1n}(\boldsymbol{\beta}) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \mathcal{X}_i [Y_i - \widehat{H}_+(X_i, V_i; \boldsymbol{\beta})], \tag{9}$$

where

$$\widehat{H}_+(X_i, V_i; \boldsymbol{\beta}) \equiv H\left[\beta_0 + \beta_1^{\mathrm{T}} X_i + \beta_2^{\mathrm{T}} Z_i + \ln \frac{\widehat{\pi}(1, V_i)[1 - \widehat{\theta}(X_i, V_i)] + \widehat{\pi}(0, V_i)\widehat{\theta}(X_i, V_i)}{\widehat{\pi}(1, V_i)\widehat{\phi}(X_i, V_i) + \widehat{\pi}(0, V_i)[1 - \widehat{\phi}(X_i, V_i)]}\right]. \tag{10}$$

The validation estimate of $\boldsymbol{\beta}$, denoted by $\widehat{\boldsymbol{\beta}}_V$, is defined as the solution to the score equation $\widehat{U}_{1n}(\boldsymbol{\beta}) = 0$ and obtained by the Newton-Raphson algorithm.

### 2.2 Joint conditional likelihood estimator

In this subsection, we develop a joint conditional estimator based on the validation and non-validation data sets. First, proceeding as Wang et al. (2002) did, we can have

$$P(Y_i = 1 | V_i) = H\left[\beta_0 + \beta_2^{\mathrm{T}} Z_i + R(V_i; \beta_1)\right], \tag{11}$$

where $R(V_i;\beta_1) = \ln\left[\mathrm{E}\left(e^{\beta_1^{\mathrm{T}}X}|Y_i = 0, V_i\right)\right]$. We can furthermore show from (11) that

$$
\begin{aligned}
&P\left(Y_i^0 = 1|V_i\right)\\
&\quad= P\left(Y_i^0 = 1, Y_i = 1|V_i\right) + P\left(Y_i^0 = 1, Y_i = 0|V_i\right)\\
&\quad= P\left(Y_i^0 = 1|Y_i = 1, V_i\right) P\left(Y_i = 1|V_i\right) + P\left(Y_i^0 = 1|Y_i = 0, V_i\right) P\left(Y_i = 0|V_i\right)\\
&\quad= [1-\theta_0(V_i)]\frac{1}{1+e^{-[\beta_0+\beta_2^{\mathrm{T}}Z_i+R(V_i;\beta_1)]}} + \phi_0(V_i)\frac{e^{-[\beta_0+\beta_2^{\mathrm{T}}Z_i+R(V_i;\beta_1)]}}{1+e^{-[\beta_0+\beta_2^{\mathrm{T}}Z_i+R(V_i;\beta_1)]}}, \qquad (12)
\end{aligned}
$$

where $\theta_0(V_i) = P\left(Y_i^0 = 0|Y_i = 1, V_i\right)$ and $\phi_0(V_i) = P\left(Y_i^0 = 1|Y_i = 0, V_i\right)$. Using (12) and calculations similar to those in (3), one can have

$$
\begin{aligned}
&P\left(Y_i^0 = 1|V_i, \delta_i = 0\right)\\
&\quad= \frac{P\left(Y_i^0 = 1, \delta_i = 0|V_i\right)}{P\left(\delta_i = 0|V_i\right)}\\
&\quad= \frac{P\left(\delta_i = 0|Y_i^0 = 1, V_i\right) P\left(Y_i^0 = 1|V_i\right)}{P\left(\delta_i = 0|Y_i^0 = 1, V_i\right) P\left(Y_i^0 = 1|V_i\right) + P\left(\delta_i = 0|Y_i^0 = 0, V_i\right) P\left(Y_i^0 = 0|V_i\right)}\\
&\quad= H\left[\ln\frac{1-\pi(1, V_i)}{1-\pi(0, V_i)} + \ln\frac{P\left(Y_i^0 = 1|V_i\right)}{P\left(Y_i^0 = 0|V_i\right)}\right]\\
&\quad= H\left[\ln\frac{1-\pi(1, V_i)}{1-\pi(0, V_i)} + \ln\frac{\left[1-\theta_0(V_i)\right] + \phi_0(V_i)e^{-[\beta_0+\beta_2^{\mathrm{T}}Z_i+R(V_i;\beta_1)]}}{\theta_0(V_i) + \left[1-\phi_0(V_i)\right]e^{-[\beta_0+\beta_2^{\mathrm{T}}Z_i+R(V_i;\beta_1)]}}\right]\\
&\quad\equiv H_-\left(V_i;\boldsymbol{\beta}\right). \qquad (13)
\end{aligned}
$$

It is clear that $U_{1n}(\boldsymbol{\beta})$ in (4) is the estimating score obtained from the likelihood of $Y$ given $(X, Z, \delta = 1)$. To have higher efficient estimators, we consider the joint conditional likelihood $L^{\delta}(Y|X, Z, \delta = 1)\, L^{1-\delta}\left(Y^0|V, \delta = 0\right)$, which is developed by conditioning on the missingness status $\delta$. Therefore, based on $H_+(X_i, V_i;\boldsymbol{\beta})$ and $H_-(V_i;\boldsymbol{\beta})$, we define the score function as follows:

$$
U_n(\boldsymbol{\beta}) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\delta_i\mathcal{X}_i\left[Y_i - H_+(X_i, V_i;\boldsymbol{\beta})\right] + (1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})\left[Y_i^0 - H_-(V_i;\boldsymbol{\beta})\right]\right\},
$$

where $\mathcal{T}_i(\boldsymbol{\beta}) = [H_1(V_i;\boldsymbol{\beta}) - H_2(V_i;\boldsymbol{\beta})]\left(1, R_{\beta_1}^{\mathrm{T}}(V_i), Z_i^{\mathrm{T}}\right)^{\mathrm{T}}$ for $H_1(V_i;\boldsymbol{\beta}) = H\left[\beta_0 + \beta_2^{\mathrm{T}}Z_i + R(V_i;\beta_1) + \ln\frac{1-\theta_0(V_i)}{\phi_0(V_i)}\right]$, $H_2(V_i;\boldsymbol{\beta}) = H\left[\beta_0 + \beta_2^{\mathrm{T}}Z_i + R(V_i;\beta_1) + \ln\frac{\theta_0(V_i)}{1-\phi_0(V_i)}\right]$, and $R_{\beta_1}(V_i) = \frac{\partial R(V_i;\beta_1)}{\partial\beta_1}$. Because $\pi\left(Y_i^0, V_i\right)$, $R(V_i;\beta_1)$, $\theta_0(V_i)$, and $\phi_0(V_i)$ are usually unknown, they must be estimated before solving $U_n(\boldsymbol{\beta}) = 0$.

To estimate these quantities $\pi\left(Y_i^0, V_i\right)$, $R\left(V_i;\beta_1\right)$, $\theta_0(V_i)$, and $\phi_0(V_i)$, we first let $A_0\left(y^0, v\right) = P\left(\delta = 1, Y = 0, Y^0 = y^0 | V = v\right)$, $B_0\left(y^0, v\right) = P\big(\delta = 1, Y = 1, Y^0 = y^0 | V = v\big)$, $r_0\left(y^0, v; \beta_1\right) = \mathrm{E}\left(e^{\beta_1^{\mathrm{T}} X} | Y = 0, Y^0 = y^0, V = v\right)$, and $r(v, \beta_1) = e^{R(v;\beta_1)} = \mathrm{E}\left(e^{\beta_1^{\mathrm{T}} X} | Y = 0, V = v\right)$. By using the following fact

$$
\begin{aligned}
&\pi(1, v)\pi(0, v)P\left(Y_i = 1, V_i = v\right) \\
&\quad = \pi(1, v)\pi(0, v)\left[P\left(Y_i^0 = 0, Y_i = 1, V_i = v\right) + P\left(Y_i^0 = 1, Y_i = 1, V_i = v\right)\right] \\
&\quad = \pi(1, v)\pi(0, v)P\left(Y_i^0 = 0, Y_i = 1, V_i = v\right) \\
&\qquad + \pi(1, v)\pi(0, v)P\left(Y_i^0 = 1, Y_i = 1, V_i = v\right) \\
&\quad = \Big[\pi(1, v)P\left(\delta_i = 1, Y_i^0 = 0, Y_i = 1 | V_i = v\right) \\
&\qquad + \pi(0, v)P(\delta_i = 1, Y_i^0 = 1, Y_i = 1 | V_i = v)\Big] P\left(V_i = v\right) \\
&\quad = \left[\pi(1, v)B_0(0, v) + \pi(0, v)B_0(1, v)\right] P(V_i = v),
\end{aligned}
$$

we can have

$$
\begin{aligned}
\theta_0(v) &= P\left(Y_i^0 = 0 | Y_i = 1, V_i = v\right) \\
&= \frac{P\left(Y_i^0 = 0, Y_i = 1, V_i = v\right)}{P\left(Y_i = 1, V_i = v\right)} \\
&= \frac{\pi(1, v)\pi(0, v)P\left(Y_i^0 = 0, Y_i = 1, V_i = v\right)}{\pi(1, v)\pi(0, v)P\left(Y_i = 1, V_i = v\right)} \\
&= \frac{\pi(1, v)B_0(0, v)}{\pi(1, v)B_0(0, v) + \pi(0, v)B_0(1, v)}.
\end{aligned} \tag{14}
$$

Similarly, we can obtain

$$
\phi_0(v) = P\left(Y_i^0 = 1 | Y_i = 0, V_i = v\right) = \frac{\pi(0, v)A_0(1, v)}{\pi(0, v)A_0(1, v) + \pi(1, v)A_0(0, v)}. \tag{15}
$$

In addition, $r(v; \beta_1)$ can be expressed as follows:

$$
\begin{aligned}
r(v; \beta_1) &= \mathrm{E}\left(e^{\beta_1^{\mathrm{T}} X_i} | Y_i = 0, V_i = v\right) \\
&= \mathrm{E}\left[\mathrm{E}\left(e^{\beta_1^{\mathrm{T}} X_i} \middle| Y_i = 0, V_i = v, Y_i^0\right) \middle| Y_i = 0, V_i = v\right] \\
&= \mathrm{E}\left(e^{\beta_1^{\mathrm{T}} X_i} | Y_i = 0, V_i = v, Y_i^0 = 0\right) P\left(Y_i^0 = 0 | Y_i = 0, V_i = v\right) \\
&\quad + \mathrm{E}\left(e^{\beta_1^{\mathrm{T}} X_i} | Y_i = 0, V_i = v, Y_i^0 = 1\right) P\left(Y_i^0 = 1 | Y_i = 0, V_i = v\right) \\
&= r_0(0, v; \beta_1)\left[1 - \phi_0(v)\right] + r_0\left(1, v; \beta_1\right)\phi_0(v).
\end{aligned} \tag{16}
$$

We estimate the probabilities $A_0\left(y^0, v\right)$, $B_0\left(y^0, v\right)$, and $r_0\left(y^0, v; \beta_1\right)$, respectively, by

$$\widehat{A}_0\left(y^0, v\right) = \frac{\sum_{i=1}^{n} \delta_i I\left(Y_i^0 = y^0, Y_i = 0, V_i = v\right)}{\sum_{i=1}^{n} I\left(V_i = v\right)}, \tag{17}$$

$$\widehat{B}_0\left(y^0, v\right) = \frac{\sum_{i=1}^{n} \delta_i I\left(Y_i^0 = y^0, Y_i = 1, V_i = v\right)}{\sum_{i=1}^{n} I\left(V_i = v\right)}, \tag{18}$$

and

$$\widehat{r}_0\left(y^0, v; \beta_1\right) = \frac{\sum_{i=1}^{n} \delta_i e^{\beta_1^{\mathrm{T}} X_i} I\left(Y_i = 0, Y_i^0 = y^0, V_i = v\right)}{\sum_{i=1}^{n} \delta_i I\left(Y_i = 0, Y_i^0 = y^0, V_i = v\right)}. \tag{19}$$

The $R\left(V_i; \beta_1\right)$, $\theta_0(V_i)$, and $\phi_0(V_i)$ can then be easily estimated by simply plugging (2), (17), (18) and (19) into (14), (15) and (16), respectively. Let $\widehat{R}(v; \beta_1) = \ln[\widehat{r}(v; \beta_1)]$ denote the estimate of $R(v; \beta_1)$. Then, the resulting estimated score function for $\boldsymbol{\beta}$ is defined as follows:

$$\widehat{U}_n(\boldsymbol{\beta}) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{\delta_i \mathcal{X}_i [Y_i - \widehat{H}_+(X_i, V_i; \boldsymbol{\beta})] + (1-\delta_i)\,\widehat{\mathcal{T}}_i(\boldsymbol{\beta})\,[Y_i^0 - \widehat{H}_-(V_i; \boldsymbol{\beta})]\right\}.$$

Here $\widehat{H}_+(X_i, V_i; \boldsymbol{\beta})$ is given in (10).

$$\widehat{H}_-(V_i; \boldsymbol{\beta}) = H\left[\ln \frac{1-\widehat{\pi}(1, V_i)}{1-\widehat{\pi}(0, V_i)} + \ln \frac{[1-\widehat{\theta}_0(V_i)] + \widehat{\phi}_0(V_i) e^{-[\beta_0 + \beta_2^{\mathrm{T}} Z_i + \widehat{R}(V_i; \beta_1)]}}{\widehat{\theta}_0(V_i) + [1-\widehat{\phi}_0(V_i)] e^{-[\beta_0 + \beta_2^{\mathrm{T}} Z_i + \widehat{R}(V_i; \beta_1)]}}\right].$$

$$\widehat{\mathcal{T}}_i(\boldsymbol{\beta}) = [\widehat{H}_1(V_i; \boldsymbol{\beta}) - \widehat{H}_2(V_i; \boldsymbol{\beta})]\left(1, \widehat{R}_{\beta_1}^{\mathrm{T}}(V_i), Z_i^{\mathrm{T}}\right)^{\mathrm{T}} \text{ for } \widehat{H}_1(V_i; \boldsymbol{\beta})$$
$$= H\left[\beta_0 + \beta_2^{\mathrm{T}} Z_i + \widehat{R}(V_i; \beta_1) + \ln \frac{1-\widehat{\theta}_0(V_i)}{\widehat{\phi}_0(V_i)}\right], \widehat{H}_2(V_i; \boldsymbol{\beta})$$
$$= H\left[\beta_0 + \beta_2^{\mathrm{T}} Z_i + \widehat{R}(V_i; \beta_1) + \ln \frac{\widehat{\theta}_0(V_i)}{1-\widehat{\phi}_0(V_i)}\right], \quad \text{and}$$
$$\widehat{R}_{\beta_1}(V_i) = \frac{\widehat{r}_{\beta_1}(0, V_i; \beta_1)[1-\widehat{\phi}_0(V_i)] + \widehat{r}_{\beta_1}(1, V_i; \beta_1)\widehat{\phi}_0(V_i)}{\widehat{r}_0(0, V_i; \beta_1)[1-\widehat{\phi}_0(V_i)] + \widehat{r}_0(1, V_i; \beta_1)\widehat{\phi}_0(V_i)}$$

with

$$\widehat{r}_{\beta_1}\left(Y_i^0, V_i; \beta_1\right) = \frac{\partial \widehat{r}_0\left(Y_i^0, V_i; \beta_1\right)}{\partial \beta_1}$$
$$= \frac{\sum_{k=1}^{n} \delta_k X_k e^{\beta_1^{\mathrm{T}} X_k} I\left(Y_k = 0, Y_k^0 = Y_i^0, V_k = V_i\right)}{\sum_{j=1}^{n} \delta_j I\left(Y_j = 0, Y_j^0 = Y_i^0, V_j = V_i\right)}.$$

Let $\widehat{\boldsymbol{\beta}}_J = \left(\hat{\beta}_{J0}, \hat{\beta}_{J1}^{\mathrm{T}}, \hat{\beta}_{J2}^{\mathrm{T}}\right)^{\mathrm{T}}$ denote the joint conditional estimator of $\boldsymbol{\beta}$, which is the root of the score equation $\widehat{U}_n(\boldsymbol{\beta}) = 0$ and obtained by the Newton-Raphson algorithm.

## 3 Asymptotic theory

To derive the asymptotic properties of the proposed estimator under the assumptions that $X$ and $Y$ are MAR and $(X^T, W^T, Z^T)$ are discrete, the following regularity conditions are required:

(A1) Let supp(V) denote the support of V. For any $y^0 = 0, 1$ and $v \in$ supp(V), the selection probability $\pi\left(y^0, v\right) > 0$.

(A2) For any $y^0 = 0, 1$ and $v \in$ supp(V), the selection probability $\pi\left(y^0, v\right) < 1$.

(A3) For any $v \in \mathrm{supp}(V)$, $\mathrm{E}\left(e^{\beta_1^{\mathrm{T}} X} | Y = 0, V = v\right)$ exists in a neighborhood of the true $\beta$.

(A4) $\mathrm{E}[\delta \mathcal{X}\mathcal{X}^{\mathrm{T}} H_+^{(1)}(X, V; \boldsymbol{\beta}) + (1-\delta)\mathcal{T}(\boldsymbol{\beta})\mathcal{T}(\boldsymbol{\beta})^{\mathrm{T}} H_-^{(1)}(V; \boldsymbol{\beta})]$ is positive definite in a neighborhood of the true $\boldsymbol{\beta}$, where $H_+^{(1)}(X, V; \boldsymbol{\beta}) = H_+(X, V; \boldsymbol{\beta})[1 - H_+(X, V; \boldsymbol{\beta})]$ and $H_-^{(1)}(V; \boldsymbol{\beta}) = H_-(V; \boldsymbol{\beta})[1 - H_-(V; \boldsymbol{\beta})]$.

(A5) The first derivatives of $U_n(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ exist almost surely in a neighborhood of the true $\boldsymbol{\beta}$. Further, in such a neighborhood, the second derivatives are bounded above by a function of $\left(Y^0, Y, X, V\right)$.

Using the estimated nonparametric selection probability, one can show that $\widehat{U}_{1n}(\boldsymbol{\beta})$ in (9) and $\widehat{U}_{2n}(\boldsymbol{\beta}) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^n (1-\delta_i)\,\widehat{\mathcal{T}_i}(\boldsymbol{\beta})[Y_i^0 - \widehat{H}_-(V_i; \boldsymbol{\beta})]$ can be expressed as the sum of independent random variables, which are stated in Lemmas 1 and 2, respectively, as follow.

**Lemma 1** *Under the condition (A1),*

$$\widehat{U}_{1n}(\boldsymbol{\beta}) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[S_c\left(Y_i, Y_i^0, X_i, V_i; \boldsymbol{\beta}\right) + \varepsilon_c\left(Y_i, Y_i^0, X_i, V_i; \boldsymbol{\beta}\right)\right] + o_p(1).$$

Here $S_c\left(Y_i, Y_i^0, X_i, V_i; \boldsymbol{\beta}\right) = \delta_i \mathcal{X}_i[Y_i - H_+(X_i, V_i; \boldsymbol{\beta})]$. The definition of $\varepsilon_c\left(Y_i, Y_i^0, X_i, V_i; \boldsymbol{\beta}\right)$ and the proof of Lemma 1 are given in the Appendix. Note that the $\varepsilon_c(\cdot)$ stands for the approximation error from the complete data score $S_c(\cdot)$, which is due to the estimation of nuisance parameters $\pi\left(Y^0, V\right)$, $\theta(X, V)$, and $\phi(X, V)$.

We linearize $\widehat{U}_{2n}(\boldsymbol{\beta})$ in the following Lemma.

**Lemma 2** *Under the conditions (A2)–(A3),*

$$\widehat{U}_{2n}(\boldsymbol{\beta}) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[S_m\left(Y_i^0, V_i; \boldsymbol{\beta}\right) + \varepsilon_m\left(Y_i, Y_i^0, X_i, V_i; \boldsymbol{\beta}\right)\right] + o_p(1).$$

Here $S_m\left(Y_i^0, V_i; \boldsymbol{\beta}\right) = (1-\delta_i)\,\mathcal{T}_i(\boldsymbol{\beta})[Y_i^0 - H_-(V_i; \boldsymbol{\beta})]$. The definition of $\varepsilon_m\left(Y_i, Y_i^0, X_i, V_i; \boldsymbol{\beta}\right)$ and the proof of Lemma 2 are given in the Appendix. Note that the $S_m(\cdot)$

stands for the score from the case with missing covariates, and $\varepsilon_m(\cdot)$ stands for the approximation error from $S_m(\cdot)$, which is due to the estimation of nuisance parameters $\pi\left(Y^0, V\right)$, $R\left(V;\beta_1\right)$, $\theta_0(V)$, and $\phi_0(V)$.

Concerning the asymptotic properties of $\widehat{\boldsymbol{\beta}}_J$, we are able to show the following.

**Theorem 1** *Under conditions (A1)–(A5), the joint conditional likelihood estimator $\widehat{\boldsymbol{\beta}}_J$ is a consistent estimate of $\boldsymbol{\beta}$, and $\sqrt{n}(\widehat{\boldsymbol{\beta}}_J - \boldsymbol{\beta})$ has an asymptotic normal distribution with mean 0 and covariance matrix $\Delta_J = G^{-1}(\boldsymbol{\beta})M(\boldsymbol{\beta})[G^{-1}(\boldsymbol{\beta})]^{\mathrm{T}}$. Here*

$$
\begin{aligned}
G(\boldsymbol{\beta}) = E\Big[&\delta\mathcal{X}\mathcal{X}^{\mathrm{T}}H_{+}^{(1)}(X, V;\boldsymbol{\beta}) \\
&+(1-\delta)\mathcal{T}(\boldsymbol{\beta})\mathcal{T}^{\mathrm{T}}(\boldsymbol{\beta})H_{-}^{(1)}(V;\boldsymbol{\beta})\Big].\\
M(\boldsymbol{\beta}) = E\Big\{\Big[&S_c\left(Y, Y^0, X, V;\boldsymbol{\beta}\right) + S_m\left(Y^0, V;\boldsymbol{\beta}\right) + \varepsilon_c\left(Y, Y^0, X, V;\boldsymbol{\beta}\right)\\
&+\varepsilon_m\left(Y, Y^0, X, V;\boldsymbol{\beta}\right)\Big]^{\otimes 2}\Big\},
\end{aligned}
$$

*where, for any vector $a$, $a^{\otimes 2} = aa^T$.*

The proof of Theorem 1 and a consistent estimator of $\Delta_J$ are given in the Appendix.

## 4 Some other estimators

In this section, we review two existing estimators, namely, the complete-case (CC) estimator and the weighted (WE) estimator.

### 4.1 Complete-case estimator

The complete-case analysis is a naive method that ignores the missing data and uses data in the validation set only. The resulting estimator, denoted by $\widehat{\boldsymbol{\beta}}_C$, is the solution of the following estimating equation:

$$
U_{cn}(\boldsymbol{\beta}) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i\mathcal{X}_i\left[Y_i - H\left(\boldsymbol{\beta}^T\mathcal{X}_i\right)\right] = 0.
$$

It is known that $\widehat{\boldsymbol{\beta}}_C$ is guaranteed to be unbiased only when the verification sample is a simple random sample of subjects in the study. If, for example, those at higher risk of disease are more likely to be selected for verification, then $\widehat{\boldsymbol{\beta}}_C$ will be biased upwards.

### 4.2 Weighted estimator

The weighted estimation approach with a class of estimators has been motivated by Horvitz and Thompson (1952). When $Y$ is MAR, Chen and Breslow (2004) showed that a semiparametric efficient estimator may also be obtained by optimizing a class of Horvitz and Thompson's estimator with estimated weights. When $X$ and $Y$ are MAR, we compute the weighted estimator, denoted by $\widehat{\boldsymbol{\beta}}_W$, which is defined as the solution to the following equation:

$$U_{wn}(\boldsymbol{\beta}) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{\pi\left(Y_i^0, V_i\right)} \mathcal{X}_i \left[Y_i - H\left(\boldsymbol{\beta}^T \mathcal{X}_i\right)\right] = 0.$$

It is not difficult to see that when the $\pi\left(Y_i^0, V_i\right)$'s are known, $\mathrm{E}\left\{\frac{\delta_i}{\pi(Y_i^0, V_i)} \mathcal{X}_i \left[Y_i - H\left(\boldsymbol{\beta}^T \mathcal{X}_i\right)\right]\right\} = 0$ and the above estimating equation is unbiased. When the $\pi\left(Y_i^0, V_i\right)$'s are unknown, the weighted estimator of $\boldsymbol{\beta}$ is now $\widehat{\boldsymbol{\beta}}_W$, which solves $\widehat{U}_{wn}(\boldsymbol{\beta}) = 0$, where $\widehat{\boldsymbol{\beta}}_W$ is defined similarly to that which is the solution to $U_{wn}(\boldsymbol{\beta}) = 0$, but with $\pi\left(Y_i^0, V_i\right)$ replaced by $\widehat{\pi}\left(Y_i^0, V_i\right)$. The $\widehat{\boldsymbol{\beta}}_W$ is consistent. $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}_W - \boldsymbol{\beta}\right)$ has an asymptotic normality with mean 0 and covariance matrix $\Delta_W$ that can be obtained by the sandwich method and consistently estimated by

$$\mathcal{G}^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}\left[\frac{\delta_i}{\widehat{\pi}(Y_i^0, V_i)}\phi_i + \left[1 - \frac{\delta_i}{\widehat{\pi}(Y_i^0, V_i)}\right]\widehat{\phi}_i^*\right]\left[\frac{\delta_i}{\widehat{\pi}(Y_i^0, V_i)}\phi_i + \left[1 - \frac{\delta_i}{\widehat{\pi}(Y_i^0, V_i)}\right]\widehat{\phi}_i^*\right]^T\right\}\mathcal{G}^{-1},$$

where $\mathcal{G} = \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}(Y_i^0, V_i)} \mathcal{X}_i \mathcal{X}_i^{\mathrm{T}} H\left(\widehat{\boldsymbol{\beta}}_W^{\mathrm{T}} \mathcal{X}_i\right)\left[1 - H\left(\widehat{\boldsymbol{\beta}}_W^{\mathrm{T}} \mathcal{X}_i\right)\right]$, $\phi_i = \mathcal{X}_i\left[Y_i - H\left(\widehat{\boldsymbol{\beta}}_W^{\mathrm{T}} \mathcal{X}_i\right)\right]$, and $\widehat{\phi}_i^* = \widehat{\mathrm{E}}\left[\phi_i | Y_i^0, V_i\right] = \sum_{j=1}^{n} \delta_j \phi_j I\left(Y_j^0 = Y_i^0, V_j = V_i\right) / \sum_{s=1}^{n} \delta_s I\left(Y_s^0 = Y_i^0, V_s = V_i\right)$. Note that $X_i$ and $Y_i$ are not observed, and the average score $\widehat{\phi}_i^*$ is obtained from the validation set.

## 5 Simulation study

In this section, we present Monte Carlo results to investigate the finite-sample performances of the four estimators including the maximum likelihood estimator from full data $\widehat{\boldsymbol{\beta}}_F$, which is used to provide a useful comparison benchmark, the CC estimator $\widehat{\boldsymbol{\beta}}_C$, the validation likelihood (VAL) estimator $\widehat{\boldsymbol{\beta}}_V$, the joint conditional likelihood (JCL) estimator $\widehat{\boldsymbol{\beta}}_J$, and the WE estimator $\widehat{\boldsymbol{\beta}}_W$. For each configuration of the experiments, 1000 replications were conducted. The sample sizes are $n = 500$ and 1000. For each estimator, we computed bias, asymptotic standard error (ASE), standard deviation (SD), and coverage probabilities (CP) of a 95% confidence interval.

We consider the case of univariate covariate $X$. First, $U$'s and $\varepsilon$'s were generated independently from normal distribution $N(0, 1)$ and $N(0, 1)$, respectively. Given $U$, a binary covariate was defined by $X = I(U \geq 0)$. Given $U$ and $\varepsilon$, a binary surrogate covariate was defined by $W = I(U + \sigma\varepsilon \geq 0)$. When $\sigma = 0.25$, the correlation coefficient of $X$ and $W$, denoted by corr$(X, W)$, was about 0.84. The response $Y$ was generated as a binary with $P(Y = 1|X) = H(\beta_0 + \beta_1 X)$, where $\boldsymbol{\beta} = (\beta_0, \beta_1) = (\ln(2), \ln(3))^{\mathrm{T}}$. The surrogate outcome $Y^0$ was a misclassified version of $Y$ such that $P\left(Y^0 = 1|Y, X\right) = 0.2 + pY$, where $0 \leq p \leq 0.8$. When

**Table 1** Simulation results (univariate covariate X)

| $mr$ | $n$ | | | $\widehat{\boldsymbol{\beta}}_F$ | $\widehat{\boldsymbol{\beta}}_C$ | $\widehat{\boldsymbol{\beta}}_V$ | $\widehat{\boldsymbol{\beta}}_J$ | $\widehat{\boldsymbol{\beta}}_W$ | $\text{RE}_1$ | $\text{RE}_2$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 34% | 500 | $\beta_0$ | bias | 0.0016 | 0.1327 | 0.0011 | 0.0017 | 0.0010 | | |
| | | | SD | 0.1376 | 0.1703 | 0.1490 | 0.1486 | 0.1509 | | |
| | | | ASE | 0.1347 | 0.1660 | 0.1448 | 0.1434 | 0.1455 | 1.020 | 1.030 |
| | | | CP | 0.9450 | 0.8750 | 0.9440 | 0.9440 | 0.9410 | | |
| | | $\beta_1$ | bias | 0.0019 | 0.0499 | 0.0070 | 0.0048 | 0.0068 | | |
| | | | SD | 0.2276 | 0.3054 | 0.2679 | 0.2620 | 0.2682 | | |
| | | | ASE | 0.2267 | 0.2998 | 0.2618 | 0.2538 | 0.2620 | 1.064 | 1.066 |
| | | | CP | 0.9510 | 0.9400 | 0.9500 | 0.9500 | 0.9490 | | |
| | 1000 | $\beta_0$ | bias | 0.0006 | 0.1348 | 0.0010 | 0.0017 | 0.0009 | | |
| | | | SD | 0.0930 | 0.1156 | 0.0987 | 0.0993 | 0.0993 | | |
| | | | ASE | 0.0950 | 0.1171 | 0.1021 | 0.1011 | 0.1025 | 1.020 | 1.028 |
| | | | CP | 0.9560 | 0.8090 | 0.9560 | 0.9550 | 0.9590 | | |
| | | $\beta_1$ | bias | 0.0018 | 0.0398 | 0.0001 | −0.0020 | 0.0001 | | |
| | | | SD | 0.1554 | 0.2146 | 0.1823 | 0.1791 | 0.1829 | | |
| | | | ASE | 0.1598 | 0.2101 | 0.1835 | 0.1784 | 0.1836 | 1.059 | 1.059 |
| | | | CP | 0.9570 | 0.9420 | 0.9480 | 0.9500 | 0.9520 | | |
| 58% | 500 | $\beta_0$ | bias | 0.0016 | 0.2278 | 0.0031 | 0.0072 | 0.0021 | | |
| | | | SD | 0.1376 | 0.2088 | 0.1636 | 0.1625 | 0.1677 | | |
| | | | ASE | 0.1347 | 0.2094 | 0.1596 | 0.1530 | 0.1619 | 1.088 | 1.120 |
| | | | CP | 0.9450 | 0.8190 | 0.9480 | 0.9320 | 0.9470 | | |
| | | $\beta_1$ | bias | 0.0019 | 0.0714 | 0.0121 | 0.0009 | 0.0140 | | |
| | | | SD | 0.2276 | 0.3895 | 0.3111 | 0.2888 | 0.3107 | | |
| | | | ASE | 0.2267 | 0.3989 | 0.3139 | 0.2823 | 0.3135 | 1.236 | 1.233 |
| | | | CP | 0.9510 | 0.9730 | 0.9460 | 0.9400 | 0.9440 | | |
| | 1000 | $\beta_0$ | bias | 0.0006 | 0.2231 | 0.0042 | 0.0058 | 0.0045 | | |
| | | | SD | 0.0930 | 0.1484 | 0.1112 | 0.1118 | 0.1133 | | |
| | | | ASE | 0.0950 | 0.1469 | 0.1124 | 0.1079 | 0.1142 | 1.085 | 1.120 |
| | | | CP | 0.9560 | 0.6790 | 0.9550 | 0.9410 | 0.9510 | | |
| | | $\beta_1$ | bias | 0.0018 | 0.0469 | 0.0070 | 0.0020 | 0.0056 | | |
| | | | SD | 0.1554 | 0.2729 | 0.2100 | 0.1981 | 0.2104 | | |
| | | | ASE | 0.1598 | 0.2768 | 0.2188 | 0.1970 | 0.2191 | 1.234 | 1.237 |
| | | | CP | 0.9570 | 0.9540 | 0.9580 | 0.9540 | 0.9640 | | |

About 58% of X are missing in 1000 replications
The true parameter vector = $(\beta_0, \beta_1)^{\mathrm{T}} = (\ln(2), \ln(3))^{\mathrm{T}}$
$\mathrm{corr}(X, W) = 0.84(\sigma = 0.25)$, $\mathrm{corr}(Y^0, Y) = 0.87(p = 0.8)$
$\mathrm{RE}_1 = \mathrm{ASV}(\widehat{\boldsymbol{\beta}}_V)/\mathrm{ASV}(\widehat{\boldsymbol{\beta}}_J)$ and $\mathrm{RE}_2 = \mathrm{ASV}(\widehat{\boldsymbol{\beta}}_W)/\mathrm{ASV}(\widehat{\boldsymbol{\beta}}_J)$

$p = 0.8$, the correlation coefficient of $Y$ and $Y^0$, denoted by $\mathrm{corr}\left(Y, Y^0\right)$, resulted in about 0.86. Given $W$ and $Y^0$, the binary indicator $\delta$ was with the probability $P\left(\delta = 1|Y^0, W\right) = H\left(\alpha_0 + \alpha_1 Y^0 + \alpha_2 W\right)$, where the values of $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)^{\mathrm{T}}$ were set at $(0.5, 0.5, -0.5)$ and $(-0.5, 0.5, -0.5)$, which resulted in about 0.34 and 0.58 of missing rate ($mr$), respectively.

Table 1 shows that the efficiencies of all estimators increase as the sample size increases. The maximum likelihood estimator from full data $\widehat{\boldsymbol{\beta}}_F$ outperforms the other estimators and is used to provide a useful comparison benchmark, but it is of limited practical value because its use requires full simulated data without missing values. The CC estimator was seriously biased. The VAL and WE estimators had similar performances because both methods are based on the complete dataset accounting for the probability of being observed. When the sample size is $n = 500$ with a missing rate of about 58%, the empirical coverage probability of JCL estimator is slightly lower than the nominal coverage 0.95, but increasing sample size always improves the coverage probability. The asymptotic standard errors (ASE) of JCL estimators are smaller than standard deviations (SD) of JCL estimators. We computed the ratio of the asymptotic variances (ASV) of $\widehat{\boldsymbol{\beta}}_V$ and $\widehat{\boldsymbol{\beta}}_W$ to that of $\widehat{\boldsymbol{\beta}}_J$. The relative efficiencies were defined by $\mathrm{RE}_1 = \mathrm{ASV}\left(\widehat{\boldsymbol{\beta}}_V\right)/\mathrm{ASV}\left(\widehat{\boldsymbol{\beta}}_J\right)$ and $\mathrm{RE}_2 = \mathrm{ASV}\left(\widehat{\boldsymbol{\beta}}_W\right)/\mathrm{ASV}\left(\widehat{\boldsymbol{\beta}}_J\right)$. Both the $\mathrm{RE}_1$ and $\mathrm{RE}_2$ were bigger than one, which implies that the $\widehat{\boldsymbol{\beta}}_J$ outperformed the two estimators $\widehat{\boldsymbol{\beta}}_V$ and $\widehat{\boldsymbol{\beta}}_W$. Moreover, when the missing rate is higher (0.58), the $\widehat{\boldsymbol{\beta}}_J$ was more efficient than the $\widehat{\boldsymbol{\beta}}_V$ and $\widehat{\boldsymbol{\beta}}_W$.

We also investigate the efficiencies of the three estimators $\widehat{\boldsymbol{\beta}}_V$, $\widehat{\boldsymbol{\beta}}_J$, and $\widehat{\boldsymbol{\beta}}_W$ under various values of $\mathrm{corr}(X, W)$ and $\mathrm{corr}\left(Y, Y^0\right)$. Tables 2 and 3 present the Monte Carlo results under the case in which $n = 500$, $\boldsymbol{\alpha} = (-0.5, 0.5, -0.5)$ and $mr = 0.58$. As shown in Table 2, the $p = 0.8$ is fixed, which implies that $\mathrm{corr}\left(Y, Y^0\right)$ was about 0.87. We selected various values of $\sigma$ so that $\mathrm{corr}(X, W)$ ranged from 0.24 to 0.84. When the $W$ was highly informative about $X$ (i.e., the $\mathrm{cor}(X, W)$ was large), the $\widehat{\boldsymbol{\beta}}_J$ was more efficient than the two estimators $\widehat{\boldsymbol{\beta}}_V$ and $\widehat{\boldsymbol{\beta}}_W$.

As seen in Table 3, the $\sigma = 0.25$ is fixed, which implies that $\mathrm{corr}(X, W)$ resulted in about 0.84. We selected various $p$ values so that $\mathrm{corr}\left(Y, Y^0\right)$ ranged from 0.26 to 0.86. Note that the missing rate increased as the $p$ decreased. When the $\mathrm{cor}\left(Y, Y^0\right)$ was not large, the $\widehat{\boldsymbol{\beta}}_J$ still performed slightly better than both the estimators $\widehat{\boldsymbol{\beta}}_V$ and $\widehat{\boldsymbol{\beta}}_W$. Overall, the efficiency of the JCL estimator increased as the missing rate, $\mathrm{corr}(X, W)$, and $\mathrm{corr}\left(Y, Y^0\right)$ increased.

## 6 Example

We apply the proposed methods to the data from a cable TV survey study in Taiwan, 2004. The 1793 respondents in the survey study are residents of three cities in Taiwan. The binary outcome variable $Y$ is the response (1 = Yes; 0 = No) to the question "Have you been given a discount on cable TV?". The covariates are the types of living places, denoted by $X$ (1. building or apartment with management committee; 2. house or condo with management committee; 3. apartment or condo without management committee) and the city of residence, denoted by $Z$ (1 = Taipei; 2 = Yunlin; 3 = Taichung). In this study, due to item non-response, $Y$ and $X$ are not available for some respondents. Therefore, there are 1455 respondents in the validation data set and the missing rate of $Y$ and $X$ is 19%. The surrogate variable of $Y$ is $Y^0$, which is the response of payment methods (1. six months or a year of payment; 0. one month or three months of payment). The surrogate variable of $X$ is $W$, which is the

**Table 2** Simulation results for various values of corr$(X, W)$ $(n = 500)$

| corr$(X, W)$ | $\sigma$ | $mr$ | | | $\widehat{\boldsymbol{\beta}}_F$ | $\widehat{\boldsymbol{\beta}}_C$ | $\widehat{\boldsymbol{\beta}}_V$ | $\widehat{\boldsymbol{\beta}}_J$ | $\widehat{\boldsymbol{\beta}}_W$ | $\mathrm{RE}_1$ | $\mathrm{RE}_2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.84 | 0.25 | 0.58 | $\beta_0$ | bias | 0.0016 | 0.2278 | 0.0031 | 0.0072 | 0.0021 | | |
| | | | | SD | 0.1376 | 0.2088 | 0.1636 | 0.1625 | 0.1677 | | |
| | | | | ASE | 0.1347 | 0.2094 | 0.1596 | 0.1530 | 0.1619 | 1.088 | 1.120 |
| | | | | CP | 0.9450 | 0.8190 | 0.9480 | 0.9320 | 0.9470 | | |
| | | | $\beta_1$ | bias | 0.0019 | 0.0714 | 0.0121 | 0.0009 | 0.0140 | | |
| | | | | SD | 0.2276 | 0.3895 | 0.3111 | 0.2888 | 0.3107 | | |
| | | | | ASE | 0.2267 | 0.3989 | 0.3139 | 0.2823 | 0.3135 | 1.236 | 1.233 |
| | | | | CP | 0.9510 | 0.9730 | 0.9460 | 0.9400 | 0.9440 | | |
| 0.66 | 0.6 | 0.58 | $\beta_0$ | bias | 0.0016 | 0.2309 | 0.0031 | 0.0088 | 0.0020 | | |
| | | | | SD | 0.1376 | 0.2119 | 0.1709 | 0.1721 | 0.1756 | | |
| | | | | ASE | 0.1347 | 0.2122 | 0.1680 | 0.1628 | 0.1716 | 1.065 | 1.111 |
| | | | | CP | 0.9450 | 0.8210 | 0.9390 | 0.9330 | 0.9440 | | |
| | | | $\beta_1$ | bias | 0.0019 | 0.0603 | 0.0172 | 0.0006 | 0.0164 | | |
| | | | | SD | 0.2276 | 0.3856 | 0.3461 | 0.3201 | 0.3455 | | |
| | | | | ASE | 0.2267 | 0.3945 | 0.3523 | 0.3228 | 0.3537 | 1.191 | 1.201 |
| | | | | CP | 0.9510 | 0.9700 | 0.9620 | 0.9540 | 0.9650 | | |
| 0.50 | 1 | 0.58 | $\beta_0$ | bias | 0.0016 | 0.2341 | 0.0028 | 0.0140 | 0.0015 | | |
| | | | | SD | 0.1376 | 0.2144 | 0.1767 | 0.1784 | 0.1820 | | |
| | | | | ASE | 0.1347 | 0.2145 | 0.1734 | 0.1695 | 0.1771 | 1.047 | 1.092 |
| | | | | CP | 0.9450 | 0.8220 | 0.9450 | 0.9360 | 0.9380 | | |
| | | | $\beta_1$ | bias | 0.0019 | 0.0489 | 0.0219 | −0.0033 | 0.0216 | | |
| | | | | SD | 0.2276 | 0.3809 | 0.3657 | 0.3413 | 0.3663 | | |
| | | | | ASE | 0.2267 | 0.3907 | 0.3705 | 0.3507 | 0.3735 | 1.116 | 1.134 |
| | | | | CP | 0.9510 | 0.9630 | 0.9620 | 0.9490 | 0.9620 | | |
| 0.24 | 2.5 | 0.58 | $\beta_0$ | bias | 0.0016 | 0.2407 | 0.0023 | 0.0229 | 0.0025 | | |
| | | | | SD | 0.1376 | 0.2203 | 0.1842 | 0.1843 | 0.1884 | | |
| | | | | ASE | 0.1347 | 0.2188 | 0.1805 | 0.1777 | 0.1828 | 1.032 | 1.058 |
| | | | | CP | 0.9450 | 0.8240 | 0.9400 | 0.9450 | 0.9430 | | |
| | | | $\beta_1$ | bias | 0.0019 | 0.0326 | 0.0257 | −0.0139 | 0.0238 | | |
| | | | | SD | 0.2276 | 0.3764 | 0.3834 | 0.3659 | 0.3847 | | |
| | | | | ASE | 0.2267 | 0.3868 | 0.3850 | 0.3807 | 0.3896 | 1.023 | 1.047 |
| | | | | CP | 0.9510 | 0.9650 | 0.9610 | 0.9580 | 0.9650 | | |

About 58% of X are missing in 1000 replications
The true parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1)^{\mathrm{T}} = (\ln(2), \ln(3))^{\mathrm{T}}$
$\mathrm{corr}(Y^0, Y) = 0.87 (p = 0.8)$
$\mathrm{RE}_1 = \mathrm{ASV}(\widehat{\boldsymbol{\beta}}_V)/\mathrm{ASV}(\widehat{\boldsymbol{\beta}}_J)$ and $\mathrm{RE}_2 = \mathrm{ASV}(\widehat{\boldsymbol{\beta}}_W)/\mathrm{ASV}(\widehat{\boldsymbol{\beta}}_J)$

response (1 = Yes; 0 = No) to the question “Would you pay extra money for additional channels?”. We consider the following logistic regression model

$$P\left(Y_i = 1|X_i, Z_i, W_i\right) = H\left(\beta_0 + \beta_1 DX_{1i} + \beta_2 DX_{2i} + \beta_3 DZ_{1i} + \beta_4 DZ_{2i}\right)$$

**Table 3** Simulation results for various values of corr$(Y^0, Y)$ $(n = 500)$

| corr$(Y^0, Y)$ | $p$ | $mr$ | | | $\widehat{\boldsymbol{\beta}}_F$ | $\widehat{\boldsymbol{\beta}}_C$ | $\widehat{\boldsymbol{\beta}}_V$ | $\widehat{\boldsymbol{\beta}}_J$ | $\widehat{\boldsymbol{\beta}}_W$ | $\text{RE}_1$ | $\text{RE}_2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.87 | 0.8 | 0.58 | $\beta_0$ | bias | 0.0016 | 0.2278 | 0.0031 | 0.0072 | 0.0021 | | |
| | | | | SD | 0.1376 | 0.2088 | 0.1636 | 0.1625 | 0.1677 | | |
| | | | | ASE | 0.1347 | 0.2094 | 0.1596 | 0.1530 | 0.1619 | 1.088 | 1.120 |
| | | | | CP | 0.9450 | 0.8190 | 0.9480 | 0.9320 | 0.9470 | | |
| | | | $\beta_1$ | bias | 0.0019 | 0.0714 | 0.0121 | 0.0009 | 0.0140 | | |
| | | | | SD | 0.2276 | 0.3895 | 0.3111 | 0.2888 | 0.3107 | | |
| | | | | ASE | 0.2267 | 0.3989 | 0.3139 | 0.2823 | 0.3135 | 1.236 | 1.233 |
| | | | | CP | 0.9510 | 0.9730 | 0.9460 | 0.9400 | 0.9440 | | |
| 0.68 | 0.7 | 0.59 | $\beta_0$ | bias | 0.0100 | 0.2109 | 0.0106 | 0.0117 | 0.0107 | | |
| | | | | SD | 0.1325 | 0.2078 | 0.1797 | 0.1826 | 0.1812 | | |
| | | | | ASE | 0.1348 | 0.2103 | 0.1808 | 0.1811 | 0.1826 | 0.997 | 1.017 |
| | | | | CP | 0.9560 | 0.8590 | 0.9460 | 0.9490 | 0.9470 | | |
| | | | $\beta_1$ | bias | −0.0013 | 0.0663 | 0.0208 | 0.0116 | 0.0208 | | |
| | | | | SD | 0.2243 | 0.4098 | 0.3713 | 0.3696 | 0.3726 | | |
| | | | | ASE | 0.2271 | 0.4010 | 0.3634 | 0.3578 | 0.3647 | 1.032 | 1.039 |
| | | | | CP | 0.9540 | 0.9520 | 0.9410 | 0.9400 | 0.9450 | | |
| 0.54 | 0.6 | 0.60 | $\beta_0$ | bias | 0.0100 | 0.1839 | 0.0108 | 0.0122 | 0.0108 | | |
| | | | | SD | 0.1325 | 0.2095 | 0.1941 | 0.1960 | 0.1952 | | |
| | | | | ASE | 0.1348 | 0.2111 | 0.1930 | 0.1931 | 0.1944 | 0.999 | 1.014 |
| | | | | CP | 0.9560 | 0.8880 | 0.9430 | 0.9390 | 0.9460 | | |
| | | | $\beta_1$ | bias | −0.0013 | 0.0632 | 0.0266 | 0.0174 | 0.0272 | | |
| | | | | SD | 0.2243 | 0.4103 | 0.3926 | 0.3907 | 0.3940 | | |
| | | | | ASE | 0.2271 | 0.4018 | 0.3827 | 0.3799 | 0.3841 | 1.015 | 1.022 |
| | | | | CP | 0.9540 | 0.9510 | 0.9470 | 0.9460 | 0.9490 | | |
| 0.26 | 0.3 | 0.63 | $\beta_0$ | bias | 0.0100 | 0.1020 | 0.0126 | 0.0134 | 0.0138 | | |
| | | | | SD | 0.1325 | 0.2109 | 0.2073 | 0.2077 | 0.2077 | | |
| | | | | ASE | 0.1348 | 0.2137 | 0.2113 | 0.2112 | 0.2122 | 1.001 | 1.009 |
| | | | | CP | 0.9560 | 0.9370 | 0.9500 | 0.9540 | 0.9480 | | |
| | | | $\beta_1$ | bias | −0.0013 | 0.0531 | 0.0341 | 0.0314 | 0.0339 | | |
| | | | | SD | 0.2243 | 0.4131 | 0.4113 | 0.4095 | 0.4121 | | |
| | | | | ASE | 0.2271 | 0.4048 | 0.4050 | 0.4045 | 0.4066 | 1.002 | 1.010 |
| | | | | CP | 0.9540 | 0.9520 | 0.9520 | 0.9560 | 0.9540 | | |

About 58% of X are missing in 1000 replications
The true parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1)^{\mathrm{T}} = (\ln(2), \ln(3))^{\mathrm{T}}$
$\text{corr}(X, W) = 0.84(\sigma = 0.25)$
$\text{RE}_1 = \text{ASV}(\widehat{\boldsymbol{\beta}}_V)/\text{ASV}(\widehat{\boldsymbol{\beta}}_J)$ and $\text{RE}_2 = \text{ASV}(\widehat{\boldsymbol{\beta}}_W)/\text{ASV}(\widehat{\boldsymbol{\beta}}_J)$

for $i = 1, 2, \ldots, 1793$. Here $DX_1$ and $DX_2$ are dummy variables for the type of living place $X$, $DX_k = 1$ if $X = k$ and 0 otherwise. $DZ_1$ and $DZ_2$ are dummy variables for the city of residence $Z$, $DZ_k = 1$ if $Z = k$ and 0 otherwise.

**Table 4** Results of cable TV survey data analysis ($mr = 19\%$)

| Variable | Parameter | $\widehat{\boldsymbol{\beta}}_V$ | $\widehat{\boldsymbol{\beta}}_C$ | $\widehat{\boldsymbol{\beta}}_W$ | $\widehat{\boldsymbol{\beta}}_J$ |
|---|---|---|---|---|---|
| Intercept | $\beta_0$ | −1.7087 | −1.7431 | −1.6911 | −1.7082 |
| | | (0.1850) | (0.1837) | (0.1853) | (0.1850) |
| $DX_1$ | $\beta_1$ | 0.7398 | 0.7404 | 0.6889 | 0.7402 |
| | | (0.1517) | (0.1513) | (0.1518) | (0.1518) |
| $DX_2$ | $\beta_2$ | 1.1542 | 1.1058 | 1.1691 | 1.1555 |
| | | (0.3191) | (0.3149) | (0.3235) | (0.3197) |
| $DZ_1$ | $\beta_3$ | 0.6409 | 0.6795 | 0.6417 | 0.6410 |
| | | (0.1838) | (0.1825) | (0.1834) | (0.1839) |
| $DZ_2$ | $\beta_4$ | 0.2936 | 0.2852 | 0.2768 | 0.2944 |
| | | (0.2223) | (0.2217) | (0.2224) | (0.2241) |

Values in parentheses, (.), are the asymptotic standard error (ASE) of the estimates. $DX_1$ and $DX_2$ are dummy variable for type of living place $X$ (1 = building or apartment with management committee; 2 = house or condo with management committee; 3 = apartment or condo without management committee), $DX_k = 1$ if $X = k$ and 0 otherwise. $DZ_1$ and $DZ_2$ are dummy variables for city of residence $Z$ (1 = Taipei; 2 = Yunlin; 3 = Taichung), $DZ_k = 1$ if $Z = k$ and 0 otherwise

To examine the missingness mechanism of the cable TV survey data, we conduct a logistic regression model with outcome variable $\delta$ and the covariate vector $\left(Y^0, W, DZ_1, DZ_2\right)$. The estimate of the corresponding regression coefficient vector is (−0.371, −0.260, 0.691, 1.022) with asymptotic standard error (0.125, 0.141, 0.145, 0.183). This result suggests that the non-response of $Y$ and $X$ are related to $Y^0$, $W$, and $Z$, indicating that MAR missingness mechanism may be plausible.

We analyze the data by assuming MAR missingness mechanism. The analysis results are given in Table 4. Recall that CC estimators are inconsistent and the analysis results from the CC regression can be misleading. The results of testing $\beta_1 = 0$, $\beta_2 = 0$, and $\beta_3 = 0$ each based on all the methods are statistically significant. Note that the estimates of $\beta_1$ and $\beta_2$ are statistically significantly different from zero and positive, which indicates that discount was more likely to occur when the living place is a building or apartment with management committee or a house or condo with management committee. The results of testing $\beta_4 = 0$ based on all the JCL, VL and WE methods are statistically insignificant, meaning that there was no difference in discount on the cable TV between the Taichung customers and Yunlin customers. Finally, the analysis results based on the JCL, VL and WE methods are close to each other, which are consistent with the results of the simulation study in Sect. 5 when the missing rate is low.

## 7 Conclusion

We have proposed a semiparametric approach to estimate the logistic regression model with missing both covariates and an outcome variable. As demonstrated in the simulation studies, the JCL estimator based on both validation and non-validation data sets performs slightly better than the other estimators except the maximum likelihood estimator from full data. When the missing rate is high, and $\mathrm{corr}(X, W)$ and $\mathrm{corr}(Y^0, Y)$ are not low, the JCL estimator can be much more efficient than both the VL and WE estimators. However, when $\mathrm{corr}(X, W)$ or $\mathrm{corr}\left(Y^0, Y\right)$ is low, the JCL estimator performs slightly better than both the VL and WE estimators. Note that the performances of the VL and WE estimators are similar. The advantage of our estimation is that it is unnecessary to make any assumptions of nuisance components such as the selection probability, misclassification probability, and conditional probability density of $X$ given $(Y, V)$.

Although the main results are presented for the case where $V$ and $X$ are discrete, it can be extended to the continuous case by using the approach of Wang and Wang (1997). Nonparametric kernel techniques are required for extending our approach because the nuisance components involve the estimators of selection probabilities, misclassification probability, and the relative risk $\mathrm{E}\left(e^{\beta_1^{\mathrm{T}} X} | Y = 0, V\right)$. Moreover, we remark that the proposed semiparametric estimation method generalizes the methods of Wang et al. (2002) and Pepe (1992). When $Y = Y^0$ with probability one, all the misclassification probabilities, i.e., $\theta(\cdot)$, $\phi(\cdot)$, $\theta_0(\cdot)$, and $\phi_0(\cdot)$, are zero, which reduces to the case considered by Wang et al. (2002). On the other hand, if $X = W$ is observable and the missing mechanism of $Y$ is MCAR, this reduces to the case considered by Pepe (1992).

**Acknowledgments** The authors thank two referees whose helpful comments improved the presentation. The research of S.M. Lee and S.H. Hsieh was supported by National Science Council grant of Taiwan, ROC, 97-2118-M-035-002-MY2 and 99-2811-M035-001 respectively. This publication was made possible by Grant Number UL1 RR024146 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research (C.S. Li).



## Appendix

We assume that the regularity Conditions (A1)–(A5) of Sect. 3 hold. In this appendix we prove Lemmas 1 and 2, and Theorem 1. Note that Conditions (A1) and (A2) are common assumptions for the existence of logarithmic transformations of selection probabilities. Condition (A3) is a usual assumption for the existence of expectation. Condition (A4) is a usual assumption for the unique solution of estimating equation. Condition (A5) is a usual assumption for the proof of consistency in estimating equation theory.

*Proof of Lemma 1* By a Taylor expansion of $\widehat{U}_{1n}(\boldsymbol{\beta})$ at $\pi(0, V_i)$, $\pi(1, V_i)$, $A(X_i, 0, V_i)$, $A(X_i, 1, V_i)$, $B(X_i, 0, V_i)$, and $B(X_i, 1, V_i)$, we can express $\widehat{U}_{1n}(\boldsymbol{\beta})$ as

$$\widehat{U}_{1n}(\boldsymbol{\beta}) - U_{1n}(\boldsymbol{\beta}) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \mathcal{X}_i \left[\widehat{H}_{+}(X_i, V_i; \boldsymbol{\beta}) - H_{+}(X_i, V_i; \boldsymbol{\beta})\right]$$

$$= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i\mathcal{X}_i H_+^{(1)}(X_i, V_i; \boldsymbol{\beta})\,\mathcal{K}_1^{-1}(X_i, V_i)\Big\{C_1(X_i, V_i) \times [\widehat{\pi}(1, V_i) - \pi(1, V_i)] + C_2(X_i, V_i)[\widehat{\pi}(0, V_i) - \pi(0, V_i)] + C_3(X_i, V_i)\big[\widehat{A}(X_i, 1, V_i) - A(X_i, 1, V_i)\big] + C_4(X_i, V_i)\big[\widehat{A}(X_i, 0, V_i) - A(X_i, 0, V_i)\big] + C_5(X_i, V_i)\big[\widehat{B}(X_i, 1, V_i) - B(X_i, 1, V_i)\big] + C_6(X_i, V_i)\big[\widehat{B}(X_i, 0, V_i) - B(X_i, 0, V_i)\big] + o_p\left(\frac{1}{\sqrt{n}}\right)\Big\}.$$

Here

$$\mathcal{K}_1(X_i, V_i) = \frac{\pi(1, V_i)[1-\theta(X_i, V_i)] + \pi(0, V_i)\theta(X_i, V_i)}{\pi(1, V_i)\phi(X_i, V_i) + \pi(0, V_i)[1-\phi(X_i, V_i)]},$$

$$C_1(X_i, V_i) = \frac{\pi(0, V_i)[1-\theta(X_i, V_i) - \phi(X_i, V_i)]}{\{\pi(1, V_i)\phi(X_i, V_i) + \pi(0, V_i)[1-\phi(X_i, V_i)]\}^2} - \frac{[\pi(1, V_i) - \pi(0, V_i)]\theta(X_i, V_i)[1-\theta(X_i, V_i)]}{\pi(1, V_i)\{\pi(1, V_i)\phi(X_i, V_i) + \pi(0, V_i)[1-\phi(X_i, V_i)]\}} + \frac{[\pi(1, V_i) - \pi(0, V_i)]\phi(X_i, V_i)[1-\phi(X_i, V_i)]\mathcal{K}_1(X_i, V_i)}{\pi(1, V_i)\{\pi(1, V_i)\phi(X_i, V_i) + \pi(0, V_i)[1-\phi(X_i, V_i)]\}},$$

$$C_2(X_i, V_i) = -\frac{\pi(1, V_i)[1-\theta(X_i, V_i) - \phi(X_i, V_i)]}{\{\pi(1, V_i)\phi(X_i, V_i) + \pi(0, V_i)[1-\phi(X_i, V_i)]\}^2} + \frac{[\pi(1, V_i) - \pi(0, V_i)]\theta(X_i, V_i)[1-\theta(X_i, V_i)]}{\pi(0, V_i)\{\pi(1, V_i)\phi(X_i, V_i) + \pi(0, V_i)[1-\phi(X_i, V_i)]\}} - \frac{[\pi(1, V_i) - \pi(0, V_i)]\phi(X_i, V_i)[1-\phi(X_i, V_i)]\mathcal{K}_1(X_i, V_i)}{\pi(0, V_i)\{\pi(1, V_i)\phi(X_i, V_i) + \pi(0, V_i)[1-\phi(X_i, V_i)]\}},$$

$$C_3(X_i, V_i) = -\frac{[\pi(1, V_i) - \pi(0, V_i)]\phi(X_i, V_i)[1-\phi(X_i, V_i)]\mathcal{K}_1(X_i, V_i)}{A(X_i, 1, V_i)\{\pi(1, V_i)\phi(X_i, V_i) + \pi(0, V_i)[1-\phi(X_i, V_i)]\}},$$

$$C_4(X_i, V_i) = \frac{[\pi(1, V_i) - \pi(0, V_i)]\phi(X_i, V_i)[1-\phi(X_i, V_i)]\mathcal{K}_1(X_i, V_i)}{A(X_i, 0, V_i)\{\pi(1, V_i)\phi(X_i, V_i) + \pi(0, V_i)[1-\phi(X_i, V_i)]\}},$$

$$C_5(X_i, V_i) = \frac{[\pi(1, V_i) - \pi(0, V_i)]\theta(X_i, V_i)[1-\theta(X_i, V_i)]}{B(X_i, 1, V_i)\{\pi(1, V_i)\phi(X_i, V_i) + \pi(0, V_i)[1-\phi(X_i, V_i)]\}},$$

and

$$C_6(X_i, V_i) = -\frac{[\pi(1, V_i) - \pi(0, V_i)]\theta(X_i, V_i)[1-\theta(X_i, V_i)]}{B(X_i, 0, V_i)\{\pi(1, V_i)\phi(X_i, V_i) + \pi(0, V_i)[1-\phi(X_i, V_i)]\}}.$$

For simplicity, we express $\widehat{U}_{1n}(\boldsymbol{\beta}) - U_{1n}(\boldsymbol{\beta})$ as follows:

$$\widehat{U}_{1n}(\boldsymbol{\beta}) - U_{1n}(\boldsymbol{\beta}) = -[R_{1n}(\boldsymbol{\beta}) + R_{2n}(\boldsymbol{\beta}) + R_{3n}(\boldsymbol{\beta}) + R_{4n}(\boldsymbol{\beta}) + R_{5n}(\boldsymbol{\beta}) + R_{6n}(\boldsymbol{\beta})] + o_p(1).$$

First of all, the $R_{1n}(\boldsymbol{\beta})$ can be expressed as

$$\begin{aligned}
R_{1n}(\boldsymbol{\beta}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i\mathcal{X}_i H_{+}^{(1)}(X_i, V_i;\boldsymbol{\beta})\mathcal{K}_1^{-1}(X_i, V_i)C_1(X_i, V_i)[\widehat{\pi}(1, V_i)-\pi(1, V_i)]\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i\mathcal{X}_i H_{+}^{(1)}(X_i, V_i;\boldsymbol{\beta})\mathcal{K}_1^{-1}(X_i, V_i)C_1(X_i, V_i)\\
&\quad\times\left[\frac{n^{-1}\sum_{j=1}^{n}[\delta_j-\pi(1, V_i)]I\left(Y_j^0=1, V_j=V_i\right)}{P\left(Y^0=1, V=V_i\right)}+o_p\left(\frac{1}{\sqrt{n}}\right)\right]\\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\delta_i\mathcal{X}_i H_{+}^{(1)}(X_i, V_i;\boldsymbol{\beta})\mathcal{K}_1^{-1}(X_i, V_i)C_1(X_i, V_i)\right.\\
&\quad\left.\times\frac{I\left(V_i=V_j\right)[\delta_j-\pi(1, V_i)]I\left(Y_j^0=1\right)}{P\left(Y^0=1, V=V_i\right)}\right\}+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{[\delta_j-\pi\left(1, V_j\right)]I\left(Y_j^0=1\right)}{P\left(Y^0=1|V=V_j\right)}\\
&\quad\times E\left[\pi\left(Y^0, V\right)\mathcal{X}H_{+}^{(1)}(X, V;\boldsymbol{\beta})\mathcal{K}_1^{-1}(X, V)C_1(X, V)|V=V_j\right]+o_p(1).
\end{aligned}$$

Likewise, we can express $R_{2n}(\boldsymbol{\beta})$, $R_{3n}(\boldsymbol{\beta})$, $R_{4n}(\boldsymbol{\beta})$, $R_{5n}(\boldsymbol{\beta})$, and $R_{6n}(\boldsymbol{\beta})$ as follows:

$$\begin{aligned}
R_{2n}(\boldsymbol{\beta}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_1\mathcal{X}_i H_{+}^{(1)}(X_i, V_i;\boldsymbol{\beta})\mathcal{K}_1^{-1}(X_i, V_i)C_2(X_i, V_i)[\widehat{\pi}(0, V_i)-\pi(0, V_i)]\\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\left[\delta_j-\pi\left(0, V_j\right)\right]I\left(Y_j^0=0\right)}{P\left(Y^0=0|V=V_j\right)}\\
&\quad\times E\left[\pi\left(Y^0, V\right)\mathcal{X}H_{+}^{(1)}(X, V;\boldsymbol{\beta})\mathcal{K}_1^{-1}(X, V)C_2(X, V)|V=V_j\right]+o_p(1),\\
R_{3n}(\boldsymbol{\beta}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\delta_i\mathcal{X}_i H_{+}^{(1)}(X_i, V_i;\boldsymbol{\beta})\mathcal{K}_1^{-1}(X_i, V_i)C_3(X_i, V_i)\left[\widehat{A}(X_i, 1, V_i)-A(X_i, 1, V_i)\right]\\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\delta_i\mathcal{X}_j H_{+}^{(1)}\left(X_j, V_j;\boldsymbol{\beta}\right)\mathcal{K}_1^{-1}\left(X_j, V_j\right)C_3\left(X_j, V_j\right)\right.\\
&\quad\left.\times\frac{\left[I\left(Y_j^0=1\right)-A\left(X_j, 1, V_j\right)\right]\delta_j I\left(Y_j=0\right)I\left(X_i=X_j, V_i=V_j\right)}{P(\delta=1, Y=0, X=X_j, V=V_j)}\right\}+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\delta_j\mathcal{X}_j H_{+}^{(1)}(X_j, V_j;\boldsymbol{\beta})\mathcal{K}_1^{-1}\left(X_j, V_j\right)C_3\left(X_j, V_j\right)\left[I\left(Y_j^0=1\right)-A\left(X_j, 1, V_j\right)\right]I\left(Y_j=0\right)}{P\left(\delta=1, Y=0, X=X_j, V=V_j\right)}\\
&\quad\times\left[\frac{1}{n}\sum_{i=1}^{n}\delta_i I\left(X_i=X_j, V_i=V_j\right)\right]+o_p(1)
\end{aligned}$$

$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\delta_j \mathcal{X}_j H_{+}^{(1)}\left(X_j, V_j; \boldsymbol{\beta}\right) \mathcal{K}_1^{-1}\left(X_j, V_j\right) C_3\left(X_j, V_j\right)\left[I\left(Y_j^0=1\right)-A\left(X_j, 1, V_j\right)\right] I\left(Y_j=0\right)}{P\left(Y=0 \mid \delta=1, X=X_j, V=V_j\right) P\left(\delta=1, X=X_j, V=V_j\right)}$$
$$\times\left[P\left(\delta=1, X=X_j, V=V_j\right)+o_p(1)\right]+o_p(1)$$
$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\delta_j \mathcal{X}_j H_{+}^{(1)}\left(X_j, V_j; \boldsymbol{\beta}\right) \mathcal{K}_1^{-1}\left(X_j, V_j\right) C_3\left(X_j, V_j\right)\left[I\left(Y_j^0=1\right)-A\left(X_j, 1, V_j\right)\right] I\left(Y_j=0\right)}{P\left(Y=0 \mid \delta=1, X=X_j, V=V_j\right)}$$
$$+o_p(1),$$

$$R_{4n}(\boldsymbol{\beta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \mathcal{X}_i H_{+}^{(1)}\left(X_i, V_i; \boldsymbol{\beta}\right) \mathcal{K}_1^{-1}\left(X_i, V_i\right) C_4\left(X_i, V_i\right)\left[\widehat{A}\left(X_i, 0, V_i\right)-A\left(X_i, 0, V_i\right)\right]$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \mathcal{X}_i H_{+}^{(1)}\left(X_i, V_i; \boldsymbol{\beta}\right) \mathcal{K}_1^{-1}\left(X_i, V_i\right) C_4\left(X_i, V_i\right)$$
$$\times\left[\frac{n^{-1} \sum_{j=1}^{n}\left[I\left(Y_j^0=0\right)-A\left(X_i, 0, V_i\right)\right] \delta_j I\left(Y_j=0, X_j=X_i, V_j=V_i\right)}{P\left(\delta=1, Y=0, X=X_i, V=V_i\right)}\right.$$
$$\left.+o_p\left(\frac{1}{\sqrt{n}}\right)\right]$$
$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\delta_j \mathcal{X}_j H_{+}^{(1)}\left(X_j, V_j; \boldsymbol{\beta}\right) \mathcal{K}_1^{-1}\left(X_j, V_j\right) C_4\left(X_j, V_j\right)\left[I\left(Y_j^0=0\right)-A\left(X_j, 0, V_j\right)\right] I\left(Y_j=0\right)}{P\left(Y=0 \mid \delta=1, X=X_j, V=V_j\right)}$$
$$+o_p(1),$$

$$R_{5n}(\boldsymbol{\beta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \mathcal{X}_i H_{+}^{(1)}\left(X_i, V_i; \boldsymbol{\beta}\right) \mathcal{K}_1^{-1}\left(X_i, V_i\right) C_5\left(X_i, V_i\right)\left[\widehat{B}\left(X_i, 1, V_i\right)-B\left(X_i, 1, V_i\right)\right]$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \mathcal{X}_i H_{+}^{(1)}\left(X_i, V_i; \boldsymbol{\beta}\right) \mathcal{K}_1^{-1}\left(X_i, V_i\right) C_5\left(X_i, V_i\right)$$
$$\times\left[\frac{n^{-1} \sum_{j=1}^{n}\left[I\left(Y_j^0=1\right)-B\left(X_i, 1, V_i\right)\right] \delta_j I\left(Y_j=1, X_j=X_i, V_j=V_i\right)}{P\left(\delta=1, Y=1, X=X_i, V=V_i\right)}\right.$$
$$\left.+o_p\left(\frac{1}{\sqrt{n}}\right)\right]$$
$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\delta_j \mathcal{X}_j H_{+}^{(1)}\left(X_j, V_j; \boldsymbol{\beta}\right) \mathcal{K}_1^{-1}\left(X_j, V_j\right) C_5\left(X_j, V_j\right)\left[I\left(Y_j^0=1\right)-B\left(X_j, 1, V_j\right)\right] I\left(Y_j=1\right)}{P\left(Y=1 \mid \delta=1, X=X_j, V=V_j\right) P\left(\delta=1, X=X_j, V=V_j\right)}$$
$$\times\left[P\left(\delta=1, X=X_j, V=V_j\right)+o_p(1)\right]+o_p(1)$$
$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\delta_j \mathcal{X}_j H_{+}^{(1)}\left(X_j, V_j; \boldsymbol{\beta}\right) \mathcal{K}_1^{-1}\left(X_j, V_j\right) C_5\left(X_j, V_j\right)\left[I\left(Y_j^0=1\right)-B\left(X_j, 1, V_j\right)\right] I\left(Y_j=1\right)}{P(Y=1 \mid \delta=1, X=X_j, V=V_j)}$$
$$+o_p(1),$$

and

$$R_{6n}(\boldsymbol{\beta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \mathcal{X}_i H_{+}^{(1)}\left(X_i, V_i; \boldsymbol{\beta}\right) \mathcal{K}_1^{-1}\left(X_i, V_i\right) C_6\left(X_i, V_i\right)\left[\widehat{B}\left(X_i, 0, V_i\right)-B\left(X_i, 0, V_i\right)\right]$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i \mathcal{X}_i H_{+}^{(1)}\left(X_i, V_i; \boldsymbol{\beta}\right) \mathcal{K}_1^{-1}\left(X_i, V_i\right) C_6\left(X_i, V_i\right)$$

$$\times \left[ \frac{n^{-1}\sum_{j=1}^{n}\left[I\left(Y_j^0=0\right)-B\left(X_i,0,V_i\right)\right]\delta_j I\left(Y_j=1,X_j=X_i,V_j=V_i\right)}{P\left(\delta=1,Y=1,X=X_i,V=V_i\right)} + o_p\left(\frac{1}{\sqrt{n}}\right)\right]$$

$$= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\delta_j \mathcal{X}_j H_+^{(1)}\left(X_j,V_j;\boldsymbol{\beta}\right)\mathcal{K}_1^{-1}\left(X_j,V_j\right)C_6\left(X_j,V_j\right)\left[I\left(Y_j^0=0\right)-B\left(X_j,0,V_j\right)\right]I\left(Y_j=1\right)}{P\left(Y=1|\delta=1,X=X_j,V=V_j\right)}$$
$$+o_p(1),$$

It can then be shown that

$$-\left[R_{1n}(\boldsymbol{\beta})+R_{2n}(\boldsymbol{\beta})+R_{3n}(\boldsymbol{\beta})+R_{4n}(\boldsymbol{\beta})+R_{5n}(\boldsymbol{\beta})+R_{6n}(\boldsymbol{\beta})\right]$$
$$=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\varepsilon_c(Y_j,Y_j^0,X_j,V_j;\boldsymbol{\beta})+o_p(1)$$

and, hence,

$$\widehat{U}_{1n}(\boldsymbol{\beta})=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[S_c\left(Y_i,Y_i^0,X_i,V_i\right)+\varepsilon_c\left(Y_i,Y_i^0,X_i,V_i;\boldsymbol{\beta}\right)\right]+o_p(1),$$

where $S_c\left(Y_i,Y_i^0,X_i,V_i;\boldsymbol{\beta}\right)=\delta_i\mathcal{X}_i[Y_i-H_+\left(X_i,V_i;\boldsymbol{\beta}\right)]$ and

$$\varepsilon_c\left(Y_i,Y_i^0,X_i,V_i;\boldsymbol{\beta}\right)=-\delta_i\mathcal{X}_i H_+^{(1)}\left(X_i,V_i;\boldsymbol{\beta}\right)\mathcal{K}_1^{-1}\left(X_i,V_i\right)$$
$$\times\left\{\frac{C_3\left(X_i,V_i\right)\left[I\left(Y_i^0=1\right)-A\left(X_i,1,V_i\right)\right]I\left(Y_i=0\right)}{P\left(Y=0|\delta=1,X=X_i,V=V_i\right)}\right.$$
$$+\frac{C_4\left(X_i,V_i\right)\left[I\left(Y_i=0\right)-A\left(X_i,0,V_i\right)\right]I\left(Y_i=0\right)}{P(Y=0|\delta=1,X=X_i,V=V_i)}$$
$$+\frac{C_5\left(X_i,V_i\right)\left[I\left(Y_i^0=1\right)-B\left(X_i,1,V_i\right)\right]I\left(Y_i=1\right)}{P\left(Y=1|\delta=1,X=X_i,V=V_i\right)}$$
$$\left.+\frac{C_6\left(X_i,V_i\right)\left[I\left(Y_i^0=0\right)-B\left(X_i,0,V_i\right)\right]I\left(Y_i=1\right)}{P\left(Y=1|\delta=1,X=X_i,V=V_i\right)}\right\}$$
$$+\frac{E\left[\pi\left(Y,V\right)\mathcal{X}H_+^{(1)}(X,V;\boldsymbol{\beta})\mathcal{K}_1^{-1}(X,V)C_1(X,V)|V=V_i\right]}{P\left(Y^0=1|V=V_i\right)}$$
$$\times[\delta_i-\pi(1,V_i)]I\left(Y_i^0=1\right)$$
$$+\frac{E\left[\pi(Y,V)\mathcal{X}H_+^{(1)}(X,V;\boldsymbol{\beta})\mathcal{K}_1^{-1}(X,V)C_2(X,V)|V=V_i\right]}{P\left(Y^0=0|V=V_i\right)}$$
$$\times[\delta_i-\pi(0,V_i)]I\left(Y_i^0=0\right).$$

The proof is completed. □

*Proof of Lemma 2* We first express $\widehat{U}_{2n}(\boldsymbol{\beta})$ as follows:

$$\begin{aligned}
\widehat{U}_{2n}(\boldsymbol{\beta}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})[Y_i^0 - \widehat{H}_-(V_i;\boldsymbol{\beta})] \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})[Y_i^0 - H_-(V_i;\boldsymbol{\beta})] \\
&\quad -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})[\widehat{H}_-(V_i;\boldsymbol{\beta}) - H_-(V_i;\boldsymbol{\beta})] \\
&\quad +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)[\widehat{\mathcal{T}}_i(\boldsymbol{\beta}) - \mathcal{T}_i(\boldsymbol{\beta})][Y_i^0 - H_-(V_i;\boldsymbol{\beta})] \\
&\quad -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)[\widehat{\mathcal{T}}_i(\boldsymbol{\beta}) - \mathcal{T}_i(\boldsymbol{\beta})][\widehat{H}_-(V_i;\boldsymbol{\beta}) - H_-(V_i;\boldsymbol{\beta})] \\
&= U_{2n}(\boldsymbol{\beta}) + Q_{1n}(\boldsymbol{\beta}) + Q_{2n}(\boldsymbol{\beta}) + Q_{3n}(\boldsymbol{\beta}).
\end{aligned}$$

Here

$$\begin{aligned}
Q_{1n}(\boldsymbol{\beta}) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})[\widehat{H}_-(V_i;\boldsymbol{\beta}) - H_-(V_i;\boldsymbol{\beta})]. \\
Q_{2n}(\boldsymbol{\beta}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)[\widehat{\mathcal{T}}_i(\boldsymbol{\beta}) - \mathcal{T}_i(\boldsymbol{\beta})][Y_i^0 - H_-(V_i;\boldsymbol{\beta})]. \\
Q_{3n}(\boldsymbol{\beta}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)[\widehat{\mathcal{T}}_i(\boldsymbol{\beta}) - \mathcal{T}_i(\boldsymbol{\beta})][\widehat{H}_-(V_i;\boldsymbol{\beta}) - H_-(V_i;\boldsymbol{\beta})].
\end{aligned}$$

Because $V_i$ has a fine support, we can show $\widehat{\mathcal{T}}_i(\boldsymbol{\beta}) - \mathcal{T}_i(\boldsymbol{\beta}) = o_p(1)$. $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)[Y_i^0 - H_-(V_i;\boldsymbol{\beta})] = O_p(1)$, so, it can be shown that $Q_{mn}(\boldsymbol{\beta}) = o_p(1)$, $m = 2, 3$. By a Taylor expansion of $Q_{1n}(\boldsymbol{\beta})$ at $\pi(0, V_i)$, $\pi(1, V_i)$, $A_0(0, V_i)$, $A_0(1, V_i)$, $B_0(0, V_i)$, $B_0(1, V_i)$, $r_0(0, V_i;\beta_1)$, and $r_0(1, V_i;\beta_1)$, one can show that

$$\begin{aligned}
Q_{1n}(\boldsymbol{\beta}) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})[\widehat{H}_-(V_i;\boldsymbol{\beta}) - H_-(V_i;\boldsymbol{\beta})] \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_-^{(1)}(V_i;\boldsymbol{\beta})D_{1i}(\boldsymbol{\beta})[\widehat{\pi}(0, V_i) - \pi(0, V_i)] \\
&\quad -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_-^{(1)}(V_i;\boldsymbol{\beta})D_{2i}(\boldsymbol{\beta})[\widehat{\pi}(1, V_i) - \pi(1, V_i)]
\end{aligned}$$

$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\,\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})D_{3i}(\boldsymbol{\beta})[\widehat{A}_0\,(0,V_i)-A_0(0,V_i)]$$
$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})D_{4i}(\boldsymbol{\beta})[\widehat{A}_0(1,V_i)-A_0(1,V_i)]$$
$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})D_{5i}(\boldsymbol{\beta})[\widehat{B}_0(0,V_i)-B_0(0,V_i)]$$
$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})D_{6i}(\boldsymbol{\beta})[\widehat{B}_0(1,V_i)-B_0(1,V_i)]$$
$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})\ D_{7i}(\boldsymbol{\beta})[\widehat{r}_0(0,V_i;\beta_1)-r_0(0,V_i,\beta_1)]$$
$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})D_{8i}(\boldsymbol{\beta})[\widehat{r}_0(1,V_i;\beta_1)-r_0\,(1,V_i;\beta_1)]+o_p(1).$$

Here

$$\begin{aligned}
D_{1i}(\boldsymbol{\beta}) &= \frac{1}{1-\pi\,(0,V_i)}+\left[\frac{H_1\,(V_i;\boldsymbol{\beta})}{1-\theta_0(V_i)}+\frac{H_2\,(V_i;\boldsymbol{\beta})}{\theta_0\,(V_i)}\right]\frac{\theta_0(V_i)[1-\theta_0(V_i)]}{\pi(0,V_i)}\\
&\quad+\frac{H_1(V_i;\boldsymbol{\beta})-H_2(V_i;\boldsymbol{\beta})}{r(V_i;\beta_1)}\frac{[r_0(1,V_i;\beta_1)-r_0(0,V_i;\beta_1)]\phi_0(V_i)[1-\phi_0(V_i)]}{\pi(0,V_i)}\\
&\quad+\left[\frac{H_1(V_i;\boldsymbol{\beta})}{1-\theta_0(V_i)}+\frac{H_2(V_i;\boldsymbol{\beta})}{\theta_0(V_i)}\right]\frac{\phi_0(V_i)[1-\phi_0(V_i)]}{\pi(0,V_i)}e^{-[\beta_0+\beta_2^{\mathrm{T}}Z_i+R(V_i;\beta_1)]},\\
D_{2i}(\boldsymbol{\beta}) &= \frac{-1}{1-\pi(1,V_i)}-\left[\frac{H_1(V_i;\boldsymbol{\beta})}{1-\theta_0(V_i)}+\frac{H_2(V_i;\boldsymbol{\beta})}{\theta_0(V_i)}\right]\frac{\theta_0(V_i)[1-\theta_0(V_i)]}{\pi(1,V_i)}\\
&\quad-\frac{H_1(V_i;\boldsymbol{\beta})-H_2(V_i;\boldsymbol{\beta})}{r(V_i;\boldsymbol{\beta}_1)}\frac{[r_0(1,V_i;\beta_1)-r_0(0,V_i;\beta_1)]\phi_0(V_i)[1-\phi_0(V_i)]}{\pi(1,V_i)}\\
&\quad-\left[\frac{H_1(V_i;\boldsymbol{\beta})}{1-\theta_0(V_i)}+\frac{H_2(V_i;\boldsymbol{\beta})}{\theta_0(V_i)}\right]\frac{\phi_0(V_i)[1-\phi_0(V_i)]}{\pi(1,V_i)}e^{-[\beta_0+\beta_2^{\mathrm{T}}Z_i+R(V_i;\beta_1)]},\\
D_{3i}(\boldsymbol{\beta}) &= -\frac{H_1(V_i;\boldsymbol{\beta})-H_2(V_i;\boldsymbol{\beta})}{r(V_i;\beta_1)}\frac{[r_0(1,V_i;\beta_1)-r_0(0,V_i;\beta_1)]\phi_0(V_i)[1-\phi_0(V_i)]}{A_0(0,V_i)}\\
&\quad-\left[\frac{H_1(V_i;\boldsymbol{\beta})}{1-\theta_0(V_i)}+\frac{H_2(V_i;\boldsymbol{\beta})}{\theta_0(V_i)}\right]\frac{\phi_0(V_i)[1-\phi_0(V_i)]}{A_0(0,V_i)}e^{-[\beta_0+\beta_2^{\mathrm{T}}Z_i+R(V_i;\beta_1)]},\\
D_{4i}(\boldsymbol{\beta}) &= \frac{H_1(V_i;\boldsymbol{\beta})-H_2(V_i;\boldsymbol{\beta})}{r(V_i;\beta_1)}\frac{[r_0(1,V_i;\boldsymbol{\beta}_1)-r_0(0,V_i;\beta_1)]\phi_0(V_i)[1-\phi_0(V_i)]}{A_0(1,V_i)}\\
&\quad+\left[\frac{H_1(V_i;\boldsymbol{\beta})}{1-\theta_0(V_i)}+\frac{H_2(V_i;\boldsymbol{\beta})}{\theta_0(V_i)}\right]\frac{\phi_0(V_i)[1-\phi_0(V_i)]}{A_0(1,V_i)}e^{-[\beta_0+\beta_2^{\mathrm{T}}Z_i+R(V_i;\beta_1)]},\\
D_{5i}(\boldsymbol{\beta}) &= -\left[\frac{H_1(V_i;\boldsymbol{\beta})}{1-\theta_0(V_i)}+\frac{H_2(V_i;\boldsymbol{\beta})}{\theta_0(V_i)}\right]\frac{\theta_0(V_i)[1-\theta_0(V_i)]}{B_0(0,V_i)},\\
D_{6i}(\boldsymbol{\beta}) &= \left[\frac{H_1(V_i;\boldsymbol{\beta})}{1-\theta_0(V_i)}+\frac{H_2(V_i;\boldsymbol{\beta})}{\theta_0(V_i)}\right]\frac{\theta_0(V_i)[1-\theta_0(V_i)]}{B_0(1,V_i)},
\end{aligned}$$

$$D_{7i}(\boldsymbol{\beta}) = \frac{[H_1(V_i;\boldsymbol{\beta}) - H_2(V_i;\boldsymbol{\beta})][1-\phi_0(V_i)]}{r(V_i;\beta_1)},$$

and

$$D_{8i}(\boldsymbol{\beta}) = \frac{[H_1(V_i;\boldsymbol{\beta}) - H_2(V_i;\boldsymbol{\beta})]\phi_0(V_i)}{r(V_i;\beta_1)}$$

with

$$H_1(V_i;\boldsymbol{\beta}) = H\left[\beta_0 + \beta_2^{\mathrm{T}} Z_i + R(V_i;\beta_1) + \ln\frac{1-\theta_0(V_i)}{\phi_0(V_i)}\right],$$

and

$$H_2(V_i;\boldsymbol{\beta}) = H\left[\beta_0 + \beta_2^{\mathrm{T}} Z_i + R(V_i;\beta_1) + \ln\frac{\theta_0(V_i)}{1-\phi_0(V_i)}\right].$$

For simplicity, we express $Q_{1n}(\boldsymbol{\beta})$ as follows:

$$\begin{aligned} Q_{1n}(\boldsymbol{\beta}) &= G_{1n}(\boldsymbol{\beta}) + G_{2n}(\boldsymbol{\beta}) + G_{3n}(\boldsymbol{\beta}) + G_{4n}(\boldsymbol{\beta}) \\ &\quad + G_{5n}(\boldsymbol{\beta}) + G_{6n}(\boldsymbol{\beta}) + G_{7n}(\boldsymbol{\beta}) + G_{8n}(\boldsymbol{\beta}) + o_p(1). \end{aligned}$$

With some algebra, we can express $G_{tn}(\boldsymbol{\beta})$'s, $t = 1, 2, \ldots, 8$, as follows:

$$\begin{aligned} G_{1n}(\boldsymbol{\beta}) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_-^{(1)}(V_i;\boldsymbol{\beta})D_{1i}(\boldsymbol{\beta})[\widehat{\pi}(0,V_i) - \pi(0,V_i)] \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_-^{(1)}(V_i;\boldsymbol{\beta})D_{1i}(\boldsymbol{\beta}) \\ &\quad \times\left\{\frac{n^{-1}\sum_{k=1}^{n}[\delta_k - \pi(0,V_i)]I(Y_k^0 = 0, V_k = V_i)}{P(Y^0 = 0, V = V_i)} + o_p\left(\frac{1}{\sqrt{n}}\right)\right\} \\ &= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left[\frac{I(Y_k^0 = 0)\mathcal{T}_k(\boldsymbol{\beta})H_-^{(1)}(V_k;\boldsymbol{\beta})D_{1k}(\boldsymbol{\beta})[\delta_k - \pi(0,V_k)]}{P(Y^0 = 0|V = V_k)}\right] \\ &\quad \times\frac{\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)I(V_i = V_k)}{P(V = V_k)} + o_p(1) \\ &= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left[\frac{I(Y_k^0 = 0)\mathcal{T}_k(\boldsymbol{\beta})H_-^{(1)}(V_k;\boldsymbol{\beta})D_{1k}(\boldsymbol{\beta})[\delta_k - \pi(0,V_k)]}{P(Y^0 = 0|V = V_k)}\right] \\ &\quad \times E[1 - \pi(Y^0, V)|V = V_k] + o_p(1) \\ &= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathcal{T}_k(\boldsymbol{\beta})H_-^{(1)}(V_k;\boldsymbol{\beta})E[1 - \pi(Y^0, V)|V = V_k]D_{1k}(\boldsymbol{\beta}) \\ &\quad \times\frac{I(Y_k^0 = 0)[\delta_k - \pi(0,V_k)]}{P(Y^0 = 0|V = V_k)} + o_p(1), \end{aligned}$$

$$
\begin{aligned}
G_{2n}(\boldsymbol{\beta}) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})D_{2i}(\boldsymbol{\beta})[\widehat{\pi}(1,V_i)-\pi(1,V_i)] \\
&= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\mathcal{T}_k(\boldsymbol{\beta})H_{-}^{(1)}(V_k;\boldsymbol{\beta})D_{2k}(\boldsymbol{\beta}) \\
&\quad\times\frac{I(Y_k^0=1)[\delta_k-\pi(1,V_k)]E[1-\pi(Y^0,V)|V=V_k]}{P(Y^0=1|V=V_k)}+o_p(1), \\
G_{3n}(\boldsymbol{\beta}) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})D_{3i}(\boldsymbol{\beta})[\widehat{A}_0(0,V_i)-A_0(0,V_i)] \\
&= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{(1-\delta_i)\mathcal{T}_k(\boldsymbol{\beta})H_{-}^{(1)}(V_k;\boldsymbol{\beta})D_{3k}(\boldsymbol{\beta})I(V_i=V_k)}{P(V=V_k)}\right. \\
&\quad\left.\times\left[\delta_k I(Y_k^0=0,Y_k=0)-A_0(0,V_k)\right]\right\}+\ o_p(1) \\
&= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left\{\mathcal{T}_k(\boldsymbol{\beta})H_{-}^{(1)}(V_k;\boldsymbol{\beta})D_{3k}(\boldsymbol{\beta})\left[\delta_k I(Y_k^0=0,Y_k=0)-A_0(0,V_k)\right]\right. \\
&\quad\left.\times\ \frac{\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)I(V_i=V_k)}{P(V=V_k)}\right\}+\ o_p(1) \\
&= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left\{\mathcal{T}_k(\boldsymbol{\beta})H_{-}^{(1)}(V_k;\boldsymbol{\beta})D_{3k}(\boldsymbol{\beta})\left[\delta_k I(Y_k^0=0,Y_k=0)-A_0(0,V_k)\right]\right. \\
&\quad\left.\times\ E[1-\pi(Y^0,V)|V=V_k)]\right\}+o_p(1), \\
G_{4n}(\boldsymbol{\beta}) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})D_{4i}(\boldsymbol{\beta})[\widehat{A}_0(1,V_i)-A_0(1,V_i)] \\
&= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left\{\mathcal{T}_k(\boldsymbol{\beta})H_{-}^{(1)}(V_k;\boldsymbol{\beta})D_{4k}(\boldsymbol{\beta})[\delta_k I(Y_k^0=1,Y_k=0)-A_0(1,V_k)]\right. \\
&\quad\left.\times\ E[1-\pi(Y^0,V)|V=V_k]\right\}+o_p(1), \\
G_{5n}(\boldsymbol{\beta}) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})D_{5i}(\boldsymbol{\beta})[\widehat{B}_0(0,V_i)-B_0(0,V_i)] \\
&= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{(1-\delta_i)\mathcal{T}_k(\boldsymbol{\beta})H_{-}^{(1)}(V_k;\boldsymbol{\beta})D_{5k}(\boldsymbol{\beta})I(V_i=V_k)}{P(V=V_k)}\right. \\
&\quad\left.\times\left[\delta_k I(Y_k^0=0,Y_k=1)-B_0(0,V_k)\right]\right\}+\ o_p(1) \\
&= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left\{\mathcal{T}_k(\boldsymbol{\beta})H_{-}^{(1)}(V_k;\boldsymbol{\beta})D_{5k}(\boldsymbol{\beta})\left[\delta_k I(Y_k^0=0,Y_k=1)-B_0(0,V_k)\right]\right.
\end{aligned}
$$

$$\times \left. \frac{\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)I(V_i=V_k)}{P(V=V_k)} \right\} + o_p(1)$$

$$= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left\{\mathcal{T}_k(\boldsymbol{\beta})H_{-}^{(1)}(V_k;\boldsymbol{\beta})D_{5k}(\boldsymbol{\beta})\left[\delta_k I(Y_k^0=0, Y_k=1)-B_0(0,V_k)\right]\right.$$

$$\left.\times E[1-\pi(Y^0,V)|V=V_k)]\right\} + o_p(1),$$

$$G_{6n}(\boldsymbol{\beta}) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})D_{6i}(\boldsymbol{\beta})[\widehat{B}_0(1,V_i)-B_0(1,V_i)]$$

$$= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left\{\mathcal{T}_k(\boldsymbol{\beta})H_{-}^{(1)}(V_k;\boldsymbol{\beta})D_{6k}(\boldsymbol{\beta})\left[\delta_k I(Y_k^0=1, Y_k=1)-B_0(1,V_k)\right]\right.$$

$$\left.\times E[1-\pi(Y^0,V)|V=V_k)]\right\} + o_p(1),$$

$$G_{7n}(\boldsymbol{\beta}) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})D_{7i}(\boldsymbol{\beta})[\widehat{r}_0(0,V_i;\boldsymbol{\beta}_1)-r_0(0,V_i;\boldsymbol{\beta}_1)]$$

$$= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})D_{7i}(\boldsymbol{\beta})$$

$$\times\left\{\frac{n^{-1}\sum_{k=1}^{n}\delta_k[e^{\boldsymbol{\beta}_1^{\mathrm{T}}X_k}-r_0(0,V_i;\boldsymbol{\beta}_1)]I(Y_k=0,Y_k^0=0,V_k=V_i)}{\pi(0,V_i)P(Y=0,Y^0=0,V=V_i)}\right.$$

$$\left.+o_p\left(\frac{1}{\sqrt{n}}\right)\right\}$$

$$= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})D_{7i}(\boldsymbol{\beta})\delta_k[e^{\boldsymbol{\beta}_1^{\mathrm{T}}X_k}-r_0(0,V_i;\boldsymbol{\beta}_1)]}{\pi(0,V_i)P(Y=0,Y^0=0,V=V_i)}\right.$$

$$\left.\times I(Y_k=0,Y_k^0=0,V_k=V_i)\right\} + o_p(1)$$

$$= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left\{\frac{\mathcal{T}_k(\boldsymbol{\beta})H_{-}^{(1)}(V_k;\boldsymbol{\beta})D_{7k}(\boldsymbol{\beta})I(Y_k=0,Y_k^0=0)\delta_k[e^{\boldsymbol{\beta}_1^{\mathrm{T}}X_k}-r_0(0,V_k;\boldsymbol{\beta}_1)]}{\pi(0,V_k)P(Y=0,Y^0=0|V=V_k)}\right.$$

$$\left.\times\frac{\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)I(V_i=V_k)}{P(V=V_k)}\right\} + o_p(1)$$

$$= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left\{\mathcal{T}_k(\boldsymbol{\beta})H_{-}^{(1)}(V_k;\boldsymbol{\beta})D_{7k}(\boldsymbol{\beta})\frac{I(Y_k=0,Y_k^0=0)\delta_k[e^{\boldsymbol{\beta}_1^{\mathrm{T}}X_k}-r_0(0,V_k;\boldsymbol{\beta}_1)]}{\pi(0,V_k)P(Y=0,Y^0=0|V=V_k)}\right.$$

$$\left.\times E[1-\pi(Y^0,V)|V=V_k]\right\} + o_p(1),$$

and

$$G_{8n}(\boldsymbol{\beta}) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})D_{8i}(\boldsymbol{\beta})[\widehat{r}_0(1,V_i;\boldsymbol{\beta}_1)-r_0(1,V_i;\boldsymbol{\beta}_1)]$$
$$= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\Bigg\{\mathcal{T}_k(\boldsymbol{\beta})H_{-}^{(1)}(V_k;\boldsymbol{\beta})D_{8k}(\boldsymbol{\beta})\frac{I(Y_k=0,Y_k^0=1)\delta_k[e^{\boldsymbol{\beta}_1^{\mathrm{T}}X_k}-r_0(1,V_k;\boldsymbol{\beta}_1)]}{\pi(1,V_k)P(Y=0,Y^0=1|V=V_k)}$$
$$\times E[1-\pi(Y^0,V)|V=V_k]\Bigg\}+o_p(1).$$

We can then show that

$$Q_{1n}(\boldsymbol{\beta}) = \sum_{k=1}^{8}G_{kn}(\boldsymbol{\beta})+o_p(1) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varepsilon_m(Y_i,Y_i^0,X_i,V_i;\boldsymbol{\beta})+o_p(1)$$

and, hence,

$$\widehat{U}_{2n}(\boldsymbol{\beta}) = U_{2n}(\boldsymbol{\beta})+Q_{1n}(\boldsymbol{\beta})+o_p(1)$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[S_m(Y_i^0,V_i;\boldsymbol{\beta})+\varepsilon_m(Y_i,Y_i^0,X_i,V_i;\boldsymbol{\beta})\right]+o_p(1),$$

where $S_m(Y_i^0,V_i;\boldsymbol{\beta}) = (1-\delta_i)\mathcal{T}_i(\boldsymbol{\beta})[Y_i^0-H_{-}(V_i;\boldsymbol{\beta})]$, and

$$\varepsilon_m(Y_i,Y_i^0,X_i,V_i;\boldsymbol{\beta}) = -\mathcal{T}_i(\boldsymbol{\beta})H_{-}^{(1)}(V_i;\boldsymbol{\beta})P(\delta=0|V=V_i)$$
$$\times\Bigg\{D_{1i}(\boldsymbol{\beta})\frac{I(Y_i^0=0)\,[\delta_i-\pi(0,V_i)]}{P(Y^0=0|V=V_i)}$$
$$+D_{2i}(\boldsymbol{\beta})\frac{I(Y_i^0=1)\,[\delta_i-\pi(1,V_i)]}{P(Y^0=1|V=V_i)}$$
$$+D_{3i}(\boldsymbol{\beta})\left[\delta_i I(Y_i^0=0,Y_i=0)-A_0(0,V_i)\right]$$
$$+D_{4i}(\boldsymbol{\beta})\left[\delta_i I(Y_i^0=1,Y_i=0)-A_0(1,V_i)\right]$$
$$+D_{5i}(\boldsymbol{\beta})\left[\delta_i I(Y_i^0=0,Y_i=1)-B_0(0,V_i)\right]$$
$$+D_{6i}(\boldsymbol{\beta})\left[\delta_i I(Y_i^0=1,Y_i=1)-B_0(1,V_i)\right]$$
$$+D_{7i}(\boldsymbol{\beta})\frac{I(Y_i^0=0,Y_i=0)\delta_i\left[e^{\beta_1^{\mathrm{T}}X_i}-r_0(0,V_i;\beta_1)\right]}{\pi(0,V_i)P(Y=0,Y^0=0|V=V_i)}$$
$$+D_{8i}(\boldsymbol{\beta})\frac{I(Y_i^0=1,Y_i=0)\delta_i\left[e^{\beta_1^{\mathrm{T}}X_i}-r_0(1,V_i;\beta_1)\right]}{\pi(1,V_i)P(Y=0,Y^0=1|V=V_i)}\Bigg\}.$$

The proof is completed. □

*Proof of Theorem 1* To show that $\widehat{\boldsymbol{\beta}}_J$ is a consistent estimator of $\boldsymbol{\beta}$, we consider

$$
\begin{aligned}
G_n(\boldsymbol{\beta}) &= \frac{1}{\sqrt{n}}\left[-\frac{\partial \widehat{U}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{\mathrm{T}}}\right] \\
&= \frac{1}{n}\sum_{j=1}^{n}\left[\delta_j \mathcal{X}_j \mathcal{X}_j^{\mathrm{T}} \widehat{H}_+^{(1)}(X_j, V_j; \boldsymbol{\beta}) + (1-\delta_j)\widehat{\mathcal{T}}_j(\boldsymbol{\beta})\widehat{\mathcal{T}}_j^{\mathrm{T}}(\boldsymbol{\beta})\widehat{H}_-^{(1)}(V_j; \boldsymbol{\beta})\right] \\
&\quad + \frac{1}{n}\sum_{j=1}^{n}(1-\delta_j)[Y_j^0 - \widehat{H}_-(V_j; \boldsymbol{\beta})]\frac{\partial}{\partial \boldsymbol{\beta}^{\mathrm{T}}}\widehat{\mathcal{T}}_j(\boldsymbol{\beta}) \\
&= \frac{1}{n}\sum_{j=1}^{n}\left[\delta_j \mathcal{X}_j \mathcal{X}_j^{\mathrm{T}} \widehat{H}_+^{(1)}(X_j, V_j; \boldsymbol{\beta}) + (1-\delta_j)\widehat{\mathcal{T}}_j(\boldsymbol{\beta})\widehat{\mathcal{T}}_j^{\mathrm{T}}(\boldsymbol{\beta})\widehat{H}_-^{(1)}(V_j; \boldsymbol{\beta})\right] \\
&\quad - \frac{1}{n}\sum_{j=1}^{n}(1-\delta_j)[Y_j^0 - \widehat{H}_-(V_j; \boldsymbol{\beta})][\widehat{H}_1^{(1)}(V_j; \boldsymbol{\beta}) - \widehat{H}_2^{(1)}(V_j; \boldsymbol{\beta})] \\
&\quad \times \begin{pmatrix} 1 & \widehat{R}_{\beta_1}^{\mathrm{T}}(V_j) & Z_j^{\mathrm{T}} \\ \widehat{R}_{\beta_1}(V_j) & \widehat{R}_{\beta_1}(V_j)\widehat{R}_{\beta_1}^{\mathrm{T}}(V_j) & \widehat{R}_{\beta_1}(V_j)Z_j^{\mathrm{T}} \\ Z_j & Z_j\widehat{R}_{\beta_1}^{\mathrm{T}}(V_j) & Z_j Z_j^{\mathrm{T}} \end{pmatrix} - \frac{1}{n}\sum_{j=1}^{n}(1-\delta_j) \\
&\quad \times [Y_j^0 - \widehat{H}_-(V_j; \boldsymbol{\beta})]\,[\widehat{H}_1(V_j; \boldsymbol{\beta}) - \widehat{H}_2(V_j; \boldsymbol{\beta})]\begin{pmatrix} 0 & \mathbf{0}^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & \widehat{R}_{\beta_1}^{(2)}(V_j) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix},
\end{aligned}
$$

where

$$
\begin{aligned}
\widehat{r}_{\beta_1}^{(2)}(Y_i^0, V_i; \beta_1) &= \frac{\partial}{\partial \beta_1}\widehat{r}_{\beta_1}(Y_i^0, V_i; \beta_1) \\
&= \frac{\sum_{k=1}^{n}\delta_k X_k X_k^{\mathrm{T}} e^{\beta_1^{\mathrm{T}} X_k} I(Y_k = 0, Y_k^0 = Y_i^0, V_k = V_i)}{\sum_{k=1}^{n}\delta_k I(Y_k = 0, Y_k^0 = y^0, V_k = V_i)}
\end{aligned}
$$

and

$$
\begin{aligned}
\widehat{R}_{\beta_1}^{(2)}(V_i) &= \frac{\partial^2}{\partial \beta_1^{\mathrm{T}} \partial \beta_1}\widehat{R}(V_i; \beta_1) \\
&= \frac{\widehat{r}_{\beta_1}^{(2)}(0, V_i; \beta_1)[1 - \widehat{\phi}_0(V_i)] + \widehat{r}_{\beta_1}^{(2)}(1, V_i; \beta_1)\widehat{\phi}_0(V_i)}{\widehat{r}_0(0, V_i; \beta_1)[1 - \widehat{\phi}_0(V_i)] + \widehat{r}_0(1, V_i; \beta_1)\widehat{\phi}_0(V_i)} - [\widehat{R}_{\beta_1}(V_i)]^{\otimes 2}.
\end{aligned}
$$

It can then be shown that $G_n(\boldsymbol{\beta}) \xrightarrow{p} G(\boldsymbol{\beta})$. By Condition (A5), the convergence of $G_n(\boldsymbol{\beta})$ to $G(\boldsymbol{\beta})$ is uniform in a neighborhood of the true $\boldsymbol{\beta}$. By the Inverse Function Theorem of Foutz (1977), along with Condition (A4), one can show that a unique consistent solution exists for the estimating equation $\widehat{U}_n(\boldsymbol{\beta}) = 0$ in a neighborhood of the true $\boldsymbol{\beta}$. Consequently it follows that $\widehat{\boldsymbol{\beta}}_J$ is a consistent estimator of $\boldsymbol{\beta}$.

Next, we derive the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_J - \boldsymbol{\beta})$. By a Taylor expansion of $\widehat{U}_n(\boldsymbol{\beta})$ at $\boldsymbol{\beta}$, we can have

$$\begin{aligned} 0 &= \widehat{U}_n(\widehat{\boldsymbol{\beta}}_J) \\ &= \widehat{U}_n(\boldsymbol{\beta}) + \frac{\partial \widehat{U}_n(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T}(\widehat{\boldsymbol{\beta}}_J - \boldsymbol{\beta}) + o_p(1) \\ &= \widehat{U}_n(\boldsymbol{\beta}) - G_n(\boldsymbol{\beta})\sqrt{n}(\widehat{\boldsymbol{\beta}}_J - \boldsymbol{\beta}) + o_p(1). \end{aligned}$$

Because $G_n(\boldsymbol{\beta}) \xrightarrow{p} G(\boldsymbol{\beta})$, it can be shown that $\sqrt{n}(\widehat{\boldsymbol{\beta}}_J - \boldsymbol{\beta}) = G^{-1}(\boldsymbol{\beta})\widehat{U}_n(\boldsymbol{\beta}) + o_p(1)$. By Lemmas 1 and 2, one can have $\mathrm{Cov}\{\widehat{U}_n(\boldsymbol{\beta})\} = M(\boldsymbol{\beta})$ as defined in Theorem 1. It is known that $S_c(Y_i, Y_i^0, X_i, V_i; \boldsymbol{\beta}) + \varepsilon_c(Y_i, Y_i^0, X_i, V_i; \boldsymbol{\beta}) + S_m(Y_i^0, V_i; \boldsymbol{\beta}) + \varepsilon_m(Y_i, Y_i^0, X_i, V_i; \boldsymbol{\beta})$ are independent variables for $i = 1, 2, \ldots, n$, so, the $\sqrt{n}(\widehat{\boldsymbol{\beta}}_J - \boldsymbol{\beta})$ is asymptotically normally distributed with mean 0 and covariance matrix $\Delta_J = G^{-1}(\boldsymbol{\beta})M(\boldsymbol{\beta})\left[G^{-1}(\boldsymbol{\beta})\right]^T$. Therefore, the proof is completed. □

In practice, we need a consistent estimator of $\Delta_J$. For this purpose, let $G_n(\widehat{\boldsymbol{\beta}}_J)$ and $M_n(\widehat{\boldsymbol{\beta}}_J)$ be an estimator of $G(\boldsymbol{\beta})$ and $M(\boldsymbol{\beta})$, respectively, as follows:

$$G_n(\widehat{\boldsymbol{\beta}}_J) = \frac{1}{n}\sum_{i=1}^{n}\left[\delta_i \mathcal{X}_i \mathcal{X}_i^{\mathrm{T}} \widehat{H}_+^{(1)}(X_i, V_i; \widehat{\boldsymbol{\beta}}_J) + (1-\delta_i)\widehat{\mathcal{T}}_i(\widehat{\boldsymbol{\beta}}_J)\widehat{\mathcal{T}}_i^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_J)\widehat{H}_-^{(1)}(V_i; \widehat{\boldsymbol{\beta}}_J)\right]$$

and

$$\begin{aligned} M_n(\widehat{\boldsymbol{\beta}}_J) = \frac{1}{n}\sum_{i=1}^{n}\Big[&\widehat{S}_c(Y_i, Y_i^0, X_i, V_i; \widehat{\boldsymbol{\beta}}_J) + \widehat{S}_m(Y_i^0, V_i; \widehat{\boldsymbol{\beta}}_J) \\ &+ \widehat{\varepsilon}_c(Y_i, Y_i^0, X_i, V_i; \widehat{\boldsymbol{\beta}}_J) + \widehat{\varepsilon}_m(Y_i, Y_i^0, X_i, V_i; \widehat{\boldsymbol{\beta}}_J)\Big]^{\otimes 2}. \end{aligned}$$

Here

$$\begin{aligned} &\widehat{S}_c(Y_i, Y_i^0, X_i, V_i; \widehat{\boldsymbol{\beta}}_J) = \delta_i \mathcal{X}_i [Y_i - \widehat{H}_+(X_i, V_i; \widehat{\boldsymbol{\beta}}_J)]. \\ &\widehat{S}_m(Y_i^0, V_i; \widehat{\boldsymbol{\beta}}_J) = (1-\delta_i)\widehat{\mathcal{T}}_i(\widehat{\boldsymbol{\beta}}_J)[Y_i^0 - \widehat{H}_-(V_i; \widehat{\boldsymbol{\beta}}_J)]. \\ &\widehat{\varepsilon}_c(Y_i, Y_i^0, X_i, V_i; \widehat{\boldsymbol{\beta}}_J) = -\mathcal{X}_i \delta_i \widehat{H}_+^{(1)}(X_i, V_i; \widehat{\boldsymbol{\beta}}_J)\widehat{\mathcal{K}}_1^{-1}(X_i, V_i) \\ &\quad \times \left\{ \frac{\widehat{C}_3(X_i, V_i)[I(Y_i^0 = 1) - \widehat{A}(X_i, 1, V_i)]I(Y_i = 0)}{\widehat{P}(Y = 0|\delta = 1, X = X_i, V = V_i)} \right. \\ &\quad + \frac{\widehat{C}_4(X_i, V_i)[I(Y_i^0 = 0) - \widehat{A}(X_i, 0, V_i)]I(Y_i = 0)}{\widehat{P}(Y = 0|\delta = 1, X = X_i, V = V_i)} \\ &\quad + \frac{\widehat{C}_5(X_i, V_i)[I(Y_i^0 = 1) - \widehat{B}(X_i, 1, V_i)]I(Y_i = 1)}{\widehat{P}(Y = 1|\delta = 1, X = X_i, V = V_i)} \end{aligned}$$

$$+\frac{\widehat{C}_6(X_i,V_i)[I(Y_i^0=0)-\widehat{B}(X_i,0,V_i)]I(Y_i=1)}{\widehat{P}(Y=1|\delta=1,X=X_i,V=V_i)}\Bigg\}$$

$$+\frac{\frac{1}{n}\sum_{j=1}^n\left[\delta_j\mathcal{X}_j\widehat{H}_+^{(1)}(X_j,V_j;\widehat{\boldsymbol{\beta}}_J)\widehat{\mathcal{K}}_1^{-1}(X_j,V_j)\widehat{C}_1(X_j,V_j)I(V_j=V_i)\right]}{\widehat{P}(Y^0=1,V=V_i)}$$

$$\times[\delta_i-\widehat{\pi}(1,V_i)]I(Y_i^0=1)$$

$$+\frac{\frac{1}{n}\sum_{j=1}^n\left[\delta_j\mathcal{X}_j\widehat{H}_+^{(1)}(X_j,V_j;\widehat{\boldsymbol{\beta}}_J)\widehat{\mathcal{K}}_1^{-1}(X_j,V_j)\widehat{C}_2(X_j,V_j)I(V_j=V_i)\right]}{\widehat{P}(Y^0=0,V=V_i)}$$

$$\times[\delta_i-\widehat{\pi}(0,V_i)]I(Y_i^0=0)$$

for

$$\widehat{P}(Y=y|\delta=1,X=X_i,V=V_i)=\frac{\sum_{s=1}^n\delta_sI(Y_s=y,X_s=X_i,V_s=V_i)}{\sum_{j=1}^n\delta_jI(X_j=X_i,V_j=V_i)}$$

and

$$\widehat{P}(Y^0=y^0,V=V_i)=\frac{1}{n}\sum_{s=1}^nI(Y_s^0=y^0,V_s=V_i).$$

$$\begin{aligned}
\widehat{\varepsilon}_m(Y_i,Y_i^0,X_i,V_i;\widehat{\boldsymbol{\beta}}_J)=&-\widehat{\mathcal{T}}_i(\widehat{\boldsymbol{\beta}}_J)\widehat{H}_-^{(1)}(V_i;\widehat{\boldsymbol{\beta}}_J)\widehat{P}(\delta=0|V=V_i)\\
&\times\left\{\widehat{D}_{1i}(\widehat{\boldsymbol{\beta}}_J)\frac{I(Y_i^0=0)\,[\delta_i-\widehat{\pi}(0,V_i)]}{\widehat{P}(Y^0=0|V=V_i)}\right.\\
&+\widehat{D}_{2i}(\widehat{\boldsymbol{\beta}}_J)\frac{I(Y_i^0=1)\left[\delta_j-\widehat{\pi}(1,V_i)\right]}{\widehat{P}(Y^0=1|V=V_i)}\\
&+\widehat{D}_{3i}(\widehat{\boldsymbol{\beta}}_J)\left[\delta_iI(Y_i^0=0,Y_i=0)-\widehat{A}_0(0,V_i)\right]\\
&+\widehat{D}_{4i}(\widehat{\boldsymbol{\beta}})\left[\delta_iI(Y_i^0=1,Y_i=0)-\widehat{A}_0(1,V_i)\right]\\
&+\widehat{D}_{5i}(\widehat{\boldsymbol{\beta}}_J)\left[\delta_iI(Y_i^0=0,Y_i=1)-\widehat{B}_0(0,V_i)\right]\\
&+\widehat{D}_{6i}(\widehat{\boldsymbol{\beta}}_J)\left[\delta_iI(Y_i^0=1,Y_i=1)-\widehat{B}_0(1,V_i)\right]\\
&+\widehat{D}_{7i}(\widehat{\boldsymbol{\beta}}_J)\frac{I(Y_i^0=0,Y_i=0)\delta_i\left[e^{\widehat{\beta}_{J1}^{\mathrm{T}}X_i}-\widehat{r}_0(0,V_i;\widehat{\beta}_{J1})\right]}{\frac{\sum_{j=1}^n\delta_jI(Y_j^0=0,Y_j=0,V_j=V_i)}{\sum_{s=1}^nI(V_s=V_i)}}\\
&\left.+\widehat{D}_{8i}(\widehat{\boldsymbol{\beta}}_J)\frac{I(Y_i^0=1,Y_i=0)\delta_i\left[e^{\widehat{\beta}_{J1}^{\mathrm{T}}X_i}-\widehat{r}_0(1,V_i;\widehat{\beta}_{J1})\right]}{\frac{\sum_{j=1}^n\delta_jI(Y_j^0=0,Y_j=1,V_j=V_i)}{\sum_{s=1}^nI(V_s=V_i)}}\right\}
\end{aligned}$$

for

$$\widehat{P}(\delta = 0|V = V_i) = \frac{\sum_{j=1}^{n}(1-\delta_j)I(V_j = V_i)}{\sum_{s=1}^{n} I(V_s = V_i)},$$

and

$$\widehat{P}(Y^0 = y^0|V = V_i) = \frac{\sum_{s=1}^{n} I(Y_s^0 = y^0, V_s = V_i)}{\sum_{s=1}^{n} I(V_s = V_i)}.$$

$$\widehat{D}_{1i}(\widehat{\boldsymbol{\beta}}_J) = \frac{1}{1-\widehat{\pi}(0, V_i)} + \left[\frac{\widehat{H}_1(V_i; \widehat{\boldsymbol{\beta}}_J)}{1-\widehat{\theta}_0(V_i)} + \frac{\widehat{H}_2(V_i; \widehat{\boldsymbol{\beta}}_J)}{\widehat{\theta}_0(V_i)}\right] \frac{\widehat{\theta}_0(V_i)[1-\widehat{\theta}_0(V_i)]}{\widehat{\pi}(0, V_i)}$$
$$+ \frac{\widehat{H}_1(V_i; \widehat{\boldsymbol{\beta}}_J) - \widehat{H}_2(V_i; \widehat{\boldsymbol{\beta}}_J)}{\widehat{r}(V_i; \widehat{\beta}_{J1})} \frac{[\widehat{r}_0(1, V_i; \widehat{\beta}_{J1}) - \widehat{r}_0(0, V_i; \widehat{\beta}_{J1})]\widehat{\phi}_0(V_i)[1-\widehat{\phi}_0(V_i)]}{\widehat{\pi}(0, V_i)}$$
$$+ \left[\frac{\widehat{H}_1(V_i; \widehat{\boldsymbol{\beta}}_J)}{1-\widehat{\theta}_0(V_i)} + \frac{\widehat{H}_2(V_i; \widehat{\boldsymbol{\beta}}_J)}{\widehat{\theta}_0(V_i)}\right] \frac{\widehat{\phi}_0(V_i)[1-\widehat{\phi}_0(V_i)]}{\widehat{\pi}(0, V_i)} e^{-[\widehat{\beta}_{J0} + \widehat{\beta}_{J2}^{\mathrm{T}} Z_i + \widehat{R}(V_i; \widehat{\beta}_{J1})]}$$

for

$$\widehat{H}_1(V_i; \widehat{\boldsymbol{\beta}}_J) = H\left[\widehat{\beta}_{J0} + \widehat{\beta}_{J2}^{\mathrm{T}} Z_i + \widehat{R}(V_i; \widehat{\beta}_{J1}) + \ln \frac{1-\widehat{\theta}_0(V_i)}{\widehat{\phi}_0(V_i)}\right]$$

and

$$\widehat{H}_2(V_i; \widehat{\boldsymbol{\beta}}_J) = H\left[\widehat{\beta}_{J0} + \widehat{\beta}_{J2}^{\mathrm{T}} Z_i + \widehat{R}(V_i; \widehat{\beta}_{J1}) + \ln \frac{\widehat{\theta}_0(V_i)}{1-\widehat{\phi}_0(V_i)}\right].$$

$$\widehat{D}_{2i}(\widehat{\boldsymbol{\beta}}_J) = \frac{-1}{1-\widehat{\pi}(1, V_i)} - \left[\frac{\widehat{H}_1(V_i; \widehat{\boldsymbol{\beta}}_J)}{1-\widehat{\theta}_0(V_i)} + \frac{\widehat{H}_2(V_i; \widehat{\boldsymbol{\beta}}_J)}{\widehat{\theta}_0(V_i)}\right] \frac{\widehat{\theta}_0(V_i)[1-\widehat{\theta}_0(V_i)]}{\widehat{\pi}(1, V_i)}$$
$$- \frac{\widehat{H}_1(V_i; \widehat{\boldsymbol{\beta}}_J) - \widehat{H}_2(V_i; \widehat{\boldsymbol{\beta}}_J)}{\widehat{r}(V_i; \widehat{\beta}_{J1})} \frac{[\widehat{r}_0(1, V_i; \widehat{\beta}_{J1}) - \widehat{r}_0(0, V_i; \widehat{\beta}_{J1})]\widehat{\phi}_0(V_i)[1-\widehat{\phi}_0(V_i)]}{\widehat{\pi}(1, V_i)}$$
$$- \left[\frac{\widehat{H}_1(V_i; \widehat{\boldsymbol{\beta}}_J)}{1-\widehat{\theta}_0(V_i)} + \frac{\widehat{H}_2(V_i; \widehat{\boldsymbol{\beta}}_J)}{\widehat{\theta}_0(V_i)}\right] \frac{\widehat{\phi}_0(V_i)[1-\widehat{\phi}_0(V_i)]}{\widehat{\pi}(1, V_i)} e^{-[\widehat{\beta}_{J0} + \widehat{\beta}_{J2}^{\mathrm{T}} Z_i + \widehat{R}(V_i; \widehat{\beta}_{J1})]}.$$

$$\widehat{D}_{3i}(\widehat{\boldsymbol{\beta}}_J) = -\frac{\widehat{H}_1(V_i; \widehat{\boldsymbol{\beta}}_J) - \widehat{H}_2(V_i; \widehat{\boldsymbol{\beta}}_J)}{\widehat{r}(V_i; \widehat{\boldsymbol{\beta}}_{J1})} \frac{[\widehat{r}_0(1, V_i; \widehat{\beta}_{J1}) - \widehat{r}_0(0, V_i; \widehat{\beta}_{J1})]\widehat{\phi}_0(V_i)[1-\widehat{\phi}_0(V_i)]}{\widehat{A}_0(0, V_i)}$$
$$- \left[\frac{\widehat{H}_1(V_i; \widehat{\boldsymbol{\beta}}_J)}{1-\widehat{\theta}_0(V_i)} + \frac{\widehat{H}_2(V_i; \widehat{\boldsymbol{\beta}}_J)}{\widehat{\theta}_0(V_i)}\right] \frac{\phi_0(V_i)[1-\phi_0(V_i)]}{\widehat{A}_0(0, V_i)} e^{-[\widehat{\beta}_{J0} + \widehat{\beta}_{J2}^{\mathrm{T}} Z_i + \widehat{R}(V_i; \widehat{\beta}_{J1})]}.$$

$$\widehat{D}_{4i}(\widehat{\boldsymbol{\beta}}_J) = \frac{\widehat{H}_1(V_i; \widehat{\boldsymbol{\beta}}_J) - \widehat{H}_2(V_i; \widehat{\boldsymbol{\beta}}_J)}{\widehat{r}(V_i; \widehat{\boldsymbol{\beta}}_{J1})} \frac{[\widehat{r}_0(1, V_i; \widehat{\beta}_{J1}) - \widehat{r}_0(0, V_i; \widehat{\beta}_{J1})]\widehat{\phi}_0(V_i)[1-\widehat{\phi}_0(V_i)]}{\widehat{A}_0(1, V_i)}$$
$$+ \left[\frac{\widehat{H}_1(V_i; \widehat{\boldsymbol{\beta}}_J)}{1-\widehat{\theta}_0(V_i)} + \frac{\widehat{H}_2(V_i; \widehat{\boldsymbol{\beta}}_J)}{\widehat{\theta}_0(V_i)}\right] \frac{\widehat{\phi}_0(V_i)[1-\widehat{\phi}_0(V_i)]}{\widehat{A}_0(1, V_i)} e^{-[\widehat{\beta}_{J0} + \widehat{\beta}_{J2}^{\mathrm{T}} Z_i + \widehat{R}(V_i; \widehat{\beta}_{J1})]}.$$

$$\widehat{D}_{5i}(\widehat{\boldsymbol{\beta}}_J) = -\left[\frac{\widehat{H}_1(V_i; \widehat{\boldsymbol{\beta}}_J)}{1-\widehat{\theta}_0(V_i)} + \frac{\widehat{H}_2(V_i; \widehat{\boldsymbol{\beta}}_J)}{\widehat{\theta}_0(V_i)}\right] \frac{\widehat{\theta}_0(V_i)[1-\widehat{\theta}_0(V_i)]}{\widehat{B}_0(0, V_i)}.$$

$$\widehat{D}_{6i}(\widehat{\boldsymbol{\beta}}_J) = \left[ \frac{\widehat{H}_1(V_i; \widehat{\boldsymbol{\beta}}_J)}{1 - \widehat{\theta}_0(V_i)} + \frac{\widehat{H}_2(V_i; \widehat{\boldsymbol{\beta}}_J)}{\widehat{\theta}_0(V_i)} \right] \frac{\widehat{\theta}_0(V_i)[1 - \widehat{\theta}_0(V_i)]}{\widehat{B}_0(1, V_i)}.$$

$$\widehat{D}_{7i}(\widehat{\boldsymbol{\beta}}_J) = \frac{[\widehat{H}_1(V_i; \widehat{\boldsymbol{\beta}}_J) - \widehat{H}_2(V_i; \widehat{\boldsymbol{\beta}}_J)][1 - \widehat{\phi}_0(V_i)]}{\widehat{r}(V_i; \widehat{\beta}_{J1})}.$$

$$\widehat{D}_{8i}(\widehat{\boldsymbol{\beta}}_J) = \frac{[\widehat{H}_1(V_i; \widehat{\boldsymbol{\beta}}_J) - \widehat{H}_2(V_i; \widehat{\boldsymbol{\beta}}_J)]\widehat{\phi}_0(V_i)}{\widehat{r}(V_i; \widehat{\beta}_{J1})}.$$

As shown in the proof of Theorem 1, $\widehat{\boldsymbol{\beta}}_J \xrightarrow{p} \boldsymbol{\beta}$. Moreover, because $G_n(\cdot)$ is the sum of $iid$ random variables, so, we can have $G_n(\widehat{\boldsymbol{\beta}}_J) - G_n(\boldsymbol{\beta}) \xrightarrow{p} 0$. By weak law of large numbers, it can be shown that $G_n(\boldsymbol{\beta}) \xrightarrow{p} G(\boldsymbol{\beta})$. Therefore, it can be justified with Slutsky's theorem that $G_n(\widehat{\boldsymbol{\beta}}_J) \xrightarrow{p} G(\boldsymbol{\beta})$. Using the same arguments as above, one can show $M_n(\widehat{\boldsymbol{\beta}}_J) \xrightarrow{p} M(\boldsymbol{\beta})$. Therefore, a consistent estimator of $\Delta_J$ is given by $\widehat{\Delta}_J = G_n^{-1}(\widehat{\boldsymbol{\beta}}_J)M_n(\widehat{\boldsymbol{\beta}}_J)[G_n^{-1}(\widehat{\boldsymbol{\beta}}_J)]^{\mathrm{T}}$. □

## References


Albert PS, Hunsberger SA, Bird FM (1997) Modeling repeated measures with monotonic ordinal responses and misclassification, with applications to studying maturation. J Am Stat Assoc 92:1304–1311

Bollinger CR, David MH (1997) Modeling discrete choice with response error: food stamp participation. J Am Stat Assoc 92:827–835

Breslow NE, Cain KC (1988) Logistic regression for two-stage case-control data. Biometrika 75:11–20

Carroll RJ, Stefanski LA (1990) Approximate quasi-likelihood estimation in models with surrogate predictors. J Am Stat Asso 85:652–663

Chen J, Breslow NE (2004) Semiparametric efficient estimation for the auxiliary outcome problem with the conditional mean model. Can J Stat 32:359–372

Cheng KF, Hsueh HM (1999) Correcting bias due to misclassification in the estimation of logistic regression models. Stat Probab Lett 44:229–240

Cheng KF, Hsueh HM (2003) Estimation of a logistic regression model with mismeasured observations. Statistica Sinica 13:111–127

Chu H, Halloran ME (2004) Estimating vaccine efficacy using auxiliary outcome data and a small validation sample. Stat Med 23:2697–2711

Cox DR (1970) The analysis of binary data. Chapman and Hall, London

Foutz RV (1977) On the unique consistent solution to the likelihood equations. J Am Stat Assoc 72:147–148

Horvitz DG, Thompson DJ (1952) A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 47:663–685

Pepe MS (1992) Inference using surrogate outcome data and a validation sample. Biometrika 79:355–365

Pepe MS, Reilly M, Fleming TR (1994) Auxiliary outcome data and the mean-score method. J Stat Plan Inference 42:137–160

Pregibon D (1981) Logistic regression diagnostics. Ann Stat 9:705–724

Reilly M, Pepe MS (1995) A mean-score method for missing and auxiliary covariate data in regression models. Biometrika 82:299–314

Rosner B, Willett WC, Spiegelman DP (1989) Correction of logistic regression relative risk estimates and confidence intervals for systemtic within-person measurement error. Stat Med 8:1051–1069

Ronser B, Spiegelman D, Willett WC (1990) Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. Am J Epidemiol 132:734–745

Rubin DB (1976) Inference and missing data. Biometrika 63:581–592

Wang CY, Wang S (1997) Semiparametric methods in logistic regression with measurement error. Statistica Sinica 7:1103–1120

Wang CY, Chen JC, Lee SM, Ou ST (2002) Joint conditional likelihood estimator in logistic regression with missing covariate data. Statistica Sinica 12:555–574