Ann Inst Stat Math (2011) 63:887–921
DOI 10.1007/s10463-009-0266-9

# Conditional likelihood estimation and efficiency comparisons in proportional odds model with missing covariates

**S. H. Hsieh · S. M. Lee · P. S. Shen · M. F. Liu**

Received: 31 March 2008 / Revised: 1 April 2009 / Published online: 19 November 2009


**Abstract** In this article, a conditional likelihood approach is developed for dealing with ordinal data with missing covariates in proportional odds model. Based on the validation data set, we propose the Breslow and Cain (Biometrika 75:11–20, 1988) type estimators using different estimates of the selection probabilities, which may be treated as nuisance parameters. Under the assumption that the observed covariates and surrogate variables are categorical, we present large sample theory for the proposed estimators and show that they are more efficient than the estimator using the true selection probabilities. Simulation results support the theoretical analysis. We also illustrate the approaches using data from a survey of cable TV satisfaction.

**Keywords** Missing value · Proportional odds model · Ordinal categorical data · Conditional likelihood

## 1 Introduction

Ordered categorical data frequently arise in biomedical research. Sometimes categories are the result of grouping continuous data, such as age, or they arise if measurement is inherently imprecise, so that an interesting continuum can only be observed on an

S. H. Hsieh · S. M. Lee (✉) · M. F. Liu
Department of Statistics, Feng Chia University,
No. 100 Wenhaw Road Seatwen,
Taichung 40724, Taiwan, ROC
e-mail: smlee@fcu.edu.tw

P. S. Shen
Department of Statistics, Tunghai University,
181 Taichung Harbor Road, Section 3,
Taichung 40704, Taiwan, ROC

ordinal scale. However, ordinal data often result from subjective assessments under ordered categories, e.g. slight, moderate and severe. Several models for analyzing data with ordinal responses have been proposed in the literature, such as the cumulative logit or proportional odds model (Walker and Duncan 1967; McCullagh 1980), the constrained and unconstrained partial proportional odds model (Peterson and Harrell 1990) and the adjacent-category logistic model (Agresti 1984). The proportional odds model has been extensively used in medical studies for analyzing data with ordered categorical outcomes (see Ashby et al. 1989; McCusker et al. 1994; Faden and Graubard 2000 and Manor et al. 2000).

Usually medical and survey studies yield missing data. In this paper, we consider a proportional odds model with missing covariates. Let $Y$ be an ordinal categorical variable with possible outcomes denoted by $1, 2, \ldots, r$ corresponding to their order and $(X, Z)$ be a vector of covariates, where $Z$ is a vector of observable covariates, and $X$ is a vector of covariates that may be missing for some subjects. Then, the proportional odds model specifies that

$$
\begin{aligned}
P(Y \leq j|X, Z) &= \frac{1}{1 + \exp\left(-\theta_j - \beta_1^T X - \beta_2^T Z\right)} \\
&= H\left(\theta_j + \beta_1^T X + \beta_2^T Z\right), \qquad (1)
\end{aligned}
$$

where $j = 1, 2, \ldots, r - 1$ and $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_{r-1}$. Developing methods for regression analysis with missing covariate data has been an active research area in the past decade (see Little 1992; Little and Rubin 1987 for a review). A closely related problem arises when $X$ is missing and a surrogate variable for $X$ is available. Several methods have been proposed for this case (see, for example, Breslow and Cain 1988; Wang et al. 1997, 2002). When missingness does not depend on either $Y$ or $X$, Carroll and Wand (1991) and Pepe and Fleming (1991) proposed estimating the likelihood non-parametrically for continuous and discrete surrogate variables. When missingness does not depend on X (i.e. missing at random (MAR, see Rubin 1976)), Flander and Greenland (1991) and Zhao and Lipsitz (1992) suggested a weighted estimator.

In this article, we consider the case where $X$ is MAR, and a surrogate variable $W$ for $X$ is available and independent of $Y$ given $(X, Z)$. Moreover, $(W, Z)$ is assumed to be discrete. Let $\delta$ indicate whether $X$ is observed $(\delta = 1)$ or not $(\delta = 0)$. The validation data set $(\delta = 1)$ consists of $(Y, X, Z, W)$ and the non-validation data set $(\delta = 0)$ consists of $(Y, Z, W)$. The probability of $X$ being observed (selection probability) depends on $(Y, Z, W)$ but not on $X$, i.e. $P(\delta = 1|Y, X, Z, W) = \pi(Y, Z, W)$, which is assumed to be strictly positive. In our problem, $\pi(Y, Z, W)$ is a nuisance component to be estimated. We compare the efficiencies of several estimators of the parameters based on different estimation approaches $\pi(Y, Z, W)$ in model (1). This paper is organized as follows. In Sect. 2, we propose the Breslow and Cain (1988) conditional likelihood (CL) estimators using different estimates of $\pi(Y, Z, W)$s. In Sect. 3, we derive the asymptotic properties of the proposed estimators and compare the asymptotic variances. In Sect. 4, we conduct a simulation study to investigate the performance of the proposed estimators. In Sect. 5, the proposed estimators are applied to a survey of cable TV satisfaction. In Sect. 6, we provide some concluding remarks.

## 2 The proposed estimators

In this section under model (1), we propose the Breslow and Cain (1988) type conditional likelihood (CL) estimators using different estimates of the selection probabilities. Let $n$ be the sample size. For $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, r-1$, we defined indicator variable $T_{ij} = I[Y_i \leq j]$, where $I[\cdot]$ is an indicator function. Observe that $T_{i1} \leq T_{i2} \leq \cdots \leq T_{i,r-1}$ and that the overall response vector is $\boldsymbol{T}_i = (T_{i1}, \ldots, T_{i,r-1})$. Based on $T_{ij}$, the model (1) can be rewritten as follows:

$$P\left(T_{ij} = 1 | X_i, Z_i\right) = H\left(\theta_j + \beta_1^T X + \beta_2^T Z\right) = H\left(\boldsymbol{\Theta}^T \mathcal{X}_{ij}\right), \tag{2}$$

where $\boldsymbol{\Theta} = (\theta_1, \theta_2, \ldots, \theta_{r-1}, \beta_1^T, \beta_2^T)^T$, $\mathcal{X}_{ij} = (h_i^{(j)^T}, X_i^T, Z_i^T)^T$ and $h_i^{(j)}$ are the $(r-1) \times 1$ vector with 1 on the $j$th row and 0 on the rest of rows. When $X$ is missing at random (MAR), the selection probability is given by

$$P(\delta_i = 1 | Y_i, X_i, Z_i, W_i) = \pi(Y_i, Z_i, W_i) = \pi(Y_i, V_i),$$

where $V_i = (Z_i^T, W_i^T)^T$. Note that $(T_{i1}, T_{i2}, \ldots, T_{i,r-1})$ is a one-to-one transformation of $Y_i$ and $Y_i = r - \sum_{j=1}^{r-1} T_{ij}$. Hence,

$$\pi(Y_i, V_i) = P\left(\delta_i = 1 | r - \sum_{j=1}^{r-1} T_{ij}, V_i\right) = \pi\left(r - \sum_{j=1}^{r-1} T_{ij}, V_i\right).$$

Now, for $a = 0, 1$, we define $\widetilde{\pi}_j(a, V_i)$ as follows:

$$\begin{aligned}\widetilde{\pi}_j(a, V_i) &= P(\delta_i = 1 | T_{ij} = a, V_i) \\ &= \sum_{y \in w_a} P(\delta_i = 1 | Y_i = y, T_{ij} = a, V_i) P(Y_i = y | T_{ij} = a, V_i) \\ &= \sum_{y \in w_a} \pi(y, V_i) P(Y_i = y | T_{ij} = a, V_i),\end{aligned}$$

where $w_0 = \{j+1, \ldots, r\}$ and $w_1 = \{1, 2, \ldots, j\}$.

For logistic regression model, Breslow and Cain (1988) and Wang et al. (2002) proposed CL estimators based on the validation data set. In the same framework, we consider

$$\begin{aligned}P(T_{ij}=1 | X_i, V_i, \delta_i=1) &= \frac{P(T_{ij}=1, X_i, V_i, \delta_i=1)}{P(T_{ij}=1, X_i, V_i, \delta_i=1) + P(T_{ij}=0, X_i, V_i, \delta_i=1)} \\ &= \frac{\widetilde{\pi}_j(1, V_i) H(\boldsymbol{\Theta}^T \mathcal{X}_{ij})}{\widetilde{\pi}_j(1, V_i) H(\boldsymbol{\Theta}^T \mathcal{X}_{ij}) + \widetilde{\pi}_j(0, V_i)[1 - H(\boldsymbol{\Theta}^T \mathcal{X}_{ij})]}\end{aligned}$$

$$= H\left\{\boldsymbol{\Theta}^T\mathcal{X}_{ij} + \ln\left[\frac{\widetilde{\pi}_j(1, V_i)}{\widetilde{\pi}_j(0, V_i)}\right]\right\}$$
$$\equiv H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi}). \tag{3}$$

When $\widetilde{\pi}_j(a, V_i)$ is known, we define $U_{1n}(\boldsymbol{\Theta}, \widetilde{\pi})$ as follows:

$$U_{1n}(\boldsymbol{\Theta}, \widetilde{\pi}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\left\{\delta_i\mathcal{X}_{ij}[T_{ij} - H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi})]\right\}. \tag{4}$$

Based on the validation data set, an unbiased CL estimator of $\boldsymbol{\Theta}$, denoted by $\widehat{\boldsymbol{\Theta}}_t$, solves $U_{1n}(\boldsymbol{\Theta}, \widetilde{\pi}) = 0$. By direct calculation we have

$$\begin{aligned}
&\mathrm{E}\left\{\delta_i\delta_s\mathcal{X}_{ij}[T_{ij} - H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi})][T_{sk} - H_{+,k}(X_s, V_s; \boldsymbol{\Theta}, \widetilde{\pi})\mathcal{X}_{sk}^T]\right\}\\
&= \mathrm{E}\left\{\mathrm{E}\left\{\delta_i\delta_s\mathcal{X}_{ij}[T_{ij} - H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi})]\right.\right.\\
&\quad \left.\left.\times [T_{sk} - H_{+,k}(X_s, V_s; \boldsymbol{\Theta}, \widetilde{\pi})]\mathcal{X}_{sk}^T \middle| X_i, V_i, X_s, V_s\right\}\right\}\\
&= \begin{cases}
\mathrm{E}\left[\delta_1\mathcal{X}_{1j}H_{+,j}^{(1)}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})\mathcal{X}_{1j}^T\right] & \text{if } i = s, j = k,\\
\mathrm{E}\left[\delta_1\mathcal{X}_{1j}H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})[1 - H_{+,k}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})]\mathcal{X}_{1k}^T\right] & \text{if } i = s, j < k,\\
\mathrm{E}\left[\delta_1\mathcal{X}_{1j}H_{+,k}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})[1 - H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})]\mathcal{X}_{1k}^T\right] & \text{if } i = s, k < j,\\
0 & \text{if } i \neq s,
\end{cases}
\end{aligned}$$

where $H^{(1)}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi}) = H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})[1 - H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})]$. For any non-singular matrix $G$, define $G^{-T} = \{G^{-1}\}^T$. Moreover, it is easy to show that $\sqrt{n}(\widehat{\boldsymbol{\Theta}}_t - \boldsymbol{\Theta})$ is asymptotically normally distributed with mean 0 and covariance matrix

$$\Delta_t = G_0^{-1}(\boldsymbol{\Theta}, \widetilde{\pi})G(\boldsymbol{\Theta}, \widetilde{\pi})G_0^{-T}(\boldsymbol{\Theta}, \widetilde{\pi}), \tag{5}$$

where

$$G_0(\boldsymbol{\Theta}, \widetilde{\pi}) = \mathrm{E}\left[\sum_{j=1}^{r-1}\delta_1\mathcal{X}_{1j}H_{+,j}^{(1)}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})\mathcal{X}_{1j}^T\right],$$
$$G_1(\boldsymbol{\Theta}, \widetilde{\pi}) = \mathrm{E}\left[\sum_{j=1}^{r-2}\sum_{k=j+1}^{r-1}\delta_1\mathcal{X}_{1j}H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})[1 - H_{+,k}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})]\mathcal{X}_{1k}^T\right],$$

and

$$G(\boldsymbol{\Theta}, \widetilde{\pi}) = G_0(\boldsymbol{\Theta}, \widetilde{\pi}) + G_1(\boldsymbol{\Theta}, \widetilde{\pi}) + G_1^T(\boldsymbol{\Theta}, \widetilde{\pi}).$$

In general, the selection probabilities $\widetilde{\pi}_j(a, V_i)$ are unknown, and are nuisance parameters that remain to be estimated, although they may be prespecified at design stage in some applications. We will propose the CL estimator of $\boldsymbol{\Theta}$ using non-parametric and parametric estimates of $\widetilde{\pi}_j(a, V_i)$.

Let $v_1, \ldots, v_m$ denote the distinct values of the $V_i$s. For $v \in (v_1, v_2, \ldots, v_m)$, we define the non-parametric estimator of $\widetilde{\pi}_j(a, v)$ as follows:

$$\widehat{\widetilde{\pi}}_j(a, v) = \frac{\sum_{i=1}^n \delta_i I(T_{ij} = a, V_i = v)}{\sum_{i=1}^n I(T_{ij} = a, V_i = v)}. \tag{6}$$

By (6), we also define

$$\widehat{U}_{1n}(\boldsymbol{\Theta}, \widehat{\widetilde{\pi}}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{r-1} \left\{\delta_i \mathcal{X}_{ij}[T_{ij} - H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \widehat{\widetilde{\pi}})]\right\}, \tag{7}$$

where $H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \widehat{\widetilde{\pi}}) = H\left\{\boldsymbol{\Theta}^T \mathcal{X}_{ij} + \ln\left[\frac{\widehat{\widetilde{\pi}}_j(1, V_i)}{\widehat{\widetilde{\pi}}_j(0, V_i)}\right]\right\}$. Hence, the CL non-parametric estimator of $\boldsymbol{\Theta}$, denoted by $\widehat{\boldsymbol{\Theta}}_{np}$, is the solution to $\widehat{U}_{1n}(\boldsymbol{\Theta}, \widehat{\widetilde{\pi}}) = 0$.

Next, we assume a parametric model for $\pi(Y_i, V_i)$ as follows:

$$\begin{aligned}\pi(Y_i, V_i; \boldsymbol{\alpha}) &= P(\delta_i = 1 | Y_i, V_i)\\ &= \frac{1}{1 + \exp\{-\alpha_0 - \alpha_1 Y_i - \alpha_2^T Z_i - \alpha_3^T W_i\}}\\ &= H(\alpha_0 + \alpha_1 Y_i + \alpha_2^T Z_i + \alpha_3^T W_i)\\ &= H(\boldsymbol{\alpha}^T \mathcal{V}_i),\end{aligned} \tag{8}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2^T, \alpha_3^T)^T$ and $\mathcal{V}_i = (1, Y_i, V_i^T)^T$. By (8), $\widetilde{\pi}_j(\cdot)$ can be parameterized as follows:

$$\begin{aligned}\widetilde{\pi}_j(0, V_i; \boldsymbol{\alpha}) &= P(\delta_i = 1 | T_{ij} = 0, V_i)\\ &= P(\delta_i = 1 | Y_i > j, V_i)\\ &= \frac{\sum_{k=j+1}^r P(\delta_i = 1 | Y_i = k, V_i) \times P(Y_i = k | V_i)}{P(Y_i > j | V_i)}\\ &= \frac{\sum_{k=j+1}^r \pi(k, V_i; \boldsymbol{\alpha}) \times P(Y_i = k | V_i)}{\sum_{k=j+1}^r P(Y_i = k | V_i)},\end{aligned} \tag{9}$$

and

$$\begin{aligned}\widetilde{\pi}_j(1, V_i; \boldsymbol{\alpha}) &= P(\delta_i = 1 | T_{ij} = 1, V_i)\\ &= \frac{\sum_{k=1}^j \pi(k, V_i; \boldsymbol{\alpha}) \times P(Y_i = k | V_i)}{\sum_{k=1}^j P(Y_i = k | V_i)}.\end{aligned} \tag{10}$$

Therefore, we have

$$P(T_{ij}=1|X_i, V_i, \delta_i=1)=H\left\{\boldsymbol{\Theta}^T\mathcal{X}_{ij}+\ln\left[\frac{\widetilde{\pi}_j(1, V_i; \boldsymbol{\alpha})}{\widetilde{\pi}_j(0, V_i; \boldsymbol{\alpha})}\right]\right\}=H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \boldsymbol{\alpha}).$$

When the selection probabilities $\pi(k, V_i; \boldsymbol{\alpha})$ and $P(Y_i = k|V_i)$ are known, it can be shown by direct calculation that $E\left\{\delta_i\mathcal{X}_{ij}[T_{ij} - H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \boldsymbol{\alpha})]\right\} = 0$. Let the estimating score $U_{2n}(\boldsymbol{\Theta}, \boldsymbol{\alpha})$ be defined as follows:

$$U_{2n}(\boldsymbol{\Theta}, \boldsymbol{\alpha}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\left\{\delta_i\mathcal{X}_{ij}[T_{ij} - H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \boldsymbol{\alpha})]\right\}. \tag{11}$$

The CL estimator with known $\pi(k, V_i; \boldsymbol{\alpha})$ and $P(Y_i = k|V_i)$, denoted by $\widehat{\boldsymbol{\Theta}}_t^*$, solves $U_{2n}(\boldsymbol{\Theta}, \boldsymbol{\alpha}) = 0$. Note that $\widehat{\boldsymbol{\Theta}}_t^*$ and $\widehat{\boldsymbol{\Theta}}_t$ are equivalent.

When $\pi(k, V_i; \boldsymbol{\alpha})$ and $P(Y_i = k|V_i)$ are unknown, we need to be estimated before solving $U_{2n}(\boldsymbol{\Theta}, \boldsymbol{\alpha}) = 0$. The maximum likelihood estimator (MLE) $\widehat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$ in (8) can be obtained by solving $M_n(\boldsymbol{\alpha}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathcal{V}_i\{\delta_i - H(\boldsymbol{\alpha}^T\mathcal{V}_i)\} = 0$. Since the estimating equation is unbiased, $\widehat{\boldsymbol{\alpha}}$ is a consistent estimator of $\boldsymbol{\alpha}$. However, the $P(Y = k|V = V_i)$s are estimated non-parametrically

$$\widehat{P}_k(V_i) = \widehat{P}(Y = k|V = V_i) = \frac{\sum_{s=1}^{n} I(Y_s = k, V_s = V_i)}{\sum_{s=1}^{n} I(V_s = V_i)}. \tag{12}$$

Based on $\widehat{\boldsymbol{\alpha}}$ and $\widehat{P}_k(V_i)$, we define $\widehat{U}_{2n}(\boldsymbol{\Theta}, \widehat{\boldsymbol{\alpha}})$ as follows:

$$\widehat{U}_{2n}(\boldsymbol{\Theta}, \widehat{\boldsymbol{\alpha}}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\left\{\delta_i\mathcal{X}_{ij}[T_{ij} - \widehat{H}_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \widehat{\boldsymbol{\alpha}})]\right\}, \tag{13}$$

where

$$\widehat{H}_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \widehat{\boldsymbol{\alpha}}) = H\left\{\boldsymbol{\Theta}^T\mathcal{X}_{ij}+\ln\left[\frac{\widehat{\widetilde{\pi}}_j(1, V_i; \widehat{\boldsymbol{\alpha}})}{\widehat{\widetilde{\pi}}_j(0, V_i; \widehat{\boldsymbol{\alpha}})}\right]\right\},$$
$$\widehat{\widetilde{\pi}}_j(0, V_i; \widehat{\boldsymbol{\alpha}}) = \frac{\sum_{k=j+1}^{r}\pi(k, V_i; \widehat{\boldsymbol{\alpha}}) \times \widehat{P}_k(V_i)}{\sum_{k=j+1}^{r}\widehat{P}_k(V_i)},$$
$$\widehat{\widetilde{\pi}}_j(1, V_i; \widehat{\boldsymbol{\alpha}}) = \frac{\sum_{k=1}^{j}\pi(k, V_i; \widehat{\boldsymbol{\alpha}}) \times \widehat{P}_k(V_i)}{\sum_{k=1}^{j}\widehat{P}_k(V_i)}.$$

Then, the CL parametric estimator of $\boldsymbol{\Theta}$, denoted by $\widehat{\boldsymbol{\Theta}}_{vp}$, is the root of $\widehat{U}_{2n}(\boldsymbol{\Theta}, \widehat{\boldsymbol{\alpha}}) = 0$. When $\boldsymbol{\alpha}$ in (8) is known, based on $\widehat{P}_k(V_i)$, we define

$$\widehat{U}_{3n}(\boldsymbol{\Theta}, \boldsymbol{\alpha}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\left\{\delta_i\mathcal{X}_{ij}[T_{ij} - \widehat{H}_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \boldsymbol{\alpha})]\right\}, \tag{14}$$

where

$$\widehat{H}_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \boldsymbol{\alpha}) = H\left\{\boldsymbol{\Theta}^T \mathcal{X}_{ij} + \ln\left[\frac{\widehat{\widetilde{\pi}}_j(1, V_i; \boldsymbol{\alpha})}{\widehat{\widetilde{\pi}}_j(0, V_i; \boldsymbol{\alpha})}\right]\right\},$$

$$\widehat{\widetilde{\pi}}_j(0, V_i; \boldsymbol{\alpha}) = \frac{\sum_{k=j+1}^{r} \pi(k, V_i; \boldsymbol{\alpha}) \times \widehat{P}_k(V_i)}{\sum_{k=j+1}^{r} \widehat{P}_k(V_i)},$$

$$\widehat{\widetilde{\pi}}_j(1, V_i; \boldsymbol{\alpha}) = \frac{\sum_{k=1}^{j} \pi(k, V_i; \boldsymbol{\alpha}) \times \widehat{P}_k(V_i)}{\sum_{k=1}^{j} \widehat{P}_k(V_i)}.$$

Hence, the CL semiparametric estimator of $\boldsymbol{\Theta}$, denoted by $\widehat{\boldsymbol{\Theta}}_{sp}$, is the solution to $\widehat{U}_{3n}(\boldsymbol{\Theta}, \boldsymbol{\alpha}) = 0$.

## 3 Asymptotic theory

In this section, we provide asymptotic results under the assumption that $V$ is discrete and $X$ is MAR. The following regularity conditions are crucial in deriving the asymptotic properties of $\widehat{\boldsymbol{\Theta}}_{np}$, $\widehat{\boldsymbol{\Theta}}_{vp}$ and $\widehat{\boldsymbol{\Theta}}_{sp}$, where $\widehat{\boldsymbol{\Theta}}_k = (\widehat{\theta}_{k1}, \widehat{\theta}_{k2}, \ldots, \widehat{\theta}_{k,r-1}, \widehat{\beta}_{k1}^T, \widehat{\beta}_{k2}^T)^T$ for $k = np,\ vp$ and $sp$.

(A1) Let $\mathrm{supp}(V)$ denote the support of V. For any $y = 1, 2, \ldots, r$ and $v \in \mathrm{supp}(V)$, the selection probability $\pi(y, v) > 0$.

(A2) For any $y = 1, 2, \ldots, r$ and $v \in \mathrm{supp}(V)$, the selection probability $\pi(y, v) < 1$.

(A3) $\mathrm{E}\left\{\sum_{j=1}^{r-1} \delta_1 \mathcal{X}_{1j} H_{+,j}^{(1)}(\cdot) \mathcal{X}_{1j}^T\right\}$ is positive definite in a neighborhood of the true $\boldsymbol{\Theta}$, where $H_{+,j}^{(1)}(\cdot) = H_{+,j}(\cdot)[1 - H_{+,j}(\cdot)]$ and $\mathcal{X}_{1j}^T = (h_1^{(j)^T}, X_1^T, Z_1^T)$.

(A4) For any $s = 1, 2, 3$, the first derivatives of $\widehat{U}_{sn}(\cdot)$ with respect to $\boldsymbol{\Theta}$ exist almost surely in a neighborhood of the true $\boldsymbol{\Theta}$. Further, in such a neighborhood, the second derivatives are bounded above by a function of $(T, X, V)$, whose expectation exists.

(A5) $\mathrm{E}\{\mathcal{V} H^{(1)}(\boldsymbol{\alpha}^T \mathcal{V}) \mathcal{V}^T\}$ is positive definite, where $H^{(1)}(\boldsymbol{\alpha}^T \mathcal{V}) = H(\boldsymbol{\alpha}^T \mathcal{V})[1 - H(\boldsymbol{\alpha}^T \mathcal{V})]$ and $\mathcal{V}^T = (1, Y, V^T)$.

Using the estimated non-parametric selection probability, the following Lemma shows that the score of the CL non-parametric estimator, $\widehat{U}_{1n}(\boldsymbol{\Theta}, \widehat{\widetilde{\pi}})$ (defined in (7)), can be expressed as the sum of independent random variables.

**Lemma 1** *Under the conditions* (A1) and (A2),

$$\widehat{U}_{1n}(\boldsymbol{\Theta}, \widehat{\widetilde{\pi}}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{r-1} \left\{S_j(T_{ij}, X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi}) - \varepsilon_j(T_{ij}, V_i; \boldsymbol{\Theta}, \widetilde{\pi})\right\} + o_p(1),$$

*where*

$$\begin{aligned} S_j(T_{ij}, X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi}) &= \delta_i \mathcal{X}_{ij}[T_{ij} - H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi})], \\ \varepsilon_j(T_{ij}, V_i; \boldsymbol{\Theta}, \widetilde{\pi}) &= E_j^*(V_i) \left\{ \frac{[\delta_i - \widetilde{\pi}_j(1, V_i)] I(T_{ij} = 1)}{\widetilde{\pi}_j(1, V_i) P(T_{1j} = 1 | V = V_i)} \right. \\ &\quad \left. - \frac{[\delta_i - \widetilde{\pi}_j(0, V_i)] I(T_{ij} = 0)}{\widetilde{\pi}_j(0, V_i) P(T_{1j} = 0 | V = V_i)} \right\}, \end{aligned}$$

*and*

$$E_j^*(V_i) = E\left\{ \delta_1 \mathcal{X}_{1j} H_{+,j}^{(1)}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi}) | V_1 = V_i \right\}.$$

The proof of the Lemma 1 is given in the Appendix. Note that $\varepsilon_j(\cdot)$ stands for the approximation error from the complete data score $S_j(\cdot)$, which is due to the estimation of the nuisance parameter $\widetilde{\pi}_j(\cdot)$.

Next, we derive the asymptotic properties of $\widehat{\boldsymbol{\Theta}}_{np}$. We define

$$\begin{aligned} G_0(\boldsymbol{\Theta}, \widetilde{\pi}) &= \mathrm{E}\left[ \sum_{j=1}^{r-1} \delta_1 \mathcal{X}_{1j} H_{+,j}^{(1)}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi}) \mathcal{X}_{1j}^T \right], \\ G_1(\boldsymbol{\Theta}, \widetilde{\pi}) &= \mathrm{E}\left[ \sum_{j=1}^{r-2} \sum_{k=j+1}^{r-1} \delta_1 \mathcal{X}_{1j} H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})[1 - H_{+,k}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})] \mathcal{X}_{1k}^T \right], \\ G(\boldsymbol{\Theta}, \widetilde{\pi}) &= G_0(\boldsymbol{\Theta}, \widetilde{\pi}) + G_1(\boldsymbol{\Theta}, \widetilde{\pi}) + G_1^T(\boldsymbol{\Theta}, \widetilde{\pi}), \\ M_0(\boldsymbol{\Theta}; \widetilde{\pi}) &= \sum_{j=1}^{r-1} \mathrm{E}\left\{ \mathrm{E}_j^*(V_1) \left[ \frac{1 - \widetilde{\pi}_j(1, V_1)}{\widetilde{\pi}_j(1, V_1) P(T_{1j} = 1 | V = V_1)} \right. \right. \\ &\quad \left. \left. + \frac{1 - \widetilde{\pi}_j(0, V_1)}{\widetilde{\pi}_j(0, V_1) P(T_{1j} = 0 | V = V_1)} \right] \mathrm{E}_j^{*^T}(V_1) \right\}, \\ M_1(\boldsymbol{\Theta}; \widetilde{\pi}) &= \sum_{j=1}^{r-2} \sum_{k=j+1}^{r-1} \mathrm{E}\left\{ \mathrm{E}_j^*(V_1) \left[ \frac{1 - \widetilde{\pi}_k(1, V_1)}{\widetilde{\pi}_k(1, V_1) P(T_{1k} = 1 | V = V_1)} \right. \right. \\ &\quad \left. \left. + \frac{1 - \widetilde{\pi}_k(0, V_1)}{\widetilde{\pi}_k(0, V_1) P(T_{1k} = 0 | V = V_1)} \right] \mathrm{E}_k^{*^T}(V_1) \right\}, \\ A_1(\boldsymbol{\Theta}; \widetilde{\pi}) &= \sum_{j=1}^{r-2} \sum_{k=j+1}^{r-1} \mathrm{E}\left\{ \mathrm{E}_j^*(V_1) \left[ \frac{1 - \widetilde{\pi}_j(0, V_1)}{\widetilde{\pi}_j(0, V_1) P(T_{1j} = 0 | V = V_1)} \right. \right. \\ &\quad \left. \left. - \frac{1 - \widetilde{\pi}_k(0, V_1)}{\widetilde{\pi}_k(0, V_1) P(T_{1k} = 0 | V = V_1)} \right] \mathrm{E}_k^{*T}(V_1) \right\}, \end{aligned}$$

$$A_2(\boldsymbol{\Theta};\widetilde{\pi}) = \sum_{k=1}^{r-2}\sum_{j=k+1}^{r-1} \mathrm{E}\left\{\mathrm{E}_j^*(V_1)\left[\frac{1-\widetilde{\pi}_j(1,V_1)}{\widetilde{\pi}_j(1,V_1)P(T_{1j}=1|V=V_1)}\right.\right.$$
$$\left.\left.-\frac{1-\widetilde{\pi}_k(1,V_1)}{\widetilde{\pi}_k(1,V_1)P(T_{1k}=1|V=V_1)}\right]\mathrm{E}_k^{*T}(V_1)\right\},$$
$$M(\boldsymbol{\Theta};\widetilde{\pi}) = M_0(\boldsymbol{\Theta};\widetilde{\pi}) + M_1(\boldsymbol{\Theta};\widetilde{\pi}) + A_1(\boldsymbol{\Theta};\widetilde{\pi}) + M_1^T(\boldsymbol{\Theta};\widetilde{\pi}) + A_2(\boldsymbol{\Theta};\widetilde{\pi}),$$
$$B_1(\boldsymbol{\Theta};\widetilde{\pi}) = \sum_{j=1}^{r-2}\sum_{k=j+1}^{r-1} \mathrm{E}\left\{\delta_1\mathcal{X}_{1j}\left[H_{+,j}(X_1,V_1;\boldsymbol{\Theta},\widetilde{\pi}) - H_{+,k}(X_1,V_1;\boldsymbol{\Theta},\widetilde{\pi})\right]\right.$$
$$\times\left[\frac{1-\widetilde{\pi}_k(1,V_1)}{\widetilde{\pi}_k(1,V_1)P(T_{1k}=1|V=V_1)}\right.$$
$$\left.\left.+\frac{1-\widetilde{\pi}_k(0,V_1)}{\widetilde{\pi}_k(0,V_1)P(T_{1k}=0|V=V_1)}\right]E_k^{*T}(V_1)\right\},$$

and

$$A(\boldsymbol{\Theta};\widetilde{\pi}) = -A_1(\boldsymbol{\Theta};\widetilde{\pi}) - A_2(\boldsymbol{\Theta};\widetilde{\pi}) + B_1(\boldsymbol{\Theta};\widetilde{\pi}) + B_1^T(\boldsymbol{\Theta};\widetilde{\pi}).$$

**Theorem 1** *Under the conditions* (A1)–(A4), $\widehat{\boldsymbol{\Theta}}_{np}$ *is a consistent estimator of* $\boldsymbol{\Theta}$ *and* $\sqrt{n}(\widehat{\boldsymbol{\Theta}}_{np}-\boldsymbol{\Theta})$ *has an asymptotic normal distribution with mean 0 and covariance matrix*

$$\Delta_{np} = G_0^{-1}(\boldsymbol{\Theta},\widetilde{\pi})\left\{G(\boldsymbol{\Theta},\widetilde{\pi}) - M(\boldsymbol{\Theta},\widetilde{\pi}) - 2A(\boldsymbol{\Theta},\widetilde{\pi})\right\}G_0^{-T}(\boldsymbol{\Theta},\widetilde{\pi}).$$

The proof of Theorem 1 is given in Appendix.

*Remark 1* When $r = 2$, it follows from the Theorem 1 that the proportional odds model is reduced to logistic regression model and

$$\Delta_{np} = G_0^{-1}(\boldsymbol{\Theta},\widetilde{\pi})\left\{G_0^{-1}(\boldsymbol{\Theta},\widetilde{\pi}) - M_0(\boldsymbol{\Theta},\widetilde{\pi})\right\}G_0^{-T}(\boldsymbol{\Theta},\widetilde{\pi}).$$

We now derive a consistent estimator of $\Delta_{np}$. Let

$$G_{0n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\delta_i\mathcal{X}_{ij}H_{+,j}^{(1)}\left(X_i,V_i;\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}\right)\mathcal{X}_{ij}^T,$$
$$G_{1n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-2}\sum_{k=j+1}^{r-1}\delta_i\mathcal{X}_{ij}H_{+,j}\left(X_i,V_i;\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}\right)$$
$$\times\left[1-H_{+,k}(X_i,V_i;\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}})\right]\mathcal{X}_{ik}^T,$$
$$G_n(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}) = G_{0n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}) + G_{1n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}) + G_{1n}^T(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}),$$
$$\widehat{\mathrm{E}}_j^*(V_k) = \frac{\sum_{i=1}^n\delta_i\mathcal{X}_{ij}H_{+,j}^{(1)}(X_i,V_i;\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}})I(V_i=V_k)}{\sum_{i=1}^n I(V_i=V_k)},$$

$$
\begin{aligned}
M_{0n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\widehat{\mathrm{E}}_j^*(V_i)\left[\frac{1-\widehat{\widetilde{\pi}}_j(1,V_i)}{\widehat{\widetilde{\pi}}_j(1,V_i)\sum_{s=1}^{j}\widehat{P}_s(V_i)}\right.\\
&\quad\left.+\frac{1-\widehat{\widetilde{\pi}}_j(0,V_i)}{\widehat{\widetilde{\pi}}_j(0,V_i)\sum_{s=j+1}^{r}\widehat{P}_s(V_i)}\right]\widehat{\mathrm{E}}_j^{*^T}(V_i),\\
M_{1n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-2}\sum_{k=j+1}^{r-1}\widehat{\mathrm{E}}_j^*(V_i)\left[\frac{1-\widehat{\widetilde{\pi}}_k(1,V_i)}{\widehat{\widetilde{\pi}}_k(1,V_i)\sum_{s=1}^{k}\widehat{P}_s(V_i)}\right.\\
&\quad\left.+\frac{1-\widehat{\widetilde{\pi}}_k(0,V_i)}{\widehat{\widetilde{\pi}}_k(0,V_i)\sum_{s=k+1}^{r}\widehat{P}_s(V_i)}\right]\widehat{\mathrm{E}}_k^{*^T}(V_i),\\
A_{1n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-2}\sum_{k=j+1}^{r-1}\widehat{\mathrm{E}}_j^*(V_i)\left[\frac{1-\widehat{\widetilde{\pi}}_j(0,V_i)}{\widehat{\widetilde{\pi}}_j(0,V_i)\sum_{s=k+1}^{r}\widehat{P}_s(V_i)}\right.\\
&\quad\left.-\frac{1-\widehat{\widetilde{\pi}}_k(0,V_i)}{\widehat{\widetilde{\pi}}_k(0,V_i)\sum_{s=k+1}^{r}\widehat{P}_s(V_i)}\right]\mathrm{E}_k^{*^T}(V_i),\\
A_{2n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{r-2}\sum_{j=k+1}^{r-1}\widehat{\mathrm{E}}_j^*(V_i)\left[\frac{1-\widehat{\widetilde{\pi}}_j(1,V_i)}{\widehat{\widetilde{\pi}}_j(1,V_i)\sum_{s=1}^{j}\widehat{P}_s(V_i)}\right.\\
&\quad\left.-\frac{1-\widehat{\widetilde{\pi}}_k(1,V_i)}{\widehat{\widetilde{\pi}}_k(1,V_i)\sum_{s=1}^{j}\widehat{P}_s(V_i)}\right]\widehat{\mathrm{E}}_k^{*^T}(V_i),\\
M_n(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}) &= M_{0n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}})+M_{1n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}})+A_{1n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}})\\
&\quad+M_{1n}^T(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}})+A_{2n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}),\\
B_{1n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-2}\sum_{k=j+1}^{r-1}\delta_i\mathcal{X}_{ij}\\
&\quad\times[H_{+,j}(X_i,V_i;\boldsymbol{\Theta},\widehat{\widetilde{\pi}})-H_{+,k}(X_i,V_i;\boldsymbol{\Theta},\widehat{\widetilde{\pi}})]\\
&\quad\times\left[\frac{[1-\widehat{\widetilde{\pi}}_k(1,V_i)]}{\widehat{\widetilde{\pi}}_k(1,V_i)\sum_{s=1}^{k}\widehat{P}_s(V_i)}+\frac{[1-\widehat{\widetilde{\pi}}_k(0,V_i)]}{\widehat{\widetilde{\pi}}_k(0,V_i)\sum_{s=k+1}^{r}\widehat{P}_s(V_i)}\right]\widehat{E}_k^{*^T}(V_i),
\end{aligned}
$$

and

$$
A_n(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}) = -A_{1n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}})-A_{2n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}})+B_{1n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}})+B_{1n}^T(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}).
$$

We already know that $\widehat{\widetilde{\pi}}_j \xrightarrow{p} \widetilde{\pi}_j$. Also the Inverse Function Theorem of Foutz (1997), along with Condition (A3), we have $\widehat{\boldsymbol{\Theta}}_{np} \xrightarrow{p} \boldsymbol{\Theta}$. It follows from continuity of function $H_{+,j}(\cdot)$, that $H_{+,j}(X,V;\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}) \xrightarrow{p} H_{+,j}(X,V;\boldsymbol{\Theta},\pi)$. Moreover, note that $G_{0n}(\cdot)$ is the sum of $iid$ random variables, and hence $G_{0n}(\widehat{\boldsymbol{\Theta}}_{np},\widetilde{\pi}) - G_{0n}(\boldsymbol{\Theta},\widetilde{\pi}) \xrightarrow{p} 0$. By weak law of large number, it is clear that $G_{0n}(\boldsymbol{\Theta},\widetilde{\pi}) \xrightarrow{p} G_0(\boldsymbol{\Theta},\widetilde{\pi})$. Therefore, the covergence of $G_{0n}(\widehat{\boldsymbol{\Theta}}_{np},\widetilde{\pi}) \xrightarrow{p} G_0(\boldsymbol{\Theta},\widetilde{\pi})$ can be justified.

Using the same process as described above, the law of large number properties are summarized as follows:

$G_{1n}(\widehat{\boldsymbol{\Theta}}_{np}, \widehat{\widetilde{\pi}}) \xrightarrow{p} G_1(\boldsymbol{\Theta}, \widetilde{\pi}), G_n(\widehat{\boldsymbol{\Theta}}_{np}, \widehat{\widetilde{\pi}}) \xrightarrow{p} G(\boldsymbol{\Theta}, \widetilde{\pi}), M_{0n}(\widehat{\boldsymbol{\Theta}}_{np}, \widehat{\widetilde{\pi}}) \xrightarrow{p} M_0(\boldsymbol{\Theta}, \widetilde{\pi})$, $M_{1n}(\widehat{\boldsymbol{\Theta}}_{np}, \widehat{\widetilde{\pi}}) \xrightarrow{p} M_1(\boldsymbol{\Theta}, \widetilde{\pi})$, $A_{1n}(\widehat{\boldsymbol{\Theta}}_{np}, \widehat{\widetilde{\pi}}) \xrightarrow{p} A_1(\boldsymbol{\Theta}, \widetilde{\pi})$, $A_{2n}(\widehat{\boldsymbol{\Theta}}_{np}, \widehat{\widetilde{\pi}}) \xrightarrow{p} A_2(\boldsymbol{\Theta}, \widetilde{\pi})$, $M_n(\widehat{\boldsymbol{\Theta}}_{np}, \widehat{\widetilde{\pi}}) \xrightarrow{p} M(\boldsymbol{\Theta}, \widetilde{\pi})$, $B_{1n}(\widehat{\boldsymbol{\Theta}}_{np}, \widehat{\widetilde{\pi}}) \xrightarrow{p} B_1(\boldsymbol{\Theta}, \widetilde{\pi})$ and $A_n(\widehat{\boldsymbol{\Theta}}_{np}, \widehat{\widetilde{\pi}}) \xrightarrow{p} A(\boldsymbol{\Theta}, \widetilde{\pi})$. Hence, a consistent estimator of $\Delta_{np}$ is given as follows:

$$\widehat{\Delta}_{np} = G_{0n}^{-1}(\widehat{\boldsymbol{\Theta}}_{np}, \widehat{\widetilde{\pi}}) \left\{G_n(\widehat{\boldsymbol{\Theta}}_{np}, \widehat{\widetilde{\pi}}) - M_n(\widehat{\boldsymbol{\Theta}}_{np}, \widehat{\widetilde{\pi}}) - 2A_n(\widehat{\boldsymbol{\Theta}}_{np}, \widehat{\widetilde{\pi}})\right\} G_{0n}^{-T}(\widehat{\boldsymbol{\Theta}}_{np}, \widehat{\widetilde{\pi}}).$$

We now present the asymptotic results of $\widehat{\boldsymbol{\Theta}}_{vp}$. Define

$$\begin{aligned}
G_0(\boldsymbol{\Theta}, \boldsymbol{\alpha}) &= \mathrm{E}\left[\sum_{j=1}^{r-1} \delta_1 \mathcal{X}_{1j} H_{+,j}^{(1)}(X_1, V_1; \boldsymbol{\Theta}, \boldsymbol{\alpha}) \mathcal{X}_{1j}^T\right], \\
G_1(\boldsymbol{\Theta}, \boldsymbol{\alpha}) &= \mathrm{E}\left[\sum_{j=1}^{r-2} \sum_{k=j+1}^{r-1} \delta_1 \mathcal{X}_{1j} H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \boldsymbol{\alpha})[1 - H_{+,k}(X_1, V_1; \boldsymbol{\Theta}, \boldsymbol{\alpha})] \mathcal{X}_{1k}^T\right], \\
G(\boldsymbol{\Theta}, \boldsymbol{\alpha}) &= G_0(\boldsymbol{\Theta}, \boldsymbol{\alpha}) + G_1(\boldsymbol{\Theta}, \boldsymbol{\alpha}) + G_1^T(\boldsymbol{\Theta}, \boldsymbol{\alpha}), \\
K(\boldsymbol{\Theta}, \boldsymbol{\alpha}) &= \mathrm{E}\left\{\sum_{j=1}^{r-1} \delta_1 \mathcal{X}_{1j} H_{+,j}^{(1)}(X_1, V_1; \boldsymbol{\Theta}, \boldsymbol{\alpha}) \left[\frac{\frac{\partial \widetilde{\pi}_j(1, V_1; \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^T}}{\widetilde{\pi}_j(1, V_1; \boldsymbol{\alpha})} - \frac{\frac{\partial \widetilde{\pi}_j(0, V_1; \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^T}}{\widetilde{\pi}_j(0, V_1; \boldsymbol{\alpha})}\right]\right\}, \\
I(\boldsymbol{\alpha}) &= \mathrm{E}\left\{\mathcal{V} H^{(1)}(\boldsymbol{\alpha}^T \mathcal{V}) \mathcal{V}^T\right\}, \\
\Omega_k(V_s; \boldsymbol{\Theta}, \boldsymbol{\alpha}) &= \mathrm{E}\left\{\sum_{j=1}^{r-1} \delta_1 \mathcal{X}_{1j} H_{+,j}^{(1)}(X_1, V_1; \boldsymbol{\Theta}, \boldsymbol{\alpha})\right. \\
&\quad \left. \times \left[\frac{\frac{\partial \widetilde{\pi}_j(1, V_1; \boldsymbol{\alpha})}{\partial P_k(V_1)}}{\widetilde{\pi}_j(1, V_1; \boldsymbol{\alpha})} - \frac{\frac{\partial \widetilde{\pi}_j(0, V_1; \boldsymbol{\alpha})}{\partial P_k(V_1)}}{\widetilde{\pi}_j(0, V_1; \boldsymbol{\alpha})}\right] \middle| V_1 = V_s\right\}, \\
J(\boldsymbol{\Theta}, \boldsymbol{\alpha}) &= \mathrm{E}\left\{\sum_{j=1}^{r-1} \mathcal{X}_{1j}[1 - H_{+,j}(X_1, V_1; \boldsymbol{\alpha}, \boldsymbol{\Theta})]\right. \\
&\quad \left. \times \left[\sum_{k=1}^{j} H^{(1)}(\boldsymbol{\alpha}^T \mathcal{V}_{k,1}) P(Y_1 = k | X_1, V_1) \mathcal{V}_{k,1}^T\right]\right\} \\
&\quad - \mathrm{E}\left\{\sum_{j=1}^{r-1} \mathcal{X}_{1j} H_{+,j}(X_1, V_1; \boldsymbol{\alpha}, \boldsymbol{\Theta})\right. \\
&\quad \left. \times \left[\sum_{k=j+1}^{r} H^{(1)}(\boldsymbol{\alpha}^T \mathcal{V}_{k,1}) P(Y_1 = k | X_1, V_1) \mathcal{V}_{k,1}^T\right]\right\}, \\
\mathcal{V}_{k,1} &= (1, k, V_1^T)^T,
\end{aligned}$$

$$Q(\boldsymbol{\Theta},\boldsymbol{\alpha}) = \mathrm{E}\left\{\sum_{j=1}^{r-1}\mathcal{X}_{1j}[1-H_{+,j}(X_1,V_1;\boldsymbol{\alpha},\boldsymbol{\Theta})]\right.$$
$$\left.\times\left[\sum_{k=1}^{j}H(\boldsymbol{\alpha}^T\mathcal{V}_{k,1})P(Y=k|X_1,V_1)\Omega_k^T(V_1;\boldsymbol{\Theta},\boldsymbol{\alpha})\right]\right\}$$
$$-\,\mathrm{E}\left\{\sum_{j=1}^{r-1}\mathcal{X}_{1j}H_{+,j}(X_1,V_1;\boldsymbol{\alpha},\boldsymbol{\Theta})\right.$$
$$\left.\times\left[\sum_{k=j+1}^{r}H(\boldsymbol{\alpha}^T\mathcal{V}_{k,1})P(Y=k|X_1,V_1)\Omega_k^T(V_1;\boldsymbol{\Theta},\boldsymbol{\alpha})\right]\right\},$$

and

$$P(\boldsymbol{\Theta},\boldsymbol{\alpha}) = \sum_{j=1}^{r}\mathrm{E}\left\{\Omega_j(V_1;\boldsymbol{\Theta},\boldsymbol{\alpha})[1-P_j(V_1)]P_j(V_1)\Omega_j^T(V_1;\boldsymbol{\Theta},\boldsymbol{\alpha})\right\}$$
$$-2\sum_{j=1}^{r}\sum_{j<k}\mathrm{E}\left\{\Omega_j(V_1;\boldsymbol{\Theta},\boldsymbol{\alpha})P_j(V_1)P_k(V_1)\Omega_k^T(V_1;\boldsymbol{\Theta},\boldsymbol{\alpha})\right\}.$$

**Theorem 2** *Under the conditions* (A1)–(A5)*, $\widehat{\boldsymbol{\Theta}}_{vp}$ is a consistent estimator of $\boldsymbol{\Theta}$ and $\sqrt{n}(\widehat{\boldsymbol{\Theta}}_{vp}-\boldsymbol{\Theta})$ has an asymptotic normal distribution with mean 0 and covariance matrix*

$$\Delta_{vp} = G_0^{-1}(\boldsymbol{\Theta},\boldsymbol{\alpha})\left\{G(\boldsymbol{\Theta},\boldsymbol{\alpha}) + K(\boldsymbol{\Theta},\boldsymbol{\alpha})I^{-1}(\boldsymbol{\alpha})K^T(\boldsymbol{\Theta},\boldsymbol{\alpha}) + P(\boldsymbol{\Theta},\boldsymbol{\alpha})\right.$$
$$\left.-\,2J(\boldsymbol{\Theta},\boldsymbol{\alpha})I^{-1}(\boldsymbol{\alpha})K^T(\boldsymbol{\Theta},\boldsymbol{\alpha}) - 2Q(\boldsymbol{\Theta},\boldsymbol{\alpha})\right\}G_0^{-T}(\boldsymbol{\Theta},\boldsymbol{\alpha}).$$

The proof of Theorem 2 is given in Appendix.

*Remark 2* When $r = 2$, it follows from Theorem 2 that $Q(\boldsymbol{\Theta},\boldsymbol{\alpha}) = P(\boldsymbol{\Theta},\boldsymbol{\alpha}) = 0$ and $J(\boldsymbol{\Theta},\boldsymbol{\alpha}) = K(\boldsymbol{\Theta},\boldsymbol{\alpha})$. Then,

$$\Delta_{vp} = G_0^{-1}(\boldsymbol{\Theta},\boldsymbol{\alpha})\left\{G_0(\boldsymbol{\Theta},\boldsymbol{\alpha}) - K(\boldsymbol{\Theta},\boldsymbol{\alpha})I^{-1}(\boldsymbol{\alpha})K^T(\boldsymbol{\Theta},\boldsymbol{\alpha})\right\}G_0^{-T}(\boldsymbol{\Theta},\boldsymbol{\alpha}).$$

We now give a consistent estimator of $\Delta_{vp}$. Let

$$G_{0n}(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\delta_i\mathcal{X}_{ij}H_{+,j}^{(1)}(X_i,V_i;\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}})\mathcal{X}_{ij}^T,$$
$$G_{1n}(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-2}\sum_{k=j+1}^{r-1}\delta_i\mathcal{X}_{ij}H_{+,j}(X_i,V_i;\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}})$$
$$\times[1-H_{+,k}(X_i,V_i;\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}})]\mathcal{X}_{ik}^T,$$

$$G_n(\widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}}) = G_{0n}(\widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}}) + G_{1n}(\widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}}) + G_{1n}^T(\widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}}),$$

$$K_n(\widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\delta_i \mathcal{X}_{ij} H_{+,j}^{(1)}(X_i, V_i; \widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}}) \times \left[\frac{\frac{\partial \widetilde{\pi}_j(1, V_i; \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^T}}{\widehat{\widetilde{\pi}}_j(1, V_i; \boldsymbol{\alpha})} - \frac{\frac{\partial \widetilde{\pi}_j(0, V_i; \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^T}}{\widehat{\widetilde{\pi}}_j(0, V_i; \boldsymbol{\alpha})}\right]_{\boldsymbol{\alpha}=\widehat{\boldsymbol{\alpha}}},$$

$$I_n(\widehat{\boldsymbol{\alpha}}) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{V}_i H(\widehat{\boldsymbol{\alpha}}^T\mathcal{V}_i)[1 - H(\widehat{\boldsymbol{\alpha}}^T\mathcal{V}_i)]\mathcal{V}_i^T,$$

$$\Omega_{n,k}(V_s; \widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\delta_i \mathcal{X}_{ij} H_{+,j}^{(1)}(X_i, V_i; \widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}}) \times \left[\frac{\frac{\partial \widetilde{\pi}_j(1, V_i; \widehat{\boldsymbol{\alpha}})}{\partial P_k(V_i)}}{\widehat{\widetilde{\pi}}_j(1, V_i; \widehat{\boldsymbol{\alpha}})} - \frac{\frac{\partial \widetilde{\pi}_j(0, V_i; \widehat{\boldsymbol{\alpha}})}{\partial P_k(V_i)}}{\widehat{\widetilde{\pi}}_j(0, V; \widehat{\boldsymbol{\alpha}})}\right]_{P_k(V_i)=\widehat{P}_k(V_i)} \frac{I(V_i = V_s)}{\widehat{P}(V = V_s)},$$

$$\widehat{P}(V = V_s) = \frac{1}{n}\sum_{i=1}^{n} I(V_i = V_s),$$

$$\begin{aligned} J_n(\widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}}) = {} & \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{j=1}^{r-1}\mathcal{X}_{ij}[1 - H_{+,j}(X_i, V_i; \widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}})] \times \left[\sum_{k=1}^{j} H^{(1)}(\widehat{\boldsymbol{\alpha}}^T\mathcal{V}_{k,i})P(Y_i = k|X_i, V_i)\mathcal{V}_{k,i}^T\right]\right\} \\ & -\frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{j=1}^{r-1}\mathcal{X}_{ij} H_{+,j}(X_i, V_i; \widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}}) \times \left[\sum_{k=j+1}^{r} H^{(1)}(\widehat{\boldsymbol{\alpha}}^T\mathcal{V}_{k,i})P(Y_i = k|X_i, V_i)\mathcal{V}_{k,i}^T\right]\right\}, \end{aligned}$$

$$\begin{aligned} Q_n(\widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}}) = {} & \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{j=1}^{r-1}\mathcal{X}_{ij}[1 - H_{+,j}(X_i, V_i; \widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}})] \times \left[\sum_{k=1}^{j} H(\widehat{\boldsymbol{\alpha}}^T\mathcal{V}_{k,i})\widehat{P}(Y_i = k|X_i, V_i)\Omega_{n,k}^T(V_i; \widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}})\right]\right\} \\ & - \frac{1}{n}\sum_{i=1}^{n}\left\{\sum_{j=1}^{r-1}\mathcal{X}_{ij} H_{+,j}(X_i, V_i; \widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}}) \times \left[\sum_{k=j+1}^{r} H(\widehat{\boldsymbol{\alpha}}^T\mathcal{V}_{k,i})\widehat{P}(Y_i = k|X_i, V_i)\Omega_{n,k}^T(V_i; \widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}})\right]\right\}, \end{aligned}$$

$$P_n(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r}\left\{\Omega_{n,j}(V_i;\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}})[1-\widehat{P}_j(V_i)]\widehat{P}_j(V_i)\Omega_{n,j}^T(V_i;\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}})\right\}$$
$$-2\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\sum_{k=j+1}^{r}\left\{\Omega_{n,j}^T(V_i;\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}})\widehat{P}_j(V_i)\widehat{P}_k(V_i)\Omega_{n,k}^T(V_i;\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}})\right\},$$

and

$$\widehat{P}(Y_i = k|X_i, V_i) = H\left(\widehat{\boldsymbol{\Theta}}_{vp}^T\mathcal{X}_{ik}\right) - H\left(\widehat{\boldsymbol{\Theta}}_{vp}^T\mathcal{X}_{i,k-1}\right).$$

Note that $\widehat{\boldsymbol{\alpha}} \xrightarrow{p} \boldsymbol{\alpha}$. Using the Inverse Function Theorem of Foutz (1997), along with Condition (A3) can be used to obtain that $\widehat{\boldsymbol{\Theta}}_{vp} \xrightarrow{p} \boldsymbol{\Theta}$. Hence, using the continuity $H_{+,j}(\cdot)$, we derived that $H_{+,j}(X, V; \widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}}) \xrightarrow{p} H_{+,j}(X, V; \boldsymbol{\Theta},\boldsymbol{\alpha})$. In addition, note that $G_{0n}(\cdot)$ is the sum of $iid$ random variables. Thus, we have $G_{0n}(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}}) - G_{0n}(\boldsymbol{\Theta},\boldsymbol{\alpha}) \xrightarrow{p} 0$. By weak law of large number, it can be show that $G_{0n}(\boldsymbol{\Theta},\boldsymbol{\alpha}) \xrightarrow{p} G_0(\boldsymbol{\Theta},\boldsymbol{\alpha})$. Therefore, the convergence of $G_{0n}(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}}) \xrightarrow{p} G_0(\boldsymbol{\Theta},\boldsymbol{\alpha})$ can be justified. All other convergences in probability can be proven similarly as follows:

$G_{1n}(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}}) \xrightarrow{p} G_1(\boldsymbol{\Theta},\boldsymbol{\alpha})$, $G_n(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}}) \xrightarrow{p} G(\boldsymbol{\Theta},\boldsymbol{\alpha})$, $K_n(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}}) \xrightarrow{p} K(\boldsymbol{\Theta},\boldsymbol{\alpha})$, $I_n(\widehat{\boldsymbol{\alpha}}) \xrightarrow{p} I(\boldsymbol{\alpha})$, $\Omega_{n,k}(V_s;\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}}) \xrightarrow{p} \Omega_k(V_s;\boldsymbol{\Theta},\boldsymbol{\alpha})$, $J_n(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}}) \xrightarrow{p} J(\boldsymbol{\Theta},\boldsymbol{\alpha})$, $Q_n(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}}) \xrightarrow{p} Q(\boldsymbol{\Theta},\boldsymbol{\alpha})$, and $P_n(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}}) \xrightarrow{p} P(\boldsymbol{\Theta},\boldsymbol{\alpha})$. Hence a consistent estimator of $\Delta_{vp}$ is given by

$$\widehat{\Delta}_{vp} = G_{0n}^{-1}(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}})\left\{G_n(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}})+K_n(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}})I_n^{-1}(\widehat{\boldsymbol{\alpha}})K_n^T(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}})+P_n(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}})\right.$$
$$\left.-\,2J_n(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}})I_n^{-1}(\widehat{\boldsymbol{\alpha}})K_n^T(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}}) - 2Q_n(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}})\right\}G_{0n}^{-T}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\boldsymbol{\alpha}}).$$

In the following, we analytically compare the asymptotic variances of the proposed estimators under different estimated selection probabilities. When $\boldsymbol{\alpha}$ in (8) is known, by (14) and Theorem 2, we have

$$\Delta_{sp} = G_0^{-1}(\boldsymbol{\Theta},\boldsymbol{\alpha})\left\{G(\boldsymbol{\Theta},\boldsymbol{\alpha}) + P(\boldsymbol{\Theta},\boldsymbol{\alpha}) - 2Q(\boldsymbol{\Theta},\boldsymbol{\alpha})\right\}G_0^{-T}(\boldsymbol{\Theta},\boldsymbol{\alpha}).$$

Similarly, replacing $\widetilde{\pi}_j(\cdot)$ in (4) with $\widetilde{\pi}_j(a, V_i;\boldsymbol{\alpha})$ results in

$$\Delta_t = G_0^{-1}(\boldsymbol{\Theta},\boldsymbol{\alpha})G(\boldsymbol{\Theta},\boldsymbol{\alpha})G_0^{-T}(\boldsymbol{\Theta},\boldsymbol{\alpha}).$$

Thus,

$$\Delta_t - \Delta_{sp} = G_0^{-1}(\boldsymbol{\Theta},\boldsymbol{\alpha})\left\{2Q(\boldsymbol{\Theta},\boldsymbol{\alpha}) - P(\boldsymbol{\Theta},\boldsymbol{\alpha})\right\}G_0^{-T}(\boldsymbol{\Theta},\boldsymbol{\alpha}), \qquad (15)$$

and

$$\Delta_t - \Delta_{vp} = G_0^{-1}(\boldsymbol{\Theta},\boldsymbol{\alpha})\left\{2Q(\boldsymbol{\Theta},\boldsymbol{\alpha}) - P(\boldsymbol{\Theta},\boldsymbol{\alpha}) - K(\boldsymbol{\Theta},\boldsymbol{\alpha})I^{-1}(\boldsymbol{\alpha})K^T(\boldsymbol{\Theta},\boldsymbol{\alpha}) + 2J(\boldsymbol{\Theta},\boldsymbol{\alpha})I^{-1}(\boldsymbol{\alpha})K^T(\boldsymbol{\Theta},\boldsymbol{\alpha})\right\} G_0^{-T}(\boldsymbol{\Theta},\boldsymbol{\alpha}). \tag{16}$$

Therefore, by (7) and Theorem 1, we have

$$\Delta_t - \Delta_{np} = G_0^{-1}(\boldsymbol{\Theta},\widetilde{\pi})\left\{M(\boldsymbol{\Theta},\widetilde{\pi}) + 2A(\boldsymbol{\Theta},\widetilde{\pi})\right\} G_0^{-T}(\boldsymbol{\Theta},\widetilde{\pi}). \tag{17}$$

As discussed in the case $r = 2$, we can obtain the $\Delta_t = \Delta_{sp} = G_0^{-1}(\boldsymbol{\Theta},\boldsymbol{\alpha})$. In addition, (16) and (17) are reduced to

$$\Delta_t - \Delta_{vp} = G_0^{-1}(\boldsymbol{\Theta},\boldsymbol{\alpha})K(\boldsymbol{\Theta},\boldsymbol{\alpha})I^{-1}(\boldsymbol{\alpha})K^T(\boldsymbol{\Theta},\boldsymbol{\alpha})G_0^{-T}(\boldsymbol{\Theta},\boldsymbol{\alpha}), \tag{18}$$

and

$$\Delta_t - \Delta_{np} = G_0^{-1}(\boldsymbol{\Theta},\widetilde{\pi})M_0(\boldsymbol{\Theta},\widetilde{\pi})G_0^{-T}(\boldsymbol{\Theta},\widetilde{\pi}). \tag{19}$$

Then $\Delta_t - \Delta_{vp}$ and $\Delta_t - \Delta_{np}$ in (18) and (19) are positive semidefinite, which implies that the efficiency of estimator of $\boldsymbol{\Theta}$ may be gained via data adjustment of $\pi(Y, V)$ (see Robins et al. 1994; Rosenbaum 1987 and Wang et al. 1997). Hence, when $r = 2$, the estimation that uses estimated selection probabilities leads to more efficient estimator compared to that using the true selection probabilities. Since we cannot analytically show that the left side of (15), (16) and (17) are positive semidefinite the case with $r \geq 3$, the asymptotic behavior of these proposed estimators are investigated through simulation study in the next section.

## 4 Simulation study

In this section, a simulation study is conducted to investigate the finite sample performance under various estimates of $\widetilde{\pi}_j(\cdot)$. The following four estimators are considered

- $\widehat{\boldsymbol{\Theta}}_t$: the CL estimator using the true $\widetilde{\pi}_j(\cdot)$.
- $\widehat{\boldsymbol{\Theta}}_{np}$: the CL non-parametric estimator.
- $\widehat{\boldsymbol{\Theta}}_{vp}$: the CL parametric estimator.
- $\widehat{\boldsymbol{\Theta}}_{sp}$: the CL semiparametric estimator.

The replication is 1,000 times and the sample size is chosen as $n = (200, 300)$ and $(500, 800)$, for univariate covariate and bivariate covariate, respectively. For each estimator, we computed bias, asymptotic standard error (ASE), standard deviation (SD), and 95% confidence interval coverage probabilities (CP).

*Case 1* We considered the case with one univariate covariate $X$ and $r = 2$. (i.e. logistic regression). First, the $X$s and $\varepsilon$s were generated from normal distributions $N(0, 1)$ and $N(0, \sigma)$, respectively. Given $X$, $\varepsilon$ and $\sigma = 0.25$, the $W$ was a binary surrogate variable with $W = 1$ if $X + \varepsilon \geq 0$ and $W = 0$ if $X + \varepsilon < 0$. The ordinal respones $Y$ was generated as binary with $P(Y = 0|X) = H(\theta_1 + \beta_1 X)$ and $\boldsymbol{\Theta} = (\theta_1, \beta_1)^T = (-1, \ln(2))^T$. Given $W$ and $Y$, the validation data indicator $\delta$ was a binary variable with $\pi(Y, W) = H(\alpha_0 + \alpha_1 Y + \alpha_2 W)$ and the values of $\boldsymbol{\alpha}$ were $(1, 0.5, -1)^T$ and $(-0.5, 0.5, -1)^T$, which resulted in about 22 and 53% of missing rate, respectively.

Simulation results (see Table 1) show that the efficiencies of all the estimators increase as sample size increases. The estimator $\widehat{\boldsymbol{\Theta}}_{np}$ performs slightly better than $\widehat{\boldsymbol{\Theta}}_{vp}$, and $\widehat{\boldsymbol{\Theta}}_{np}$ outperforms the other two estimators $\widehat{\boldsymbol{\Theta}}_{sp}$ and $\widehat{\boldsymbol{\Theta}}_t$. Moreover, when the missing rate is high (0.53), $\widehat{\beta}_{1_{np}}$ is more efficient than the other three estimators, $\widehat{\beta}_{1_{vp}}$, $\widehat{\beta}_{1_{sp}}$ and $\widehat{\beta}_{1_t}$. Overall, the efficiencies of all the estimators increase as the missing rate decreases. Moreover, when the sample size is small, all the proposed estimators perform poorly for large values of $\sigma$, which is the case when $W$ provides little informative about $X$.

*Case 2* The distribution of all the variables were the same as those used in case 1, except that the values of parameter $\boldsymbol{\Theta}$ were $(0, -\ln(2))^T$ and $(2, -\ln(2))^T$, which resulted in $P(Y = 0) \cong 0.5$ and $P(Y = 0) \cong 0.13$, respectively. Given $W$ and $Y$, the validation data indicator $\delta$ was a binary variable with $\pi(Y, W) = H(-0.5+0.5Y-W)$. On average, about 58% of the missing rate.

**Table 1** Simulation results of Case 1

| *mr* | | | $n = 200$ | | | | $n = 300$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\widehat{\boldsymbol{\Theta}}_{np}$ | $\widehat{\boldsymbol{\Theta}}_{vp}$ | $\widehat{\boldsymbol{\Theta}}_{sp}$ | $\widehat{\boldsymbol{\Theta}}_t$ | $\widehat{\boldsymbol{\Theta}}_{np}$ | $\widehat{\boldsymbol{\Theta}}_{vp}$ | $\widehat{\boldsymbol{\Theta}}_{sp}$ | $\widehat{\boldsymbol{\Theta}}_t$ |
| | $(\alpha_0, \alpha_1, \alpha_2) = (1, 0.5, -1)$ | | | | | | | | | |
| 22% | $\theta_1$ | Bias | −0.022 | −0.023 | −0.028 | −0.028 | −0.006 | −0.006 | −0.012 | −0.012 |
| | | SD | 0.178 | 0.180 | 0.199 | 0.199 | 0.136 | 0.136 | 0.157 | 0.157 |
| | | ASE | 0.176 | 0.172 | 0.193 | 0.200 | 0.140 | 0.137 | 0.157 | 0.162 |
| | | CP | 0.948 | 0.939 | 0.940 | 0.950 | 0.960 | 0.946 | 0.952 | 0.962 |
| | $\beta_1$ | Bias | 0.023 | 0.022 | 0.023 | 0.023 | 0.016 | 0.013 | 0.014 | 0.014 |
| | | SD | 0.204 | 0.219 | 0.222 | 0.222 | 0.157 | 0.166 | 0.168 | 0.168 |
| | | ASE | 0.196 | 0.206 | 0.209 | 0.213 | 0.159 | 0.167 | 0.169 | 0.172 |
| | | CP | 0.946 | 0.941 | 0.938 | 0.942 | 0.955 | 0.962 | 0.964 | 0.965 |
| | $(\alpha_0, \alpha_1, \alpha_2) = (-0.5, 0.5, -1)$ | | | | | | | | | |
| 53% | $\theta_1$ | Bias | −0.016 | −0.024 | −0.030 | −0.030 | −0.004 | −0.008 | −0.023 | −0.023 |
| | | SD | 0.173 | 0.176 | 0.277 | 0.277 | 0.144 | 0.146 | 0.222 | 0.222 |
| | | ASE | 0.176 | 0.172 | 0.264 | 0.268 | 0.148 | 0.145 | 0.217 | 0.221 |
| | | CP | 0.948 | 0.945 | 0.946 | 0.948 | 0.956 | 0.953 | 0.949 | 0.953 |
| | $\beta_1$ | Bias | 0.040 | 0.044 | 0.047 | 0.047 | 0.020 | 0.026 | 0.025 | 0.025 |
| | | SD | 0.381 | 0.482 | 0.487 | 0.487 | 0.201 | 0.240 | 0.242 | 0.242 |
| | | ASE | 0.370 | 0.466 | 0.470 | 0.475 | 0.192 | 0.235 | 0.237 | 0.239 |
| | | CP | 0.951 | 0.955 | 0.950 | 0.951 | 0.940 | 0.944 | 0.950 | 0.951 |

*mr* is the average missing rate over 1,000 replicates. The true parameters $\boldsymbol{\Theta} = (1, -\ln(2))^T$

**Table 2** Simulation results of Case 2 ($n = 300$)

| $mr$ | | | $P(Y=0) \cong 0.5$ | | | | $P(Y=0) \cong 0.13$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\widehat{\boldsymbol{\Theta}}_{np}$ | $\widehat{\boldsymbol{\Theta}}_{vp}$ | $\widehat{\boldsymbol{\Theta}}_{sp}$ | $\widehat{\boldsymbol{\Theta}}_{t}$ | $\widehat{\boldsymbol{\Theta}}_{np}$ | $\widehat{\boldsymbol{\Theta}}_{vp}$ | $\widehat{\boldsymbol{\Theta}}_{sp}$ | $\widehat{\boldsymbol{\Theta}}_{t}$ |
| 58% | $\theta_1$ | Bias | −0.001 | −0.002 | −0.006 | −0.006 | 0.036 | 0.046 | 0.055 | 0.055 |
| | | SD | 0.133 | 0.137 | 0.188 | 0.188 | 0.226 | 0.233 | 0.317 | 0.317 |
| | | ASE | 0.132 | 0.132 | 0.191 | 0.193 | 0.214 | 0.210 | 0.282 | 0.288 |
| | | CP | 0.948 | 0.943 | 0.962 | 0.963 | 0.947 | 0.938 | 0.929 | 0.938 |
| | $\beta_1$ | Bias | −0.016 | −0.019 | −0.021 | −0.021 | −0.034 | −0.041 | −0.045 | −0.045 |
| | | SD | 0.176 | 0.219 | 0.221 | 0.221 | 0.251 | 0.295 | 0.299 | 0.299 |
| | | ASE | 0.175 | 0.211 | 0.214 | 0.215 | 0.243 | 0.283 | 0.287 | 0.291 |
| | | CP | 0.944 | 0.943 | 0.947 | 0.952 | 0.952 | 0.957 | 0.956 | 0.959 |

$mr$ is the average missing rate over 1,000 replicates. The true parameters $\boldsymbol{\Theta} = (0, -\ln(2))^T$ and $(2, -\ln(2))^T$

Simulation results show (see Table 2) that all the estimators perform better when $P(Y=0) \cong 0.5$ compared to $P(Y=0) \cong 0.13$. In general, $\widehat{\boldsymbol{\Theta}}_{np}$ is asymptotically more efficient than $\widehat{\boldsymbol{\Theta}}_{vp}$, $\widehat{\boldsymbol{\Theta}}_{sp}$ and $\widehat{\boldsymbol{\Theta}}_{t}$. In addition, $\widehat{\boldsymbol{\Theta}}_{sp}$ and $\widehat{\boldsymbol{\Theta}}_{t}$ are asymptotically equivalent.

*Case 3* We considered a univariate covariate $X$ and $r = 3$. First, the $X$s were generated from a uniform distribution $[-1, 1]$. Given $X$, the $W$ was a binary surrogate variable with $W = 1$ if $X \geq 0$ and $W = 0$ if $X < 0$. The ordinal response $Y$ was generated as three categories with $P(Y \leq j|X) = H(\theta_j + \beta_1 X)$, where $j = 1, 2$ and $\boldsymbol{\Theta} = (-\ln 2, -\ln 2 + 1, \ln 3)^T$. Given $W$ and $Y$, the validation data indicator $\delta$ was a binary variable with $\pi(Y, W) = H(\alpha_0 + \alpha_1 Y + \alpha_2 W)$ and the values of $\boldsymbol{\alpha}$ were set at $(-0.5, 0.5, -1)^T$ and $(-0.5, 0.5, 0)^T$, which resulted in about 51 and 37% of missing rate, respectively. Note that when $\alpha_2 = 0$, there was no surrogate variable for $X$.

Simulation results (see Table 3) show that the efficiencies of all the estimators increase as the sample size increases. When the surrogate variable $W$ exists, the estimator $\widehat{\boldsymbol{\Theta}}_{np}$ performs slightly better than $\widehat{\boldsymbol{\Theta}}_{vp}$, and $\widehat{\boldsymbol{\Theta}}_{np}$ outperforms the other two estimators $\widehat{\boldsymbol{\Theta}}_{sp}$ and $\widehat{\boldsymbol{\Theta}}_{t}$. When $W$ does not exist, the selection probability only depends on $Y$. In this case, for the estimation of $\beta_1$, all the proposed estimators are very similar.

*Case 4* The distribution of all the variables were the same as those use in case 2, except that the covariate $Z$ was a binary variable with $P(Z=1) = P(Z=0) = 0.5$ and parameter $\boldsymbol{\Theta} = (-\ln(2), -\ln(2) + 1, 0.7, \ln(3))^T$. The binary variable $\delta$ was generated as $\pi(Y, W) = H(\alpha_0 + \alpha_1 Y + \alpha_2 W + \alpha_3 Z)$, where the values of $\boldsymbol{\alpha}$ were set at $(-0.5, 0.5, 0.5, -1)^T$ and $(-0.5, 0.5, 0, -1)^T$. The settings resulted in about 54 and 51% of missing rate, respectively.

Simulation results (see Table 4) show that the efficiencies of all the estimators increase as sample size increases. When surrogate variable $W$ exists, the estimator $\widehat{\boldsymbol{\Theta}}_{np}$ is more efficient than the other three estimators $\widehat{\boldsymbol{\Theta}}_{vp}$, $\widehat{\boldsymbol{\Theta}}_{sp}$ and $\widehat{\boldsymbol{\Theta}}_{t}$. When $W$ does not exist, the selection probability only depends on $Y$ and $Z$. In this case, specially

**Table 3** Simulation results of Case 3 (with univariate covariate $X$)

| $mr$ | | | $n=300$ | | | | $n=600$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\widehat{\boldsymbol{\Theta}}_{np}$ | $\widehat{\boldsymbol{\Theta}}_{vp}$ | $\widehat{\boldsymbol{\Theta}}_{sp}$ | $\widehat{\boldsymbol{\Theta}}_{t}$ | $\widehat{\boldsymbol{\Theta}}_{np}$ | $\widehat{\boldsymbol{\Theta}}_{vp}$ | $\widehat{\boldsymbol{\Theta}}_{sp}$ | $\widehat{\boldsymbol{\Theta}}_{t}$ |
| | $(\alpha_0, \alpha_1, \alpha_2) = (-0.5, 0.5, -1)$ | | | | | | | | | |
| 51% | $\theta_1$ | Bias | −0.003 | −0.008 | −0.017 | −0.020 | −0.004 | −0.006 | −0.011 | −0.014 |
| | | SD | 0.135 | 0.150 | 0.198 | 0.199 | 0.096 | 0.107 | 0.143 | 0.144 |
| | | ASE | 0.134 | 0.145 | 0.196 | 0.200 | 0.094 | 0.103 | 0.138 | 0.140 |
| | | CP | 0.953 | 0.955 | 0.949 | 0.952 | 0.938 | 0.937 | 0.945 | 0.948 |
| | $\theta_2$ | Bias | 0.009 | 0.009 | 0.005 | 0.004 | −0.003 | −0.002 | −0.005 | −0.005 |
| | | SD | 0.122 | 0.134 | 0.172 | 0.172 | 0.088 | 0.097 | 0.122 | 0.122 |
| | | ASE | 0.126 | 0.133 | 0.171 | 0.173 | 0.089 | 0.095 | 0.120 | 0.122 |
| | | CP | 0.959 | 0.956 | 0.949 | 0.957 | 0.956 | 0.944 | 0.951 | 0.953 |
| | $\beta_1$ | Bias | 0.018 | 0.014 | 0.015 | 0.017 | 0.009 | 0.012 | 0.012 | 0.015 |
| | | SD | 0.242 | 0.298 | 0.303 | 0.306 | 0.159 | 0.209 | 0.212 | 0.213 |
| | | ASE | 0.234 | 0.290 | 0.295 | 0.298 | 0.164 | 0.204 | 0.207 | 0.209 |
| | | CP | 0.941 | 0.950 | 0.947 | 0.947 | 0.959 | 0.941 | 0.950 | 0.950 |
| | $(\alpha_0, \alpha_1, \alpha_2) = (-0.5, 0.5, 0)$ | | | | | | | | | |
| 37% | $\theta_1$ | Bias | −0.010 | −0.013 | −0.016 | −0.031 | 0.001 | 0.003 | 0.000 | −0.014 |
| | | SD | 0.134 | 0.145 | 0.177 | 0.179 | 0.094 | 0.099 | 0.120 | 0.121 |
| | | ASE | 0.133 | 0.137 | 0.169 | 0.173 | 0.094 | 0.097 | 0.119 | 0.121 |
| | | CP | 0.948 | 0.935 | 0.941 | 0.943 | 0.947 | 0.935 | 0.945 | 0.950 |
| | $\theta_2$ | Bias | −0.006 | −0.004 | −0.004 | 0.001 | 0.004 | 0.002 | 0.002 | 0.007 |
| | | SD | 0.127 | 0.132 | 0.153 | 0.154 | 0.092 | 0.097 | 0.113 | 0.113 |
| | | ASE | 0.127 | 0.146 | 0.151 | 0.154 | 0.089 | 0.103 | 0.107 | 0.108 |
| | | CP | 0.953 | 0.974 | 0.958 | 0.963 | 0.945 | 0.965 | 0.941 | 0.945 |
| | $\beta_1$ | Bias | 0.003 | 0.003 | 0.003 | 0.008 | −0.006 | −0.006 | −0.006 | −0.001 |
| | | SD | 0.251 | 0.251 | 0.251 | 0.252 | 0.190 | 0.190 | 0.190 | 0.191 |
| | | ASE | 0.261 | 0.261 | 0.261 | 0.261 | 0.183 | 0.183 | 0.183 | 0.183 |
| | | CP | 0.962 | 0.962 | 0.962 | 0.960 | 0.945 | 0.945 | 0.945 | 0.944 |

$mr$ is the average missing rate over 1,000 replicates. The true parameters $\boldsymbol{\Theta} = (-\ln 2, -\ln 2 + 1, \ln 3)^T$. There was no surrogate variable $W$, which means $\alpha_2 = 0$

for the estimation of $\beta_1$, all the proposed estimators are close to one another. For the estimation of $\beta_2$, the $\widehat{\beta}_{2_{np}}$ outperforms the other three estimators $\widehat{\beta}_{2_{vp}}$, $\widehat{\beta}_{2_{sp}}$ and $\widehat{\beta}_{2_t}$. All the proposed estimators perform better when surrogate variable $W$ exists compared to the case when $W$ does not exist.

In summary, based on Tables 3 and 4, we concluded that $\widehat{\boldsymbol{\Theta}}_{np}$ is asymptotically more efficient than $\widehat{\boldsymbol{\Theta}}_{vp}$, $\widehat{\boldsymbol{\Theta}}_{sp}$ and $\widehat{\boldsymbol{\Theta}}_{t}$. Moreover, the $\widehat{\boldsymbol{\Theta}}_{vp}$ is asymptotically more efficient than $\widehat{\boldsymbol{\Theta}}_{sp}$ and $\widehat{\boldsymbol{\Theta}}_{t}$. In addition, the efficiencies of all the estimators are impacted by the values of sample size and missing rate. However, when the missing rate is high and sample size is small, we are not able to obtain enough information for estimating parameters $\boldsymbol{\Theta}$. In this case, the estimation procedure can fail to converge.

**Table 4** Simulation results of Case 4 (with bivariate covariate $X$ and $Z$)

| $mr$ | | | $n = 500$ | | | | $n = 800$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $\widehat{\boldsymbol{\Theta}}_{np}$ | $\widehat{\boldsymbol{\Theta}}_{vp}$ | $\widehat{\boldsymbol{\Theta}}_{sp}$ | $\widehat{\boldsymbol{\Theta}}_{t}$ | $\widehat{\boldsymbol{\Theta}}_{np}$ | $\widehat{\boldsymbol{\Theta}}_{vp}$ | $\widehat{\boldsymbol{\Theta}}_{sp}$ | $\widehat{\boldsymbol{\Theta}}_{t}$ |
| | $(\alpha_0, \alpha_1, \alpha_2, \alpha_3) = (-0.5, 0.5, 0.5, -1)$ | | | | | | | | | |
| 54% | $\theta_1$ | Bias | −0.004 | −0.009 | −0.012 | −0.012 | 0.003 | 0.002 | −0.004 | −0.004 |
| | | SD | 0.135 | 0.154 | 0.167 | 0.167 | 0.104 | 0.117 | 0.130 | 0.130 |
| | | ASE | 0.131 | 0.148 | 0.164 | 0.167 | 0.103 | 0.117 | 0.129 | 0.131 |
| | | CP | 0.954 | 0.944 | 0.949 | 0.953 | 0.948 | 0.948 | 0.945 | 0.946 |
| | $\theta_2$ | Bias | 0.001 | −0.001 | −0.002 | −0.003 | 0.007 | 0.009 | 0.005 | 0.004 |
| | | SD | 0.129 | 0.147 | 0.159 | 0.159 | 0.099 | 0.110 | 0.121 | 0.121 |
| | | ASE | 0.127 | 0.139 | 0.152 | 0.154 | 0.100 | 0.110 | 0.120 | 0.122 |
| | | CP | 0.955 | 0.936 | 0.943 | 0.948 | 0.960 | 0.954 | 0.951 | 0.954 |
| | $\beta_1$ | Bias | 0.004 | 0.009 | 0.010 | 0.013 | 0.000 | −0.002 | −0.001 | 0.002 |
| | | SD | 0.172 | 0.208 | 0.210 | 0.212 | 0.136 | 0.161 | 0.164 | 0.165 |
| | | ASE | 0.174 | 0.204 | 0.207 | 0.209 | 0.137 | 0.161 | 0.163 | 0.165 |
| | | CP | 0.947 | 0.948 | 0.951 | 0.946 | 0.957 | 0.953 | 0.949 | 0.949 |
| | $\beta_2$ | Bias | 0.009 | 0.019 | 0.018 | 0.017 | 0.004 | 0.005 | 0.005 | 0.004 |
| | | SD | 0.180 | 0.244 | 0.249 | 0.251 | 0.146 | 0.187 | 0.191 | 0.192 |
| | | ASE | 0.181 | 0.238 | 0.244 | 0.248 | 0.143 | 0.188 | 0.193 | 0.195 |
| | | CP | 0.957 | 0.947 | 0.945 | 0.950 | 0.955 | 0.954 | 0.956 | 0.956 |
| | $(\alpha_0, \alpha_1, \alpha_2, \alpha_3) = (-0.5, 0.5, 0, -1)$ | | | | | | | | | |
| 51% | $\theta_1$ | Bias | −0.004 | −0.004 | −0.004 | −0.014 | −0.003 | −0.002 | −0.004 | −0.014 |
| | | SD | 0.132 | 0.157 | 0.175 | 0.177 | 0.106 | 0.125 | 0.140 | 0.141 |
| | | ASE | 0.133 | 0.153 | 0.172 | 0.175 | 0.105 | 0.121 | 0.136 | 0.138 |
| | | CP | 0.948 | 0.945 | 0.948 | 0.950 | 0.943 | 0.951 | 0.939 | 0.942 |
| | $\theta_2$ | Bias | 0.005 | 0.004 | 0.006 | 0.011 | 0.002 | 0.003 | 0.002 | 0.007 |
| | | SD | 0.131 | 0.148 | 0.163 | 0.164 | 0.100 | 0.115 | 0.128 | 0.129 |
| | | ASE | 0.129 | 0.148 | 0.158 | 0.161 | 0.101 | 0.118 | 0.125 | 0.127 |
| | | CP | 0.940 | 0.951 | 0.933 | 0.939 | 0.955 | 0.946 | 0.947 | 0.943 |
| | $\beta_1$ | Bias | 0.013 | 0.013 | 0.013 | 0.016 | 0.017 | 0.017 | 0.017 | 0.020 |
| | | SD | 0.221 | 0.221 | 0.221 | 0.222 | 0.172 | 0.172 | 0.172 | 0.173 |
| | | ASE | 0.220 | 0.220 | 0.220 | 0.220 | 0.174 | 0.174 | 0.174 | 0.174 |
| | | CP | 0.944 | 0.944 | 0.944 | 0.943 | 0.951 | 0.951 | 0.951 | 0.952 |
| | $\beta_2$ | Bias | 0.002 | 0.003 | 0.005 | 0.005 | 0.005 | 0.001 | 0.002 | 0.003 |
| | | SD | 0.177 | 0.250 | 0.257 | 0.261 | 0.146 | 0.203 | 0.208 | 0.211 |
| | | ASE | 0.185 | 0.249 | 0.259 | 0.263 | 0.145 | 0.197 | 0.205 | 0.207 |
| | | CP | 0.958 | 0.943 | 0.950 | 0.950 | 0.950 | 0.946 | 0.955 | 0.953 |

$mr$ is the average missing rate over 1,000 replicates. The true parameters $\boldsymbol{\Theta} = (-\ln 2, -\ln 2 + 1, 0.7, \ln 3)^T$. There was no surrogate variable $W$, which means $\alpha_2 = 0$

## 5 Example

To illustrate the proposed approaches, we analyze the data set from a survey of cable TV satisfaction in Taiwan. The ordinal response $Y$ is the satisfaction level of cable TV channels with three categories (1. Satisfied; 2. Acceptable; 3. Unacceptable). The covariate $X$ is the response (1. Yes; 2. No) of the following question:

"Have you been given a discount on cable TV?"

In this study, due to item non-response, $X$ is not available for some subjects. The surrogate variable $W$ is the response (1. Yes; 2. No) of the following question:

"Do you want to pay extra expense for additional channels?"

The covariate $Z$ is the three cities in Taiwan (1. Taipei; 2. Taichung; 3. Yunlin). It can be written in vector notation as $\beta_2 DZ_1 + \beta_3 DZ_2$, where $DZ_i (i = 1, 2)$ are dummy variables whose components take the value 1 if the unit $Z$ has level $i$, and zero otherwise. A random sample of 2000 subjects are obtained from three cities and there are 1566 subjects in the validation data set.

To verify if the missingness mechanism is $MAR$, we conducted a logistic regression analysis with outcome variables $\delta_i$, $i = 1, \ldots, 2000$, and five covariates ($DY_1$, $DY_2$, $W$, $DZ_1$, $DZ_2$), where $DY_i (i = 1, 2)$ are dummy variables whose components take the value 1 if the unit $Y$ have level $i$, and zero otherwise. The parameter estimates for these five covariates were equal to (−0.017, 0.437, 0.332, 0.326, 0.849), with corresponding standard errors (SE) (0.148, 0.114, 0.114, 0.143, 0.099). The missingness mechanism only depends on the joint distributions of $Y$, $W$ and $Z$ (i.e. MAR).

We examine the performance of the CL estimator using non-parametric and parametric estimates of selection probabilities. The estimator $\widehat{\beta}_3$ is not significantly different from zero, which means that there is no different between Taichung and Yunlin cities in terms of channel customer satisfaction. Hence, we redefine the covariate $Z$ by combining the two cities as (1. Taipei; 2. Taichung or Yunlin). Table 5 shows the estimates and the corresponding estimated standard errors (SE) of the estimators. For the estimation of $\theta_1$, $\theta_2$, $\beta_1$ and $\beta_2$, all the estimators show significant results. The two proposed methods are close to one another, expect that estimator $\widehat{\theta}_{2_{vp}}$ is larger than $\widehat{\theta}_{2_{np}}$. The estimated SE of the CL non-parametric estimator is smaller than the CL parametric estimator, which is consistent with simulation results in Sect. 4.

**Table 5** Results of the cable TV satisfaction data

| | $\widehat{\boldsymbol{\Theta}}_{np}$ | $\widehat{\boldsymbol{\Theta}}_{vp}$ |
|---|---|---|
| $\theta_1$ | −1.6755 | −1.6643 |
| (SE) | (0.0697) | (0.0744) |
| $\theta_2$ | 0.4583 | 0.6173 |
| (SE) | (0.0561) | (0.0597) |
| $\beta_1$ | 0.1619 | 0.1624 |
| (SE) | (0.0577) | (0.0578) |
| $\beta_2$ | 0.2180 | 0.2127 |
| (SE) | (0.0467) | (0.0532) |

## 6 Discussion

We have developed CL approaches for the estimation of the proportional odds model with missing covariates. Specifically, we have proposed the CL estimators using different estimates of the selection probabilities. Based on theoretical analysis and simulation studies, we conclude that the proposed estimator $\widehat{\boldsymbol{\Theta}}_{np}$ outperforms all the other estimators, $\widehat{\boldsymbol{\Theta}}_{vp}$, $\widehat{\boldsymbol{\Theta}}_{sp}$ and $\widehat{\boldsymbol{\Theta}}_{t}$. Moreover, the estimator $\widehat{\boldsymbol{\Theta}}_{vp}$ is more efficient than both $\widehat{\boldsymbol{\Theta}}_{sp}$ and $\widehat{\boldsymbol{\Theta}}_{t}$ estimators. In our approaches, the CL non-parametric estimators do not require assumptions either for the selection probability or for the conditional probability of $Y$ given $V$.

Although the main results were presented for the case where $V$ is discrete, extension to the continuous case involves the estimation of density $\widetilde{\pi}_j(a, V)$ and $P_k(V)$, which can be obtained by kernel methods (e.g. see Wang et al. 1997). Furthermore, we have performed some simulation studies to investigate this case. The simulation results show that $\widehat{\boldsymbol{\Theta}}_{np}$ outperforms all the others proposed estimators. The same behavior is observed in the case where $V$ is discrete. Further research is required to derive the asymptotic properties of the kernel-based estimators.

In this paper, our focus is primarily on efficiency comparisons of several conditional likelihood estimators based on the validation set. We have studied the joint conditional likelihood estimator of Wang et al. (2002) in the proportional odds model, and our preliminary results showed that it is a promising estimator. The detailed work of the joint conditional likelihood estimator will be presented elsewhere in the future. Another possible extension is to apply our approaches to the adjacent category logistic model with missing covariates.

**Acknowledgments** The research of Sing-Ming Lee was supported by the National Science Council grant of Taiwan, ROC, 97-2118-M-035-002-MY2. The authors thank associate editor and two anonymous referees for their valuable comments which greatly improved the presentation of this paper.

## Appendix

*Proof of Lemma 1* By Taylor expansion of $\widehat{U}_{1n}(\boldsymbol{\Theta}, \widehat{\widetilde{\pi}})$ at $\widetilde{\pi}_j(0, V_i)$ and $\widetilde{\pi}_j(1, V_i)$, we have

$$\begin{aligned}
&\widehat{U}_{1n}(\boldsymbol{\Theta}, \widehat{\widetilde{\pi}}) - U_{1n}(\boldsymbol{\Theta}, \widetilde{\pi}) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{r-1} \left\{ \delta_i \mathcal{X}_{ij} [H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi}) - H_{+j}(X_i, V_i; \boldsymbol{\Theta}, \widehat{\widetilde{\pi}})] \right\} \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{r-1} \delta_i \mathcal{X}_{ij} H^{(1)}_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi}) \left\{ -\frac{\widehat{\widetilde{\pi}}_j(1, V_i) - \widetilde{\pi}_j(1, V_i)}{\widetilde{\pi}_j(1, V_i)} \right. \\
&\qquad + \frac{\widehat{\widetilde{\pi}}_j(0, V_i) - \widetilde{\pi}_j(0, V_i)}{\widetilde{\pi}_j(0, V_i)} + o_p(\widehat{\widetilde{\pi}}_j(1, V_i) - \widetilde{\pi}_j(1, V_i)) \\
&\qquad \left. + o_p(\widehat{\widetilde{\pi}}_j(0, V_i) - \widetilde{\pi}_j(0, V_i)) \right\}
\end{aligned}$$

$$= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\delta_i\mathcal{X}_{ij}H^{(1)}_{+,j}(X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})\left\{\frac{\widehat{\widetilde{\pi}}_j(1,V_i)-\widetilde{\pi}_j(1,V_i)}{\widetilde{\pi}_j(1,V_i)}\right\}$$
$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\delta_i\mathcal{X}_{ij}H^{(1)}_{+}(X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})\left\{\frac{\widehat{\widetilde{\pi}}_j(0,V_i)-\widetilde{\pi}_j(0,V_i)}{\widetilde{\pi}_j(0,V_i)}\right\}+o_p(1)$$
$$= B_{1n}(\boldsymbol{\Theta})+B_{2n}(\boldsymbol{\Theta})+o_p(1).$$

Now,

$$B_{1n}(\boldsymbol{\Theta}) = -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\delta_i\mathcal{X}_{ij}H^{(1)}_{+,j}(X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})\left\{\frac{\widehat{\widetilde{\pi}}_j(1,V_i)-\widetilde{\pi}_j(1,V_i)}{\widetilde{\pi}_j(1,V_i)}\right\}$$
$$= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\delta_i\mathcal{X}_{ij}H^{(1)}_{+,j}(X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})$$
$$\times\left\{\frac{\sum_{k=1}^{n}[\delta_k-\widetilde{\pi}_j(1,V_i)]I(T_{kj}=1,V_k=V_i)}{\widetilde{\pi}_j(1,V_i)\sum_{s=1}^{n}I(T_{sj}=1,V_s=V_i)}\right\}$$
$$= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\sum_{j=1}^{r-1}\frac{[\delta_k-\widetilde{\pi}_j(1,V_k)]I(T_{kj}=1)}{\widetilde{\pi}_j(1,V_k)P(T_{1j}=1|V=V_k)}$$
$$\times\left\{\frac{\frac{1}{n}\sum_{i=1}^{n}\delta_i\mathcal{X}_{ij}H^{(1)}_{+,j}(X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})I(V_i=V_k)}{P(V=V_k)}\right\}+o_p(1)$$
$$= -\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\sum_{j=1}^{r-1}\frac{[\delta_k-\widetilde{\pi}_j(1,V_k)]I(T_{kj}=1)}{\widetilde{\pi}_j(1,V_k)P(T_{1j}=1|V=V_k)}E^*_j(V_k)+o_p(1),$$

since

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\delta_i\mathcal{X}_{ij}H^{(1)}_{+,j}(X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})I(V_i=V_k)}{P(V=V_k)}\xrightarrow{p}\mathrm{E}\left\{\delta_1\mathcal{X}_{1j}H^{(1)}_{+,j}(X_1,V_1;\boldsymbol{\Theta})|V_1=V_k\right\}$$
$$\equiv E^*_j(V_k).$$

Similarly,

$$B_{2n}(\boldsymbol{\Theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\delta_i\mathcal{X}_{ij}H^{(1)}_{+,j}(X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})-\left\{\frac{\widehat{\widetilde{\pi}}_j(0,V_i)-\widetilde{\pi}_j(0,V_i)}{\widetilde{\pi}_j(0,V_i)}\right\}$$
$$= \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\sum_{j=1}^{r-1}\frac{[\delta_k-\widetilde{\pi}_j(0,V_k)]I(T_{kj}=0)}{\widetilde{\pi}_j(0,V_k)P(T_{1j}=0|V=V_k)}E^*_j(V_k)+o_p(1).$$

Hence,

$$
\begin{aligned}
&\widehat{U}_{1n}(\boldsymbol{\Theta},\widehat{\widetilde{\pi}})\\
&\quad = U_{1n}(\boldsymbol{\Theta},\widetilde{\pi}) - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{r-1} E_j^*(V_i)\left\{\frac{[\delta_i - \widetilde{\pi}_j(1,V_i)]I(T_{ij}=1)}{\widetilde{\pi}_j(1,V_i)P(T_{1j}=1|V=V_i)}\right.\\
&\qquad \left. - \frac{[\delta_i - \widetilde{\pi}_j(0,V_i)]I(T_{ij}=0)}{\widetilde{\pi}_j(0,V_i)P(T_{1j}=0|V=V_i)}\right\} + o_p(1)\\
&\quad = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\sum_{j=1}^{r-1}\left\{S_j(T_{ij},X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi}) - \varepsilon_j(T_{ij},V_i;\boldsymbol{\Theta},\widetilde{\pi})\right\} + o_p(1),
\end{aligned}
$$

where $S_j(T_{ij}, X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi})$ and $\varepsilon_j(T_{ij}, V_i; \boldsymbol{\Theta}, \widetilde{\pi})$ are defined in Lemma 1. □

*Proof of Theorem 1* First, by direct calculation we have,

$$
\begin{aligned}
&\mathrm{E}\left\{\delta_i\delta_s\mathcal{X}_{ij}[T_{ij} - H_{+,j}(X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})][T_{sk} - H_{+,k}(X_s,V_s;\boldsymbol{\Theta},\widetilde{\pi})]\mathcal{X}_{sk}^T\right\}\\
&\quad = \mathrm{E}\left\{\mathrm{E}\left\{\delta_i\delta_s\mathcal{X}_{ij}[T_{ij} - H_{+,j}(X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})]\right.\right.\\
&\qquad \left.\left.\times[T_{sk} - H_{+,k}(X_s,V_s;\boldsymbol{\Theta},\widetilde{\pi})]\mathcal{X}_{sk}^T\right| X_i,V_i,X_s,V_s\right\}\right\}\\
&\quad = \begin{cases}
\mathrm{E}\left[\delta_1\mathcal{X}_{1j}H_{+,j}^{(1)}(X_1,V_1;\boldsymbol{\Theta},\widetilde{\pi})\mathcal{X}_{1j}^T\right], & \text{if } i=s,\ j=k\\
\mathrm{E}\left[\delta_1\mathcal{X}_{1j}H_{+,j}(X_1,V_1;\boldsymbol{\Theta},\widetilde{\pi})[1 - H_{+,k}(X_1,V_1;\boldsymbol{\Theta},\widetilde{\pi})]\mathcal{X}_{1k}^T\right], & \text{if } i=s,\ j<k\\
\mathrm{E}\left[\delta_1\mathcal{X}_{1j}H_{+,k}(X_1,V_1;\boldsymbol{\Theta},\widetilde{\pi})[1 - H_{+,j}(X_1,V_1;\boldsymbol{\Theta},\widetilde{\pi})]\mathcal{X}_{1k}^T\right], & \text{if } i=s,\ k<j\\
0, & \text{if } i\neq s.
\end{cases}
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
&\mathrm{Var}\left\{\sum_{j=1}^{r-1} S_j(T_{ij},X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})\right\}\\
&\quad = \sum_{j=1}^{r-1}\mathrm{Var}\left\{S_j(T_{ij},X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})\right\}\\
&\qquad + \sum_{j=1}^{r-2}\sum_{k=j+1}^{r-1}\mathrm{Cov}\left\{S_j(T_{ij},X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi}), S_k(T_{ik},X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})\right\}\\
&\qquad + \sum_{k=1}^{r-2}\sum_{j=k+1}^{r-1}\mathrm{Cov}\left\{S_j(T_{ij},X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi}), S_k(T_{ik},X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \mathrm{E}\left[\sum_{j=1}^{r-1} \delta_1 \mathcal{X}_{1j} H_{+,j}^{(1)}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi}) \mathcal{X}_{1j}^T\right] \\
&\quad + \mathrm{E}\left[\sum_{j=1}^{r-2} \sum_{k=j+1}^{r-1} \delta_1 \mathcal{X}_{1j} H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})[1 - H_{+,k}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})] \mathcal{X}_{1k}^T\right] \\
&\quad + \mathrm{E}\left[\sum_{k=1}^{r-2} \sum_{j=k+1}^{r-1} \delta_1 \mathcal{X}_{1j} H_{+,k}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})[1 - H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})] \mathcal{X}_{1k}^T\right] \\
&= G_0(\boldsymbol{\Theta}, \widetilde{\pi}) + G_1(\boldsymbol{\Theta}, \widetilde{\pi}) + G_1^T(\boldsymbol{\Theta}, \widetilde{\pi}) \equiv G(\boldsymbol{\Theta}, \widetilde{\pi}),
\end{aligned}
$$

where $G_0(\boldsymbol{\Theta}, \widetilde{\pi})$, $G_1(\boldsymbol{\Theta}, \widetilde{\pi})$ and $G(\boldsymbol{\Theta}, \widetilde{\pi})$ are defined in Theorem 1.

Second, we get

$$
\begin{aligned}
&\mathrm{E}\left\{E_j^*(V_i) \frac{[\delta_i - \widetilde{\pi}_j(1, V_i)] I(T_{ij} = 1)}{\widetilde{\pi}_j(1, V_i) P(T_{1j} = 1 | V = V_i)} \frac{[\delta_s - \widetilde{\pi}_k(1, V_s)] I(T_{sk} = 1)}{\widetilde{\pi}_k(1, V_s) P(T_{1k} = 1 | V = V_s)} E_k^{*T}(V_s)\right\} \\
&\quad = \mathrm{E}\left\{\mathrm{E}\left\{E_j^*(V_i) \frac{[\delta_i \delta_s - \widetilde{\pi}_j(1, V_i) \widetilde{\pi}_k(1, V_s)] I(T_{ij} = 1, T_{sk} = 1)}{\widetilde{\pi}_j(1, V_i) P(T_{1j} = 1 | V = V_i) \widetilde{\pi}_k(1, V_s) P(T_{1k} = 1 | V = V_s)}\right.\right. \\
&\qquad \left.\left. \times E_k^{*T}(V_s) \middle| V_i, V_s, T_{sk} = 1, T_{ij} = 1\right\}\right\} \\
&\quad = \begin{cases}
\mathrm{E}\left[\mathrm{E}_j^*(V_1) \frac{1 - \widetilde{\pi}_j(1, V_1)}{\widetilde{\pi}_j(1, V_1) P(T_{1j} = 1 | V = V_1)} \mathrm{E}_j^{*T}(V_1)\right], & \text{if } i = s, j = k \\
\mathrm{E}\left[\mathrm{E}_j^*(V_1) \frac{1 - \widetilde{\pi}_k(1, V_1)}{\widetilde{\pi}_k(1, V_1) P(T_{1k} = 1 | V = V_1)} \mathrm{E}_k^{*T}(V_1)\right], & \text{if } i = s, j < k \\
\mathrm{E}\left[\mathrm{E}_j^*(V_1) \frac{1 - \widetilde{\pi}_j(1, V_1)}{\widetilde{\pi}_j(1, V_1) P(T_{1j} = 1 | V = V_1)} \mathrm{E}_k^{*T}(V_1)\right], & \text{if } i = s, k < j \\
0, & \text{if } i \neq s
\end{cases}
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{E}\left\{E_j^*(V_i) \frac{[\delta_i - \widetilde{\pi}_j(0, V_i)] I(T_{ij} = 0)}{\widetilde{\pi}_j(0, V_i) P(T_{1j} = 0 | V = V_i)} \frac{[\delta_s - \widetilde{\pi}_k(0, V_s)] I(T_{sk} = 0)}{\widetilde{\pi}_k(0, V_s) P(T_{1k} = 0 | V = V_s)} E_k^{*T}(V_s)\right\} \\
&\quad = \mathrm{E}\left\{\mathrm{E}\left\{E_j^*(V_i) \frac{[\delta_i \delta_s - \widetilde{\pi}_j(0, V_i) \widetilde{\pi}_k(0, V_s)] I(T_{ij} = 0, T_{sk} = 0)}{\widetilde{\pi}_j(0, V_i) P(T_{1j} = 0 | V = V_i) \widetilde{\pi}_k(0, V_s) P(T_{1k} = 0 | V = V_s)}\right.\right. \\
&\qquad \left.\left. \times E_k^{*T}(V_s) \middle| V_i, V_s, T_{sk} = 0, T_{ij} = 0\right\}\right\} \\
&\quad = \begin{cases}
\mathrm{E}\left[\mathrm{E}_j^*(V_1) \frac{1 - \widetilde{\pi}_j(0, V_1)}{\widetilde{\pi}_j(0, V_1) P(T_{1j} = 0 | V = V_1)} \mathrm{E}_j^{*T}(V_1)\right], & \text{if } i = s, j = k \\
\mathrm{E}\left[\mathrm{E}_j^*(V_1) \frac{1 - \widetilde{\pi}_j(0, V_1)}{\widetilde{\pi}_j(0, V_1) P(T_{1j} = 0 | V = V_1)} \mathrm{E}_k^{*T}(V_1)\right], & \text{if } i = s, j < k \\
\mathrm{E}\left[\mathrm{E}_j^*(V_1) \frac{1 - \widetilde{\pi}_k(0, V_1)}{\widetilde{\pi}_k(0, V_1) P(T_{1k} = 0 | V = V_1)} \mathrm{E}_k^{*T}(V_1)\right], & \text{if } i = s, k < j \\
0, & \text{if } i \neq s.
\end{cases}
\end{aligned}
$$

Hence, we have

$$\operatorname{Var}\left\{\sum_{j=1}^{r-1}\varepsilon_j(T_{ij},V_i;\boldsymbol{\Theta},\widetilde{\pi})\right\}$$

$$=\sum_{j=1}^{r-1}\operatorname{Var}\left\{\varepsilon_j(T_{ij},V_i;\boldsymbol{\Theta},\widetilde{\pi})\right\}$$

$$+\sum_{j=1}^{r-2}\sum_{k=j+1}^{r-1}\operatorname{Cov}\left\{\varepsilon_j(T_{ij},V_i;\boldsymbol{\Theta},\widetilde{\pi}),\varepsilon_k(T_{ik},V_i;\boldsymbol{\Theta},\widetilde{\pi})\right\}$$

$$+\sum_{k=1}^{r-2}\sum_{j=k+1}^{r-1}\operatorname{Cov}\left\{\varepsilon_j(T_{ij},V_i;\boldsymbol{\Theta},\widetilde{\pi}),\varepsilon_k(T_{ik},V_i;\boldsymbol{\Theta},\widetilde{\pi})\right\}$$

$$=\sum_{j=1}^{r-1}\mathrm{E}\left\{\mathrm{E}_j^*(V_1)\left[\frac{1-\widetilde{\pi}_j(1,V_1)}{\widetilde{\pi}_j(1,V_1)P(T_{1j}=1|V=V_1)}\right.\right.$$

$$\left.\left.+\ \frac{1-\widetilde{\pi}_j(0,V_1)}{\widetilde{\pi}_j(0,V_i)P(T_{1j}=0|V=V_1)}\right]\mathrm{E}_j^{*^T}(V_1)\right\}$$

$$+\sum_{j=1}^{r-2}\sum_{k=j+1}^{r-1}\mathrm{E}\left\{\mathrm{E}_j^*(V_1)\left[\frac{1-\widetilde{\pi}_k(1,V_1)}{\widetilde{\pi}_k(1,V_1)P(T_{1k}=1|V=V_1)}\right.\right.$$

$$\left.\left.+\ \frac{1-\widetilde{\pi}_j(0,V_1)}{\widetilde{\pi}_j(0,V_1)P(T_{1j}=0|V=V_1)}\right]\mathrm{E}_k^{*^T}(V_1)\right\}$$

$$+\sum_{k=1}^{r-2}\sum_{j=k+1}^{r-1}\mathrm{E}\left\{\mathrm{E}_j^*(V_1)\left[\frac{1-\widetilde{\pi}_j(1,V_1)}{\widetilde{\pi}_j(1,V_1)P(T_{1j}=1|V=V_1)}\right.\right.$$

$$\left.\left.+\ \frac{1-\widetilde{\pi}_k(0,V_1)}{\widetilde{\pi}_k(0,V_1)P(T_{1k}=0|V=V_1)}\right]\mathrm{E}_k^{*^T}(V_1)\right\}$$

$$=\sum_{j=1}^{r-1}\mathrm{E}\left\{\mathrm{E}_j^*(V_1)\left[\frac{1-\widetilde{\pi}_j(1,V_1)}{\widetilde{\pi}_j(1,V_1)P(T_{1j}=1|V=V_1)}\right.\right.$$

$$\left.\left.+\ \frac{1-\widetilde{\pi}_j(0,V_1)}{\widetilde{\pi}_j(0,V_1)P(T_{1j}=0|V=V_1)}\right]\mathrm{E}_j^{*^T}(V_1)\right\}$$

$$+\sum_{j=1}^{r-2}\sum_{k=j+1}^{r-1}\mathrm{E}\left\{\mathrm{E}_j^*(V_1)\left[\frac{1-\widetilde{\pi}_k(1,V_1)}{\widetilde{\pi}_k(1,V_1)P(T_{1k}=1|V=V_1)}\right.\right.$$

$$
\begin{aligned}
&+ \frac{1-\widetilde{\pi}_k(0,V_1)}{\widetilde{\pi}_k(0,V_1)P(T_{1k}=0|V=V_1)}\Bigg]\mathrm{E}_k^{*^T}(V_1)\Bigg\} \\
&+\sum_{j=1}^{r-2}\sum_{k=j+1}^{r-1}\mathrm{E}\Bigg\{\mathrm{E}_j^*(V_1)\Bigg[\frac{1-\widetilde{\pi}_j(0,V_1)}{\widetilde{\pi}_j(0,V_1)P(T_{1j}=0|V=V_1)} \\
&- \frac{1-\widetilde{\pi}_k(0,V_1)}{\widetilde{\pi}_k(0,V_1)P(T_{1k}=0|V=V_1)}\Bigg]\mathrm{E}_k^{*T}(V_1)\Bigg\} \\
&+\sum_{k=1}^{r-2}\sum_{j=k+1}^{r-1}\mathrm{E}\Bigg\{\mathrm{E}_j^*(V_1)\Bigg[\frac{1-\widetilde{\pi}_k(1,V_1)}{\widetilde{\pi}_k(1,V_1)P(T_{1k}=1|V=V_1)} \\
&+ \frac{1-\widetilde{\pi}_k(0,V_1)}{\widetilde{\pi}_k(0,V_1)P(T_{1k}=0|V=V_1)}\Bigg]\mathrm{E}_k^{*^T}(V_1)\Bigg\} \\
&+\sum_{k=1}^{r-2}\sum_{j=k+1}^{r-1}\mathrm{E}\Bigg\{\mathrm{E}_j^*(V_1)\Bigg[\frac{1-\widetilde{\pi}_j(1,V_1)}{\widetilde{\pi}_j(1,V_1)P(T_{1j}=1|V=V_1)} \\
&- \frac{1-\widetilde{\pi}_k(1,V_1)}{\widetilde{\pi}_k(1,V_1)P(T_{1k}=1|V=V_1)}\Bigg]\mathrm{E}_k^{*T}(V_1)\Bigg\} \\
&= M_0(\boldsymbol{\Theta};\widetilde{\pi}) + M_1(\boldsymbol{\Theta};\widetilde{\pi}) + A_1(\boldsymbol{\Theta};\widetilde{\pi}) + M_1^T(\boldsymbol{\Theta};\widetilde{\pi}) + A_2(\boldsymbol{\Theta};\widetilde{\pi}) \equiv M(\boldsymbol{\Theta};\widetilde{\pi}),
\end{aligned}
$$

where $M_0(\boldsymbol{\Theta},\widetilde{\pi})$, $M_1(\boldsymbol{\Theta},\widetilde{\pi})$, $A_1(\boldsymbol{\Theta},\widetilde{\pi})$ and $A_2(\boldsymbol{\Theta},\widetilde{\pi})$ are defined in Theorem 1.

Finally, we derive $\mathrm{Cov}\left\{S_j(T_{ij},X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi}),\varepsilon_k(T_{sk},V_s;\boldsymbol{\Theta},\widetilde{\pi})\right\}$ as follows. Note that

$$
\begin{aligned}
&E_j^*(V_i)\left\{\frac{[\delta_i-\widetilde{\pi}_j(1,V_i)]I(T_{ij}=1)}{\widetilde{\pi}_j(1,V_i)P(T_{1j}=1|V=V_i)} - \frac{[\delta_i-\widetilde{\pi}_j(0,V_i)]I(T_{ij}=0)}{\widetilde{\pi}_j(0,V_i)P(T_{1j}=0|V=V_i)}\right\} \\
&= E_j^*(V_i)\left\{\frac{[\delta_i-\widetilde{\pi}_j(1,V_i)]T_{ij}}{\widetilde{\pi}_j(1,V_i)P(T_{1j}=1|V=V_i)} - \frac{[\delta_i-\widetilde{\pi}_j(0,V_i)](1-T_{ij})}{\widetilde{\pi}_j(0,V_i)P(T_{1j}=0|V=V_i)}\right\} \\
&= E_j^*(V_i)\left\{\frac{\delta_i-\widetilde{\pi}_j(1,V_i)}{\widetilde{\pi}_j(1,V_i)P(T_{1j}=1|V=V_i)} + \frac{\delta_i-\widetilde{\pi}_j(0,V_i)}{\widetilde{\pi}_j(0,V_i)P(T_{1j}=0|V=V_i)}\right\}T_{ij} \\
&\quad -E_j^*(V_i)\frac{\delta_i-\widetilde{\pi}_j(0,V_i)}{\widetilde{\pi}_j(0,V_i)P(T_{1j}=0|V=V_i)}.
\end{aligned}
$$

Then, we have

$$
\begin{aligned}
&\mathrm{Cov}\left\{S_j(T_{ij},X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi}),\varepsilon_k(T_{sk},V_s;\boldsymbol{\Theta},\widetilde{\pi})\right\} \\
&= \mathrm{E}\left\{\delta_i\mathcal{X}_{ij}[T_{ij}-H_{+,j}(X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})]\left[\frac{[\delta_s-\widetilde{\pi}_k(1,V_s)]I(T_{sk}=1)}{\widetilde{\pi}_k(1,V_s)P(T_{1k}=1|V=V_s)}\right.\right. \\
&\left.\left. - \frac{[\delta_s-\widetilde{\pi}_k(0,V_s)]I(T_{sk}=0)}{\widetilde{\pi}_k(0,V_s)P(T_{1k}=0|V=V_s)}\right]E_k^{*^T}(V_s)\right\}
\end{aligned}
$$

$$= \mathrm{E}\left\{\delta_i \mathcal{X}_{ij}[T_{ij} - H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi})]T_{sk}\left[\frac{\delta_s - \widetilde{\pi}_k(1, V_s)}{\widetilde{\pi}_k(1, V_s)P(T_{1k} = 1|V = V_s)} + \frac{\delta_s - \widetilde{\pi}_k(0, V_s)}{\widetilde{\pi}_k(0, V_s)P(T_{1k} = 0|V = V_s)}\right] E_k^{*^T}(V_s)\right\}$$
$$-\mathrm{E}\left\{\delta_i \mathcal{X}_{ij}[T_{ij} - H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi})]\frac{\delta_s - \widetilde{\pi}_k(0, V_s)}{\widetilde{\pi}_k(0, V_s)P(T_{1k} = 0|V = V_s)} E_k^{*^T}(V_s)\right\}$$
$$= \mathrm{E}\left\{\mathrm{E}\left\{\delta_i \mathcal{X}_{ij}[T_{ij} - H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi})]T_{sk} \times \left[\frac{\delta_s - \widetilde{\pi}_k(1, V_s)}{\widetilde{\pi}_k(1, V_s)P(T_{1k} = 1|V = V_s)} + \frac{\delta_s - \widetilde{\pi}_k(0, V_s)}{\widetilde{\pi}_k(0, V_s)P(T_{1k} = 0|V = V_s)}\right] \times E_k^{*^T}(V_s)\Big| X_i, V_i, , X_s, V_s, \delta_i = 1, \delta_s = 1\right\}\right\}$$
$$-\mathrm{E}\left\{\mathrm{E}\left[\delta_i \mathcal{X}_{ij}[T_{ij} - H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi})]\frac{\delta_s - \widetilde{\pi}_k(0, V_s)}{\widetilde{\pi}_k(0, V_s)P(T_{1k} = 0|V = V_s)} \times E_k^{*^T}(V_s)\Big| X_i, V_i, X_s, V_s, \delta_i = 1, \delta_s = 1 \frac{\delta_s - \widetilde{\pi}_k(1, V_s)}{\widetilde{\pi}_k(1, V_s)P(T_{1k} = 1|V = V_s)}\right]\right\}$$

$$= \begin{cases} \mathrm{E}\left\{\delta_1 \mathcal{X}_{1j} H_{+,j}^{(1)}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})\left[\frac{1-\widetilde{\pi}_j(1,V_1)}{\widetilde{\pi}_j(1,V_1)P(T_{1j}=1|V=V_1)} + \frac{1-\widetilde{\pi}_j(0,V_1)}{\widetilde{\pi}_j(0,V_1)P(T_{1j}=0|V=V_1)}\right] E_k^{*^T}(V_1)\right\}, & \text{if } i=s,\ j=k \\ \mathrm{E}\left\{\delta_1 \mathcal{X}_j H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})[1 - H_{+,k}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})] \times \left[\frac{1-\widetilde{\pi}_k(1,V_1)}{\widetilde{\pi}_k(1,V_1)P(T_{1k}=1|V=V_1)} + \frac{1-\widetilde{\pi}_k(0,V_1)}{\widetilde{\pi}_k(0,V_1)P(T_{1k}=0|V=V_1)}\right] E_k^{*^T}(V_1)\right\}, & \text{if } i=s,\ j<k \\ \mathrm{E}\left\{\delta_1 \mathcal{X}_{1j} H_{+,k}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})[1 - H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})] \times \left[\frac{1-\widetilde{\pi}_k(1,V_1)}{\widetilde{\pi}_k(1,V_1)P(T_{1k}=1|V=V_1)} + \frac{1-\widetilde{\pi}_k(0,V_1)}{\widetilde{\pi}_k(0,V_1)P(T_{1k}=0|V=V_1)}\right] E_k^{*^T}(V_1)\right\}, & \text{if } i=s, k<j \\ 0, & \text{if } i \neq s \end{cases}$$

$$= \begin{cases} \mathrm{E}\left\{E_j^*(V_1)\left[\frac{1-\widetilde{\pi}_j(1,V_1)}{\widetilde{\pi}_j(1,V_1)P(T_{1j}=1|V=V_1)} + \frac{1-\widetilde{\pi}_j(0,V_1)}{\widetilde{\pi}_j(0,V_1)P(T_{1j}=0|V=V_1)}\right] E_k^{*^T}(V_1)\right\}, & \text{if } i=s,\ j=k \\ \mathrm{E}\left\{E_j^*(V_1)\left[\frac{1-\widetilde{\pi}_k(1,V_1)}{\widetilde{\pi}_k(1,V_1)P(T_{1k}=1|V=V_1)} + \frac{1-\widetilde{\pi}_k(0,V_1)}{\widetilde{\pi}_k(0,V_1)P(T_{1k}=0|V=V_1)}\right] E_k^{*^T}(V_1)\right\} \\ \quad +\mathrm{E}\left\{\delta_1 \mathcal{X}_{1j}[H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi}) - H_{+,k}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})] \times \left[\frac{1-\widetilde{\pi}_k(1,V_1)}{\widetilde{\pi}_k(1,V_1)P(T_{1k}=1|V=V_1)} + \frac{1-\widetilde{\pi}_k(0,V_1)}{\widetilde{\pi}_k(0,V_1)P(T_{1k}=0|V=V_1)}\right] E_k^{*^T}(V_1)\right\}, & \text{if } i=s,\ j<k \\ \mathrm{E}\left\{E_j^*(V_1)\left[\frac{1-\widetilde{\pi}_k(1,V_1)}{\widetilde{\pi}_k(1,V_1)P(T_{1k}=1|V=V_1)} + \frac{1-\widetilde{\pi}_k(0,V_1)}{\widetilde{\pi}_k(0,V_1)P(T_{1k}=0|V=V_1)}\right] E_k^{*^T}(V_1)\right\} \\ \quad +\mathrm{E}\left\{\delta_1 \mathcal{X}_{1j}[H_{+,k}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi}) - H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})] \times \left[\frac{1-\widetilde{\pi}_k(1,V_1)}{\widetilde{\pi}_k(1,V_1)P(T_{1k}=1|V=V_1)} + \frac{1-\widetilde{\pi}_k(0,V_1)}{\widetilde{\pi}_k(0,V_1)P(T_{1k}=0|V=V_1)}\right] E_k^{*^T}(V_1)\right\}, & \text{if } i=s, k<j \\ 0, & \text{if } i \neq s. \end{cases}$$

Also,

$$\operatorname{Cov}\left\{\sum_{j=1}^{r-1} S_j(T_{ij}, X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi}), \sum_{j=1}^{r-1} \varepsilon_j(T_{ij}, V_i; \boldsymbol{\Theta}, \widetilde{\pi})\right\}$$

$$= \sum_{j=1}^{r-1} \operatorname{Cov}\left\{S_j(T_{ij}, X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi}), \varepsilon_j(T_{ij}, V_i; \boldsymbol{\Theta}, \widetilde{\pi})\right\}$$

$$+ \sum_{j=1}^{r-2} \sum_{k=j+1}^{r-1} \operatorname{Cov}\left\{S_j(T_{ij}, X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi}), \varepsilon_k(T_{ik}, V_i; \boldsymbol{\Theta}, \widetilde{\pi})\right\}$$

$$+ \sum_{k=1}^{r-2} \sum_{j=k+1}^{r-1} \operatorname{Cov}\left\{S_j(T_{ij}, X_i, V_i; \boldsymbol{\Theta}, \widetilde{\pi}), \varepsilon_k(T_{ik}, V_i; \boldsymbol{\Theta}, \widetilde{\pi})\right\}$$

$$= \sum_{j=1}^{r-1} \mathrm{E}\left\{E_j^*(V_1)\left[\frac{1-\widetilde{\pi}_j(1, V_1)}{\widetilde{\pi}_j(1, V_1) P(T_{1j}=1|V=V_1)}\right.\right.$$

$$\left.\left.+ \frac{1-\widetilde{\pi}_j(0, V_1)}{\widetilde{\pi}_j(0, V_1) P(T_{1j}=0|V=V_1)}\right] E_k^{*^T}(V_1)\right\}$$

$$+ \sum_{j=1}^{r-2} \sum_{k=j+1}^{r-1} \mathrm{E}\left\{E_j^*(V_1)\left[\frac{1-\widetilde{\pi}_k(1, V_1)}{\widetilde{\pi}_k(1, V_1) P(T_{1k}=1|V=V_1)}\right.\right.$$

$$\left.\left.+ \frac{1-\widetilde{\pi}_k(0, V_1)}{\widetilde{\pi}_k(0, V_1) P(T_{1k}=0|V=V_1)}\right] E_k^{*^T}(V_1)\right\}$$

$$+ \sum_{j=1}^{r-2} \sum_{k=j+1}^{r-1} \mathrm{E}\left\{\delta_1 \mathcal{X}_{1j}[H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi}) - H_{+,k}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})]\right.$$

$$\left.\times\left[\frac{1-\widetilde{\pi}_k(1, V_1)}{\widetilde{\pi}_k(1, V_1) P(T_{1k}=1|V=V_1)} + \frac{1-\widetilde{\pi}_k(0, V_1)}{\widetilde{\pi}_k(0, V_1) P(T_{1k}=0|V=V_1)}\right] E_k^{*^T}(V_1)\right\}$$

$$+ \sum_{k=1}^{r-2} \sum_{j=k+1}^{r-1} \mathrm{E}\left\{E_j^*(V_1)\left[\frac{1-\widetilde{\pi}_k(1, V_1)}{\widetilde{\pi}_k(1, V_1) P(T_{1k}=1|V=V_1)}\right.\right.$$

$$\left.\left.+ \frac{1-\widetilde{\pi}_k(0, V_1)}{\widetilde{\pi}_k(0, V_1) P(T_{1k}=0|V=V_1)}\right] E_k^{*^T}(V_1)\right\}$$

$$+ \sum_{k=1}^{r-2} \sum_{j=k+1}^{r-1} \mathrm{E}\left\{\delta_1 \mathcal{X}_{1j}[H_{+,k}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi}) - H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \widetilde{\pi})]\right.$$

$$\left.\times\left[\frac{1-\widetilde{\pi}_k(1, V_1)}{\widetilde{\pi}_k(1, V_1) P(T_{1k}=1|V=V_1)} + \frac{1-\widetilde{\pi}_k(0, V_1)}{\widetilde{\pi}_k(0, V_1) P(T_{1k}=0|V=V_1)}\right] E_k^{*^T}(V_1)\right\}$$

$$
\begin{aligned}
&= M_0(\boldsymbol{\Theta};\widetilde{\pi}) + M_1(\boldsymbol{\Theta};\widetilde{\pi}) + B_1(\boldsymbol{\Theta};\widetilde{\pi}) + M_1^T(\boldsymbol{\Theta};\widetilde{\pi}) + B_1^T(\boldsymbol{\Theta};\widetilde{\pi})\\
&= M(\boldsymbol{\Theta};\widetilde{\pi}) - A_1(\boldsymbol{\Theta};\widetilde{\pi}) - A_2(\boldsymbol{\Theta};\widetilde{\pi}) + B_1(\boldsymbol{\Theta};\widetilde{\pi}) + B_1^T(\boldsymbol{\Theta};\widetilde{\pi})\\
&= M(\boldsymbol{\Theta};\widetilde{\pi}) + A(\boldsymbol{\Theta};\widetilde{\pi}),
\end{aligned}
$$

where $B_1(\boldsymbol{\Theta},\widetilde{\pi})$ and $A(\boldsymbol{\Theta};\widetilde{\pi})$ are defined in Theorem 1.

Hence,

$$
\begin{aligned}
&\mathrm{Var}\left\{\sum_{i=1}^{n}\sum_{j=1}^{r-1}[S_j(T_{ij},X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi}) - \varepsilon_j(T_{ij},V_i;\boldsymbol{\Theta},\widetilde{\pi})]\right\}\\
&\quad = \mathrm{Var}\left\{\sum_{j=1}^{r-1}S_j(T_{ij},X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})\right\} + \mathrm{Var}\left\{\sum_{j=1}^{r-1}\varepsilon_j(T_{ij},V_i;\boldsymbol{\Theta},\widetilde{\pi})\right\}\\
&\qquad -2\mathrm{Cov}\left\{\sum_{j=1}^{r-1}S_j(T_{ij},X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi}),\sum_{j=1}^{r-1}\varepsilon_j(T_{ij},V_i;\boldsymbol{\Theta},\widetilde{\pi})\right\}\\
&\quad = G(\boldsymbol{\Theta};\widetilde{\pi}) + M(\boldsymbol{\Theta};\widetilde{\pi}) - 2\{M(\boldsymbol{\Theta};\widetilde{\pi}) + A(\boldsymbol{\Theta};\widetilde{\pi})\}\\
&\quad = G(\boldsymbol{\Theta};\widetilde{\pi}) - M(\boldsymbol{\Theta};\widetilde{\pi}) - 2A(\boldsymbol{\Theta};\widetilde{\pi}).
\end{aligned}
$$

Moreover, we first prove the $\widehat{\boldsymbol{\Theta}}_{np}$ is a consistent estimator of $\boldsymbol{\Theta}$. Now,

$$
\begin{aligned}
G_{0n}(\boldsymbol{\Theta},\widehat{\widetilde{\pi}}) &= \frac{1}{\sqrt{n}}\left[-\frac{\partial\widehat{U}_{1n}(\boldsymbol{\Theta},\widehat{\widetilde{\pi}})}{\partial\boldsymbol{\Theta}^T}\right]\\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\delta_i\mathcal{X}_{ij}\mathcal{X}_{ij}^T H_{+,j}^{(1)}(X_i,V_i;\boldsymbol{\Theta},\widehat{\widetilde{\pi}}).
\end{aligned}
$$

Then it can be shown that $G_{0n}(\boldsymbol{\Theta},\widehat{\widetilde{\pi}}) \xrightarrow{p} G_0(\boldsymbol{\Theta},\widetilde{\pi}) = \mathrm{E}[\delta_1\mathcal{X}_{1j}H_{+,j}^{(1)}(X_1,V_1;\boldsymbol{\Theta},\widetilde{\pi})\mathcal{X}_{1j}^T]$. By Condition (A4), the convergence of $G_{0n}(\boldsymbol{\Theta},\widehat{\widetilde{\pi}})$ to $G_0(\boldsymbol{\Theta},\widetilde{\pi})$ is uniform in a neighborhood of the true $\boldsymbol{\Theta}$. By the Inverse Function Theorem of Foutz (1977), along with Condition (A3), there exists a unique consistent solution to the estimating equation $U_{1n}(\boldsymbol{\Theta},\widetilde{\pi}) = 0$ in a neighborhood of the true $\boldsymbol{\Theta}$. Hence, it follows that $\widehat{\boldsymbol{\Theta}}_{np}$ is a consistent estimator of $\boldsymbol{\Theta}$.

Next, we derive the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\Theta}}_{np} - \boldsymbol{\Theta})$. By Taylor series expansion of $U_{1n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}})$ at $(\boldsymbol{\Theta},\widehat{\widetilde{\pi}})$, we have

$$
\begin{aligned}
0 = \widehat{U}_{1n}(\widehat{\boldsymbol{\Theta}}_{np},\widehat{\widetilde{\pi}}) &= \widehat{U}_{1n}(\boldsymbol{\Theta},\widehat{\widetilde{\pi}}) + \frac{\partial\widehat{U}_{1n}(\boldsymbol{\Theta},\widehat{\widetilde{\pi}})}{\partial\boldsymbol{\Theta}^T}(\widehat{\boldsymbol{\Theta}}_{np} - \boldsymbol{\Theta}) + o_p(1)\\
&= \widehat{U}_{1n}(\boldsymbol{\Theta},\widehat{\widetilde{\pi}}) - G_0(\boldsymbol{\Theta},\widetilde{\pi})\sqrt{n}(\widehat{\boldsymbol{\Theta}}_{np} - \boldsymbol{\Theta}) + o_p(1).
\end{aligned}
$$

Hence,

$$
\sqrt{n}(\widehat{\boldsymbol{\Theta}}_{np} - \boldsymbol{\Theta}) = G_0^{-1}(\boldsymbol{\Theta},\widetilde{\pi})\widehat{U}_{1n}(\boldsymbol{\Theta},\widehat{\widetilde{\pi}}) + o_p(1).
$$

By Lemma 1, we have

$$\mathrm{Cov}\{\widehat{U}_{1n}(\boldsymbol{\Theta},\widehat{\widetilde{\pi}})\}=\mathrm{Var}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\left\{S_j(T_{ij},X_i,V_i;\boldsymbol{\Theta},\widetilde{\pi})-\varepsilon_j(T_{ij},V_i;\boldsymbol{\Theta},\widetilde{\pi})\right\}\right\}$$
$$= G(\boldsymbol{\Theta};\widetilde{\pi})-M(\boldsymbol{\Theta};\widetilde{\pi})-2A(\boldsymbol{\Theta};\widetilde{\pi})+o(1),$$

and $\widehat{U}_{1n}(\boldsymbol{\Theta},\widehat{\widetilde{\pi}})$ can be expressed as the sum of independent random variable. Using the Central Limit Theorem, $\sqrt{n}(\widehat{\boldsymbol{\Theta}}_{np}-\boldsymbol{\Theta})$ is asymptotically normally distributed with mean 0 and covariance matrix

$$\Delta_{np}=G_0^{-1}(\boldsymbol{\Theta},\widetilde{\pi})\left\{G(\boldsymbol{\Theta};\widetilde{\pi})-M(\boldsymbol{\Theta};\widetilde{\pi})-2A(\boldsymbol{\Theta};\widetilde{\pi})\right\}G_0^{-T}(\boldsymbol{\Theta},\widetilde{\pi}).$$

□

*Proof of Theorem 2* We first prove that $\widehat{\boldsymbol{\Theta}}_{vp}$ is a consistent estimator of $\boldsymbol{\Theta}$. Now,

$$G_{0n}(\boldsymbol{\Theta},\boldsymbol{\alpha})=\frac{1}{\sqrt{n}}\left[-\frac{\partial U_{2n}(\boldsymbol{\Theta},\boldsymbol{\alpha})}{\partial\boldsymbol{\Theta}^T}\right]=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\delta_i\mathcal{X}_{ij}\mathcal{X}_{ij}^T H_{+,j}^{(1)}(X_i,V_i;\boldsymbol{\alpha},\boldsymbol{\Theta}),$$

and

$$K_n(\boldsymbol{\Theta},\boldsymbol{\alpha})=\frac{1}{\sqrt{n}}\left[-\frac{\partial U_{2n}(\boldsymbol{\Theta},\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}^T}\right]$$
$$=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\delta_i\mathcal{X}_{ij}H_{+,j}^{(1)}(X_i,V_i;\boldsymbol{\Theta},\boldsymbol{\alpha})\left[\frac{\frac{\partial\widetilde{\pi}_j(1,V_i;\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}^T}}{\widetilde{\pi}_j(1,V_i;\boldsymbol{\alpha})}-\frac{\frac{\partial\widetilde{\pi}_j(0,V_i;\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}^T}}{\widetilde{\pi}_j(0,V_i;\boldsymbol{\alpha})}\right].$$

Then it can be shown that $G_{0n}(\boldsymbol{\Theta},\boldsymbol{\alpha})\xrightarrow{p}G_0(\boldsymbol{\Theta},\boldsymbol{\alpha})=\mathrm{E}\left\{\sum_{j=1}^{r-1}\delta_1\mathcal{X}_{1j}H_{+,j}^{(1)}(X_1,V_1;\boldsymbol{\Theta},\boldsymbol{\alpha})\mathcal{X}_{1j}^T\right\}$ and

$$K_n(\boldsymbol{\Theta},\boldsymbol{\alpha})\xrightarrow{p}\mathrm{E}\left\{\sum_{j=1}^{r-1}\delta_1\mathcal{X}_{1j}H_{+,j}^{(1)}(X_1,V_1;\boldsymbol{\Theta},\boldsymbol{\alpha})\left[\frac{\frac{\partial\widetilde{\pi}_j(1,V_1;\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}^T}}{\widetilde{\pi}_j(1,V_1;\boldsymbol{\alpha})}-\frac{\frac{\partial\widetilde{\pi}_j(0,V_1;\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}^T}}{\widetilde{\pi}_j(0,V_1;\boldsymbol{\alpha})}\right]\right\}$$
$$\equiv K(\boldsymbol{\Theta},\boldsymbol{\alpha}).$$

By Condition (A4), the convergence of $G_{0n}(\boldsymbol{\Theta},\widehat{\boldsymbol{\alpha}})$ to $G_0(\boldsymbol{\Theta},\boldsymbol{\alpha})$ is uniform in a neighborhood of the true $\boldsymbol{\Theta}$. By the Inverse Function Theorem of Foutz (1977), along with Condition (A3), there exists a unique consistent solution to the estimating equation $\widehat{U}_{2n}(\boldsymbol{\Theta},\widehat{\boldsymbol{\alpha}})=0$ in a neighborhood of the true $\boldsymbol{\Theta}$. Hence, it follows that $\widehat{\boldsymbol{\Theta}}_{vp}$ is a consistent estimator of $\boldsymbol{\Theta}$.

Next, we derive the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\Theta}}_{vp}-\boldsymbol{\Theta})$. By Taylor series expansion of $\widehat{U}_{2n}(\widehat{\boldsymbol{\Theta}}_{vp},\widehat{\boldsymbol{\alpha}})$ at $(\boldsymbol{\Theta},\boldsymbol{\alpha},P_1(V_i),\ldots,P_r(V_i))$, we have

$$
\begin{aligned}
0 &= \widehat{U}_{2n}(\widehat{\boldsymbol{\Theta}}_{vp}, \widehat{\boldsymbol{\alpha}}) \\
&= U_{2n}(\boldsymbol{\Theta}, \boldsymbol{\alpha}) + \frac{1}{\sqrt{n}}\left[\frac{\partial U_{2n}(\boldsymbol{\Theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\Theta}^T}\right]\sqrt{n}(\widehat{\boldsymbol{\Theta}}_{vp} - \boldsymbol{\Theta}) \\
&\quad + \frac{1}{\sqrt{n}}\left[\frac{\partial U_{2n}(\boldsymbol{\Theta}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^T}\right]\sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \\
&\quad + \sum_{i=1}^{n}\sum_{k=1}^{r}\frac{1}{\sqrt{n}}\left[\frac{\partial U_{2n}(\boldsymbol{\Theta}, \boldsymbol{\alpha})}{\partial P_k(V_i)}\right]\sqrt{n}[\widehat{P}_k(V_i) - P_k(V_i)] + o_p(1) \\
&= U_{2n}(\boldsymbol{\Theta}, \boldsymbol{\alpha}) - G_0(\boldsymbol{\Theta}, \boldsymbol{\alpha})\sqrt{n}(\widehat{\boldsymbol{\Theta}}_{vp} - \boldsymbol{\Theta}) - K(\boldsymbol{\Theta}, \boldsymbol{\alpha})\sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \\
&\quad - \frac{1}{\sqrt{n}}\sum_{s=1}^{n}\sum_{k=1}^{r}\Omega_{n,k}(V_s; \boldsymbol{\Theta}, \boldsymbol{\alpha})[I(Y_s = k) - P_k(V_s)] + o_p(1).
\end{aligned}
$$

Note that

$$
\begin{aligned}
&\sum_{i=1}^{n}\sum_{k=1}^{r}\frac{1}{\sqrt{n}}\left[-\frac{\partial U_{2n}(\boldsymbol{\Theta}, \boldsymbol{\alpha})}{\partial P_k(V_i)}\right]\sqrt{n}[\widehat{P}_k(V_i) - P_k(V_i)] \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\delta_i \mathcal{X}_{ij} H_{+,j}^{(1)}(X_i, V_i; \boldsymbol{\alpha}, \boldsymbol{\Theta})\sum_{k=1}^{r}\left[\frac{\frac{\partial \widetilde{\pi}_j(1, V_i; \boldsymbol{\alpha})}{\partial P_k(V_i)}}{\widetilde{\pi}_j(1, V_i; \boldsymbol{\alpha})} - \frac{\frac{\partial \widetilde{\pi}_j(0, V_i; \boldsymbol{\alpha})}{\partial P_k(V_i)}}{\widetilde{\pi}_j(0, V; \boldsymbol{\alpha})}\right] \\
&\quad \times \frac{\frac{1}{\sqrt{n}}\sum_{s=1}^{n}[I(Y_s = k) - P_k(V_s)]I(V_s = V_i)}{\frac{1}{n}\sum_{s=1}^{n} I(V_s = V_i)} \\
&= \frac{1}{\sqrt{n}}\sum_{s=1}^{n}\sum_{k=1}^{r} \\
&\quad \times \left\{\frac{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\delta_i \mathcal{X}_{ij} H_{+,j}^{(1)}(X_i, V_i; \boldsymbol{\alpha}, \boldsymbol{\Theta})\left[\frac{\frac{\partial \widetilde{\pi}_j(1, V_i; \boldsymbol{\alpha})}{\partial P_k(V_i)}}{\widetilde{\pi}_j(1, V_i; \boldsymbol{\alpha})} - \frac{\frac{\partial \widetilde{\pi}_j(0, V_i; \boldsymbol{\alpha})}{\partial P_k(V_i)}}{\widetilde{\pi}_j(0, V; \boldsymbol{\alpha})}\right] I(V_i = V_s)}{P(V = V_s)}\right\} \\
&\quad \times [I(Y_s = k) - P_k(V_s)] + o_p(1) \\
&= \frac{1}{\sqrt{n}}\sum_{s=1}^{n}\sum_{k=1}^{r}\Omega_{n,k}(V_s; \boldsymbol{\Theta}, \boldsymbol{\alpha})[I(Y_s = k) - P_k(V_s)] + o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
&\Omega_{n,k}(V_s; \boldsymbol{\Theta}, \boldsymbol{\alpha}) \\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\delta_i \mathcal{X}_{ij} H_{+,j}^{(1)}(X_i, V_i; \boldsymbol{\Theta}, \boldsymbol{\alpha})\left[\frac{\frac{\partial \widetilde{\pi}_j(1, V_i; \boldsymbol{\alpha})}{\partial P_k(V_i)}}{\widetilde{\pi}_j(1, V_i; \boldsymbol{\alpha})} - \frac{\frac{\partial \widetilde{\pi}_j(0, V_i; \boldsymbol{\alpha})}{\partial P_k(V_i)}}{\widetilde{\pi}_j(0, V; \boldsymbol{\alpha})}\right]\frac{I(V_i = V_s)}{P(V = V_s)}.
\end{aligned}
$$

In addition, under model (8), the MLE of $\boldsymbol{\alpha}$ is the solution of $M_n(\boldsymbol{\alpha}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathcal{V}_i\{\delta_i - H(\boldsymbol{\alpha}^T\mathcal{V}_i)\} = 0$. Since the estimating equation is unbiased, $\widehat{\boldsymbol{\alpha}}$ is consistent estimator of $\boldsymbol{\alpha}$. By Taylor series expansion $\sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = I_n^{-1}(\boldsymbol{\alpha})M_n(\boldsymbol{\alpha}) + o_p(1)$, where $I_n(\boldsymbol{\alpha}) = \frac{1}{\sqrt{n}}[-\partial M_n(\boldsymbol{\alpha})/\partial \boldsymbol{\alpha}^T] = n^{-1}\sum_{i=1}^n \mathcal{V}_i \mathcal{V}_i^T H(\boldsymbol{\alpha}^T \mathcal{V}_i)[1–H(\boldsymbol{\alpha}^T \mathcal{V}_i)]$. Since $I_n(\boldsymbol{\alpha}) \xrightarrow{p} I(\boldsymbol{\alpha})$ and $M_n(\boldsymbol{\alpha})$ converges in distribution to $N(\mathbf{0}, I(\boldsymbol{\alpha}))$, where $I(\boldsymbol{\alpha})$ was defined in Theorem 2, we have $\sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) = I^{-1}(\boldsymbol{\alpha})M_n(\boldsymbol{\alpha}) + o_p(1)$. Hence, $\sqrt{n}(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})$ is asymptotically normally distributed with mean $\mathbf{0}$ and variance $I^{-1}(\boldsymbol{\alpha})$. Since

$$\begin{aligned}\Omega_{n,k}(V_s; \boldsymbol{\Theta}, \boldsymbol{\alpha}) \xrightarrow{p} \mathrm{E}\Bigg\{ & \sum_{j=1}^{r-1} \delta_1 \mathcal{X}_{1j} H_{+,j}^{(1)}(X_1, V_1; \boldsymbol{\Theta}, \boldsymbol{\alpha}) \\ & \times \left[ \frac{\frac{\partial \widetilde{\pi}_j(1, V_1; \boldsymbol{\alpha})}{\partial P_k(V_1)}}{\widetilde{\pi}_j(1, V_1; \boldsymbol{\alpha})} - \frac{\frac{\partial \widetilde{\pi}_j(0, V_1; \boldsymbol{\alpha})}{\partial P_k(V_1)}}{\widetilde{\pi}_j(0, V_1; \boldsymbol{\alpha})} \right] \Bigg| V_1 = V_s \Bigg\} \\ \equiv {} & \Omega_k(V_s; \boldsymbol{\Theta}, \boldsymbol{\alpha}),\end{aligned}$$

we have

$$\begin{aligned}\sqrt{n}(\widehat{\boldsymbol{\Theta}}_{vp} - \boldsymbol{\Theta}) = G_0^{-1}(\boldsymbol{\Theta}, \boldsymbol{\alpha}) \Bigg\{ & U_{2n}(\boldsymbol{\Theta}, \boldsymbol{\alpha}) - K(\boldsymbol{\Theta}, \boldsymbol{\alpha}) I^{-1}(\boldsymbol{\alpha}) M_n(\boldsymbol{\alpha}) \\ & - \frac{1}{\sqrt{n}} \sum_{s=1}^n \sum_{k=1}^r \Omega_k(V_s; \boldsymbol{\Theta}, \boldsymbol{\alpha})[I(Y_s = k) - P_k(V_s)] \Bigg\} + o_p(1).\end{aligned}$$

Using the Central Limit Theorem, $\sqrt{n}(\widehat{\boldsymbol{\Theta}}_{vp} - \boldsymbol{\Theta})$ is asymptotically normally distributed with mean 0 and covariance matrix

$$\begin{aligned}\Delta_{vp} = G_0^{-1}(\boldsymbol{\Theta}, \boldsymbol{\alpha}) \Big\{ & G(\boldsymbol{\Theta}, \boldsymbol{\alpha}) + K(\boldsymbol{\Theta}, \boldsymbol{\alpha}) I^{-1}(\boldsymbol{\alpha}) K^T(\boldsymbol{\Theta}, \boldsymbol{\alpha}) + P(\boldsymbol{\Theta}, \boldsymbol{\alpha}) \\ & - 2J(\boldsymbol{\Theta}, \boldsymbol{\alpha}) I^{-1}(\boldsymbol{\alpha}) K^T(\boldsymbol{\Theta}, \boldsymbol{\alpha}) - 2Q(\boldsymbol{\Theta}, \boldsymbol{\alpha}) \Big\} G_0^{-T}(\boldsymbol{\Theta}, \boldsymbol{\alpha}).\end{aligned}$$

We now derive the covariance matrix $\Delta_{vp}$. Note that $G(\boldsymbol{\Theta}, \boldsymbol{\alpha})$, $J(\boldsymbol{\Theta}, \boldsymbol{\alpha})$, $Q(\boldsymbol{\Theta}, \boldsymbol{\alpha})$, and $P(\boldsymbol{\Theta}, \boldsymbol{\alpha})$ are defined in Theorem 2. We will show

$$\mathrm{Cov}\{U_{2n}(\boldsymbol{\Theta}, \boldsymbol{\alpha}), M_n(\boldsymbol{\alpha})\} = J(\boldsymbol{\Theta}, \boldsymbol{\alpha}),$$

$$\mathrm{Cov}\left\{ U_{2n}(\boldsymbol{\Theta}, \boldsymbol{\alpha}), \frac{1}{\sqrt{n}} \sum_{s=1}^n \sum_{k=1}^r \Omega_k(V_s; \boldsymbol{\Theta}, \boldsymbol{\alpha})[I(Y_s = k) - P_k(V_s)] \right\} = Q(\boldsymbol{\Theta}, \boldsymbol{\alpha}),$$

and

$$\mathrm{Cov}\left\{ \frac{1}{\sqrt{n}} \sum_{s=1}^n \sum_{k=1}^r \Omega_k(V_s; \boldsymbol{\Theta}, \boldsymbol{\alpha})[I(Y_s = k) - P_k(V_s)] \right\} = P(\boldsymbol{\Theta}, \boldsymbol{\alpha}).$$

First,

$$
\begin{aligned}
&\operatorname{Cov}\{U_{2n}(\boldsymbol{\Theta},\boldsymbol{\alpha}), M_n(\boldsymbol{\alpha})\}\\
&\quad= \frac{1}{n}\sum_{i=1}^{n}\operatorname{Cov}\left\{\sum_{j=1}^{r-1}\delta_i\mathcal{X}_{ij}[T_{ij}-H_{+,j}(X_i,V_i;\boldsymbol{\Theta},\boldsymbol{\alpha})],\mathcal{V}_i[\delta_i-H(\boldsymbol{\alpha}^T\mathcal{V}_i)]\right\}\\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left\{\sum_{j=1}^{r-1}\delta_i\mathcal{X}_{ij}[I(Y_i\le j)-H_{+,j}(X_i,V_i;\boldsymbol{\Theta},\boldsymbol{\alpha})]\mathcal{V}_i[\delta_i-H(\boldsymbol{\alpha}^T\mathcal{V}_i)]\right\}\\
&\quad=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left\{\sum_{j=1}^{r-1}\sum_{k=1}^{j}\delta_i\mathcal{X}_{ij}[1-H_{+,j}(X_i,V_i;\boldsymbol{\Theta},\boldsymbol{\alpha})][1-H(\boldsymbol{\alpha}^T\mathcal{V}_{k,i})]I(Y_i=k)\mathcal{V}_{k,i}^T\right\}\\
&\qquad-\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left\{\sum_{j=1}^{r-1}\sum_{k=j+1}^{r}\delta_i\mathcal{X}_{ij}H_{+,j}(X_i,V_i;\boldsymbol{\Theta},\boldsymbol{\alpha})[1-H(\boldsymbol{\alpha}^T\mathcal{V}_{k,i})]I(Y_i=k)\mathcal{V}_{k,i}^T\right\}\\
&\quad=\mathrm{E}\left\{\sum_{j=1}^{r-1}\mathcal{X}_{1j}[1-H_{+,j}(X_1,V_1;\boldsymbol{\Theta},\boldsymbol{\alpha})]\left[\sum_{k=1}^{j}H^{(1)}(\boldsymbol{\alpha}^T\mathcal{V}_{k,1})P(Y_1=k|X_1,V_1)\mathcal{V}_{k,1}^T\right]\right\}\\
&\qquad-\mathrm{E}\left\{\sum_{j=1}^{r-1}\mathcal{X}_{1j}H_{+,j}(X_1,V_{1i};\boldsymbol{\Theta},\boldsymbol{\alpha})\left[\sum_{k=j+1}^{r}H^{(1)}(\boldsymbol{\alpha}^T\mathcal{V}_{k,1})P(Y_1=k|X_1,V_1)\mathcal{V}_{k,1}^T\right]\right\}\\
&\quad\equiv J(\boldsymbol{\Theta},\boldsymbol{\alpha}).
\end{aligned}
$$

Also,

$$
\begin{aligned}
&\operatorname{Cov}\left\{U_{2n}(\boldsymbol{\Theta},\boldsymbol{\alpha}),\frac{1}{\sqrt{n}}\sum_{s=1}^{n}\sum_{k=1}^{r}\Omega_k(V_s;\boldsymbol{\Theta},\boldsymbol{\alpha})[I(Y_s=k)-P_k(V_s)]\right\}\\
&\quad=\frac{1}{n}\sum_{i=1}^{n}\operatorname{Cov}\left\{\sum_{j=1}^{r-1}\delta_i\mathcal{X}_{ij}[T_{ij}-H_{+,j}(X_i,V_i;\boldsymbol{\Theta},\boldsymbol{\alpha})],\sum_{j=1}^{r}[I(Y_i=j)\right.\\
&\qquad\left.-P_j(V_i)]\Omega_j^T(V_i;\boldsymbol{\Theta},\boldsymbol{\alpha})\right\}\\
&\quad=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\mathrm{E}\left\{\delta_i\mathcal{X}_{ij}[T_{ij}-H_{+,j}(X_i,V_i;\boldsymbol{\Theta},\boldsymbol{\alpha})]I(Y_i=j)\Omega_j^T(V_i;\boldsymbol{\Theta},\boldsymbol{\alpha})\right\}\\
&\qquad+\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\sum_{k=j+1}^{r}\mathrm{E}\left\{\delta_i\mathcal{X}_{ij}[T_{ij}-H_{+,j}(X_i,V_i;\boldsymbol{\Theta},\boldsymbol{\alpha})]I(Y_i=k)\Omega_k^T(V_i;\boldsymbol{\Theta},\boldsymbol{\alpha})\right\}\\
&\qquad+\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{r-2}\sum_{j=k+1}^{r-1}\mathrm{E}\left\{\delta_i\mathcal{X}_{ij}[T_{ij}-H_{+,j}(X_i,V_i;\boldsymbol{\Theta},\boldsymbol{\alpha})]I(Y_i=k)\Omega_k^T(V_i;\boldsymbol{\Theta},\boldsymbol{\alpha})\right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1} \mathrm{E}\left\{\delta_i \mathcal{X}_{ij}[1 - H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \boldsymbol{\alpha})] I(Y_i = j)\Omega_j^T(V_i; \boldsymbol{\Theta}, \boldsymbol{\alpha})\right\} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r-1}\sum_{k=j+1}^{r} \mathrm{E}\left\{\delta_i \mathcal{X}_{ij}[0 - H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \boldsymbol{\alpha})] I(Y_i = k)\Omega_k^T(V_i; \boldsymbol{\Theta}, \boldsymbol{\alpha})\right\} \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{r-2}\sum_{j=k+1}^{r-1} \mathrm{E}\left\{\delta_i \mathcal{X}_{ij}[1 - H_{+,j}(X_i, V_i; \boldsymbol{\Theta}, \boldsymbol{\alpha})] I(Y_i = k)\Omega_k^T(V_i; \boldsymbol{\Theta}, \boldsymbol{\alpha})\right\} \\
&= \mathrm{E}\left\{\sum_{j=1}^{r-1} \mathcal{X}_{1j}[1 - H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \boldsymbol{\alpha})]\right. \\
&\quad \times \left.\left[\sum_{k=1}^{j} H(\boldsymbol{\alpha}^T \mathcal{V}_{k,1}) P(Y_1 = k | X_1, V_1)\Omega_k^T(V_1; \boldsymbol{\Theta}, \boldsymbol{\alpha})\right]\right\} \\
&\quad - \mathrm{E}\left\{\sum_{j=1}^{r-1} \mathcal{X}_{1j} H_{+,j}(X_1, V_1; \boldsymbol{\Theta}, \boldsymbol{\alpha})\right. \\
&\quad \times \left.\left[\sum_{k=j+1}^{r} H(\boldsymbol{\alpha}^T \mathcal{V}_{k,1}) P(Y_1 = k | X_1, V_1)\Omega_k^T(V_1; \boldsymbol{\Theta}, \boldsymbol{\alpha})\right]\right\} \equiv Q(\boldsymbol{\Theta}, \boldsymbol{\alpha}).
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
&\mathrm{Cov}\left\{\frac{1}{\sqrt{n}}\sum_{s=1}^{n}\sum_{k=1}^{r} \Omega_k(V_s; \boldsymbol{\Theta}, \boldsymbol{\alpha})[I(Y_s = k) - P_k(V_s)]\right\} \\
&\quad = \sum_{j=1}^{r} \mathrm{E}\left\{\Omega_j(V_1; \boldsymbol{\Theta}, \boldsymbol{\alpha})[1 - P_j(V_1)] P_j(V_1)\Omega_j^T(V_1; \boldsymbol{\Theta}, \boldsymbol{\alpha})\right\} \\
&\quad\quad -2\sum_{j=1}^{r}\sum_{j<k} \mathrm{E}\left\{\Omega_j(V_1; \boldsymbol{\Theta}, \boldsymbol{\alpha}) P_j(V_1) P_k(V_1)\Omega_k^T(V_1; \boldsymbol{\Theta}, \boldsymbol{\alpha})\right\} \equiv P(\boldsymbol{\Theta}, \boldsymbol{\alpha}).
\end{aligned}
$$

In addition, it is easy to see that

$$
\mathrm{Cov}\{U_{2n}(\boldsymbol{\Theta}, \boldsymbol{\alpha})\} = G(\boldsymbol{\Theta}, \boldsymbol{\alpha}),
$$
$$
\mathrm{Cov}\left\{M_n(\boldsymbol{\alpha}), \frac{1}{\sqrt{n}}\sum_{s=1}^{n}\sum_{k=1}^{r} \Omega_k(V_s; \boldsymbol{\Theta}, \boldsymbol{\alpha})[I(Y_s = k) - P_k(V_s)]\right\} = 0.
$$

Therefore, we can obtain

$$\Delta_{vp} = G_0^{-1}(\boldsymbol{\Theta}, \boldsymbol{\alpha}) \Big\{ G(\boldsymbol{\Theta}, \boldsymbol{\alpha}) + K(\boldsymbol{\Theta}, \boldsymbol{\alpha}) I^{-1}(\boldsymbol{\alpha}) K^T(\boldsymbol{\Theta}, \boldsymbol{\alpha}) + P(\boldsymbol{\Theta}, \boldsymbol{\alpha}) - 2J(\boldsymbol{\Theta}, \boldsymbol{\alpha}) I^{-1}(\boldsymbol{\alpha}) K^T(\boldsymbol{\Theta}, \boldsymbol{\alpha}) - 2Q(\boldsymbol{\Theta}, \boldsymbol{\alpha}) \Big\} G_0^{-T}(\boldsymbol{\Theta}, \boldsymbol{\alpha}).$$

The proof is completed. □

## References


Agresti, A. (1984). *Analysis of ordinal categorical data*. New York: Wiley.

Ashby, D., et al. (1989). The ordered logistic regression model in psychiatry: Rising prevalence of dementia in old people homes. *Statistics in Medicine, 8*, 1317–1326.

Breslow, N. E., Cain, K. C. (1988). Logistic regression for two-stage case–control data. *Biometrika, 75*, 11–20.

Carroll, R. J., Wand, M. P. (1991). Semiparametric estimation in logistic measurement error models. *Journal of the Royal Statistical Society, Series B, 53*, 573–585.

Faden, V. B., Graubard, B. I. (2000). Maternal substance use during pregnancy and developmental outcome at age three. *Journal of Substance Abuse, 12*, 329–340.

Flander, W. D., Greenland, S. (1991). Analytic methods for two stage case-control studies and other stratified designs. *Statistics in Medicine, 10*, 739–747.

Foutz, R. V. (1977). On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association, 72*, 147–148.

Little, R. J. A. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association, 87*, 1227–1237.

Little, R. J. A., Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: Wiley.

Manor, O., et al. (2000). Dichotomous or categorical response? Analysing self-rated health and lifetime social class. *International Journal of Epidemiology, 57*, 149–157.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society, Series B, 42*, 109–142.

McCusker, J., et al. (1994). Human immunodeficiency virus type 1 antibody status and changes in risk behavior among drug users. *Annals of Epidemiology, 4*, 466–471.

Pepe, M. S., Fleming, T. R. (1991). A nonparametric method for dealing with mismeasured covariate data. *Journal of the American Statistical Association, 86*, 108–113.

Peterson, B. L., Harrell, F. E. (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics, 39*, 205–217.

Robins, J. M., Rotnitzky, A., Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association, 89*, 846–866.

Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association, 82*, 387–394.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*, 581–592.

Walker, S. H., Duncan, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika, 54*, 167–179.

Wang, C. Y., et al. (1997). Weighted semiparametric estimation in regression analysis with missing covariate data. *Journal of the American Statistical Association, 92*, 512–525.

Wang, C. Y., et al. (2002). Joint conditional likelihood estimator in logistic regression with missing covariate data. *Statistica Sinica, 12*, 555–574.

Zhao, L. P., Lipsitz, S. (1992). Designs and analysis of two-stage studies. *Statistics in Medicine, 11*, 769–782.