

Statistical Inference in Marginalized Zero-inflated Poisson Regression Models with Missing Data in Covariates

Kouakou Mathias Amani^{1*}, Ouagnina Hili^{1**}, and Konan Jean Geoffroy Kouakou^{1***}

¹*UMRI-Mathématiques et Nouvelles Technologies de l'Information, Institut National Polytechnique Félix Houphouët-Boigny (INP-HB) de Yamoussoukro, Yamoussoukro, Côte d'Ivoire*

Received April 17, 2023; revised July 1, 2023; accepted July 23, 2023

Abstract—The marginalized zero-inflated poisson (MZIP) regression model quantifies the effects of an explanatory variable in the mixture population. Also, in practice the variables are usually partially observed. Thus, we first propose to study the maximum likelihood estimator when all variables are observed. Then, assuming that the probability of selection is modeled using mixed covariates (continuous, discrete and categorical), we propose a semiparametric inverse-probability weighted (SIPW) method for estimating the parameters of the MZIP model with covariates missing at random (MAR). The asymptotic properties (consistency, asymptotic normality) of the proposed estimators are established under certain regularity conditions. Through numerical studies, the performance of the proposed estimators was evaluated. Then the results of the SIPW are compared to the results obtained by semiparametric inverse-probability weighted kernel-based (SIPWK) estimator method. Finally, we apply our methodology to a dataset on health care demand in the United States.

DOI: 10.3103/S1066530723040038

Keywords: count data, missing at random, semiparametric inverse-probability weighted kernel-based, semiparametric inverse probability weighting

1. INTRODUCTION

Although Poisson models (or binomial models) are the most widely used tools for modeling count data, we are seeing more and more count data with zero inflation in several fields such as economics, biomedical studies, criminology, insurance, sociology and political science. When the number of observed zeros is greater than that predicted by standard counting distributions, zero inflation (ZI) regression models are an alternative for modeling such data. For more information on ZI regression models, see Lambert [1], Diallo et al. [4, 7], Kouakou et al. [27], Ali et al. [28]. The ZIP distribution proposed by Lambert has gained popularity. ZIP regression models have been used successfully in a variety of important applications, see for example Dietz et al. [2], Yau et al. [3], and Cheung et al. [5].

However, the ZIP distribution has two group regression parameters, one for the probability of being an zero-inflation and the other for the Poisson mean. The parameters have latent class interpretations, these latent classes are often thought to classify some not-at-risk group and the at-risk group indicating a difference in susceptibility between the two populations. Because entire population parameters interpretations are desired, Long et al. [16] introduced the marginalized zero-inflation Poisson regression (MZIP).

In practice, the data are most often partially observed. In this context, the basic method used is called the complete cases method which consists in removing the individuals who have at least one missing data. This method is simple to implement. However, when the proportion of individuals who have a missing data is higher than 5% this method gives bad results. Two other alternatives to the complete case for the treatment of missing data are the Monte Carlo EM algorithm (MCEM) and multiple imputation (MI). The MCEM and MI methods are efficient but require quite high computational loads. Finally, the

*E-mail: amanimathias@gmail.com

**E-mail: o_hili@yahoo.fr

***E-mail: geoffroy.kouakou@inphb.ci

IPW method that is often used requires that we find the right model for the selection probability. see for example Diallo et al. [7] and Benecha et al. [18]. To circumvent these modeling difficulties while proposing a non-numerical method, Lukusa et al. [8, 9] proposed weighted semiparametric estimators that are suitable when the selection probabilities are expressed in terms of covariates of the same nature. However, there is little work on the estimation of the MZIP model in the context of missing data. This work aims to fill this gap. In this article, we propose a semiparametric approach in which the probability of selection that is a function of continuous, discrete and categorical covariates is estimated nonparametrically. This alternative consists in discretizing the continuous covariates using Jenks's method to have categorical covariates.

The rest of the paper is organized as follows. In the Section 2, We present the MZIP regression model and its maximum likelihood estimator. We present the SIPWK and SIPW estimation methods of MZIP model when the covariates are missing at random (MAR) and the consistency and asymptotic normality of the SIPW estimators are established in Section 3. The performance of the presented estimators are evaluated in Section 4. As an illustration, we apply these methods to real data in Section 5. A discussion and some perspectives are presented in Section 6. The technical proofs are reported in an Appendix.

2. MARGINALIZED ZIP MODELS

The ZIP distribution is used to model the counting variable of interest, namely $Y_i, i = 1 \dots n$. Y_i takes the value of from a Poisson distribution, with a mean of μ_i , with a probability of $1 - \psi_i$, or is drawn to zero from a Bernoulli distribution, with a probability of ψ_i . For example in dental caries research, the marginal mean ν_i caries count is often of more interest than the mean caries count μ_i of a susceptible latent group of individuals see Preisser [17].

Because entire population parameter interpretations are desired, the marginal mean ν_i can be modeled directly to give overall exposure effect estimates. Given that $\mu_i = \nu_i / (1 - \psi_i)$ the representation of the MZIP distribution is

$$\mathbb{P}(Y_i = k) = \begin{cases} \psi_i + (1 - \psi_i) \exp(-\nu_i / (1 - \psi_i)), & k = 0 \\ (1 - \psi_i) \frac{\exp(-\nu_i / (1 - \psi_i)) [\nu_i / (1 - \psi_i)]^k}{k!}, & k > 0. \end{cases} \tag{2.1}$$

In the MZIP model, Long et al. [16] links regression parameters directly to the marginal mean ν_i , while employing another set of parameters to model the probability of being an excess zero (i.e., ψ_i). The parameters ν_i and ψ_i of MZIP model are modeling by

$$\text{logit}(\psi_i) = \mathbf{Z}_i^T \gamma \quad \text{and} \quad \text{log}(\nu_i) = \mathbf{X}_i^T \alpha, \tag{2.2}$$

where $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_q)^T$ is a $(q \times 1)$ column have the same interpretation as in ZIP model, $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)^T$ is a $(p \times 1)$ vector of regression parameters for ν_i having interpretations as the log-incidence density ratio (IDR) for the entire sample population and $\mathbf{X}_{i(p \times 1)}$ and $\mathbf{Z}_{i(q \times 1)}$ denote the vectors of covariates for the i th individual. Let $\theta = (\gamma^T, \alpha^T)^T$. Consider that we observe a sample of n independent copies $(Y_1, \mathbf{X}_1, \mathbf{Z}_1), (Y_2, \mathbf{X}_2, \mathbf{Z}_2), \dots, (Y_n, \mathbf{X}_n, \mathbf{Z}_n)$ of $(Z, \mathbf{X}, \mathbf{Z})$. Then, the log-likelihood of θ is

$$l_n(\theta) = \sum_{i=1}^n -\log(1 + e^{\mathbf{Z}_i^T \gamma}) + J_i \log \left(e^{\mathbf{Z}_i^T \gamma} + e^{-(1 + \exp(\mathbf{Z}_i^T \gamma)) \exp(\mathbf{X}_i^T \alpha)} \right) + \sum_{i=1}^n (1 - J_i) \left(-(1 + e^{\mathbf{Z}_i^T \gamma}) e^{\mathbf{X}_i^T \alpha} + Y_i \log(1 + e^{\mathbf{Z}_i^T \gamma}) + \mathbf{X}_i^T \alpha Y_i - \log(Y_i!) \right),$$

where $J_i = 1_{\{Y_i=0\}}$. The maximum likelihood estimator $\hat{\theta}_{F,n} = (\hat{\gamma}_n^T, \hat{\alpha}_n^T)^T$ of θ is the solution of the equation $U_{F,n}(\theta) = 0$, with

$$U_{F,n}(\theta) = \frac{1}{\sqrt{n}} \frac{\partial l_n(\theta)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial l_i(\theta)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}_i(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbf{Z}_i B_i(\theta), \mathbf{X}_i A_i(\theta))^T, \tag{2.3}$$

where

$$A_i(\theta) = (Y_i - e^{\mathbf{X}_i^T \alpha} (1 + e^{\mathbf{Z}_i^T \gamma})) (1 - J_i) - \frac{e^{\mathbf{X}_i^T \alpha} (1 + e^{\mathbf{Z}_i^T \gamma}) J_i}{e^{\mathbf{Z}_i^T \gamma + h_i(\theta)} + 1},$$

$$B_i(\theta) = \frac{J_i e^{\mathbf{Z}_i^T \gamma} (e^{h_i(\theta)} - e^{\mathbf{X}_i^T \alpha})}{e^{\mathbf{Z}_i^T \gamma + h_i(\theta)} + 1} + \frac{e^{\mathbf{Z}_i^T \gamma} (Y_i - 1)}{1 + e^{\mathbf{Z}_i^T \gamma}} - (1 - J_i) e^{\mathbf{X}_i^T \alpha + \mathbf{Z}_i^T \gamma},$$

and

$$h_i(\theta) = (1 + \exp(\mathbf{Z}_i^T \gamma)) \exp(\mathbf{X}_i^T \alpha).$$

3. ESTIMATING PARAMETERS WITH MISSING COVARIATES

Let \mathbf{X} and \mathbf{Z} be the vectors covariates with missing data and Y always observed. Let Δ_i be a dummy variable that is 1 when $\{\mathbf{Z}_i, \mathbf{X}_i\}$ is completely observed, 0 otherwise, see Rubin [12] for details. We consider covariates mixed (continuous, discrete, and categorical). Let $\mathbf{V} = (Y, \mathbf{S}^D, \mathbf{S}^C)^T$, where $\mathbf{S}^D = (\mathbf{X}^{D(\text{obs}),T}, \mathbf{Z}^{D(\text{obs}),T})$ denote the vector of discrete variables that are always observed on each individual, $\mathbf{S}^C = (\mathbf{X}^{C(\text{obs}),T}, \mathbf{Z}^{C(\text{obs}),T})$ denote the vector of continuous variables that are always observed on each individual and $\{\mathbf{X}^{(\text{miss}),T}, \mathbf{Z}^{(\text{miss}),T}\}$ the missing components of $\{\mathbf{X}, \mathbf{Z}\}$. Under the MAR mechanism, define the selection probability

$$\pi(\mathbf{V}_i) = \mathbb{P}(\Delta_i = 1 | Y_i, \mathbf{X}_i, \mathbf{Z}_i) = \mathbb{P}(\Delta_i = 1 | \mathbf{V}_i).$$

3.1. Kernel-Based Weighting Estimator of a MZIP Model

Let $\mathbf{D} = (\mathbf{X}^{(\text{obs}),T}, \mathbf{Z}^{(\text{obs}),T})$ and $d \in \{d_1, d_2, \dots, d_m\}$ denote the distinct values of the \mathbf{D} . We consider $\hat{\pi}(y, d)$ a Nadaraya–Watson (N-W) [22, 24] type estimator of $\pi(y, d)$ defined by

$$\hat{\pi}(y, d) = \frac{\sum_{k=1}^n \Delta_k K_h(Y_k = y, \mathbf{D}_k - d)}{\sum_{i=1}^n K_h(Y_i = y, \mathbf{D}_i - d)},$$

where K_h is a kernel function and h is a bandwidth satisfying some conditions stated in Wang [23]. The resulting semiparametric kernel-assisted weighting (SIPWK) estimator $\hat{\theta}_n^{wsk}$ of θ in models 2.1 and 2.2 is the solution of the equation

$$U_{w,n}(\theta, \hat{\pi}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\Delta_i}{\hat{\pi}(Y_i, \mathbf{D}_i)} \dot{l}_i(\theta) = 0. \tag{3.1}$$

In the following section, we present another weighted semiparametric estimation of a MZIP regression model.

3.2. Semiparametric IPW (SIPW) Estimator of a MZIP Model

We recall that $\mathbf{S}^C = (\mathbf{X}^{C(\text{obs}),T}, \mathbf{Z}^{C(\text{obs}),T})$ is the set of observed continuous covariates. Inspired by Jenks' method [26], we discretize this set. Using Herbert's method [25], we obtain the number of optimal classes. Jenk's method is based on the similarity principle. The method minimizes the intraclass variance. This method allows to have new categorical covariates \mathbf{S}'^D .

Let $s_1^D, s_2^D, \dots, s_m^D$ denote the distinct values of the \mathbf{S}_i^D s, $s_1'^D, s_2'^D, \dots, s_m'^D$ denote the distinct values of the \mathbf{S}'^D s. The nonparametric estimator of $\pi(y, s^D, s'^D)$ is given by the following expression:

$$\hat{\pi}(y, s^D, s'^D) = \frac{\sum_{k=1}^n \Delta_k I(Y_k = y, \mathbf{S}_k^D = s^D, \mathbf{S}'_k{}^D = s'^D)}{\sum_{i=1}^n I(Y_i = y, \mathbf{S}_i^D = s^D, \mathbf{S}'_i{}^D = s'^D)},$$

where $y = 0, 1, 2, \dots, s^D \in \{s_1^D, s_2^D, \dots, s_m^D\}$ and $s'^D \in \{s_1'^D, s_2'^D, \dots, s_m'^D\}$.

Thus, in this context, the SIPW estimator $\hat{\theta}_n^{ws}$ of θ in models 2.1 and 2.2 is the solution of the equation

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\Delta_i}{\hat{\pi}(Y_i, \mathbf{S}_i^D, \mathbf{S}'_i{}^D)} \dot{l}_i(\theta) = 0. \tag{3.2}$$

We study the asymptotic properties of $\hat{\theta}_n^F$ and $\hat{\theta}_n^{ws}$ in the following section.

3.3. Asymptotic Results

To establish the asymptotic properties of $\hat{\theta}_n^F$ and $\hat{\theta}_n^{ws}$ we give conditions of regularity.

- H1.** The true parameter value $\theta_0 := (\gamma_0^T, \alpha_0^T)^T$ lies in the interior of some known compact set of $\mathbb{R}^p \times \mathbb{R}^q$.
- H2.** Let $\text{supp}(\mathbf{S}^D)$ denote's the support of \mathbf{S}^D and $\text{supp}(\mathbf{S}'^D)$ denote's the support of \mathbf{S}'^D . Assume $\text{supp}(\mathbf{S}^D)$ and $\text{supp}(\mathbf{S}'^D)$ does not depend on θ . Furthermore, for any $y = 0, 1, \dots$, for $s^D \in \text{supp}(\mathbf{S}^D)$ and for $s'^D \in \text{supp}(\mathbf{S}'^D)$, the selection probability $\pi(y, s^D, s'^D) > 0$.
- H3.** $\mathbb{E} \left[\frac{\dot{l}_i(\theta) \dot{l}_i(\theta)^T}{\pi(\mathbf{V}_i)} \right]$ is finite and positive definite in neighborhood of the true θ .
- H4.** In a neighborhood of the true θ , the first and second derivatives of $U_{F,n}(\theta)$ with respect to θ exist almost surely and are uniformly bounded above by a function of $(Y, \mathbf{X}, \mathbf{Z})$, whose expectations exist.
- H5.** The first derivatives of $U_{w,n}(\theta, \pi)$ with respect to θ exist almost surely in a neighborhood of θ_0 . Additionally, in such a neighborhood, these first derivatives are uniformly bounded above by a function of $(Y, \mathbf{X}, \mathbf{Z})$, whose expectations exist.

The asymptotic properties of $\hat{\theta}_n^F$ and $\hat{\theta}_n^{ws}$ are stated in Theorems 1 and 2, respectively. The detailed of proofs of Theorem 1 in the Appendix A and Theorem 2 in the Appendix B.

Before studying the asymptotic properties of the estimators, we define by

$$\Sigma_n(\theta) = -n^{-1/2} \frac{\partial U_{F,n}}{\partial \theta^T} = -\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\partial^2 l_i(\theta)}{\partial \theta \partial \theta^T} \right\} \text{ and } Q_F(\theta_0) = \mathbb{E} \left[\dot{l}_1(\theta_0) \dot{l}_1(\theta_0)^T \right].$$

Because each component of $\Sigma_n(\theta)$ is a mean of independent and identically distributed random variables, we have $\mathbb{E} [\Sigma_n(\theta)] = \mathbb{E} \left[-\frac{\partial^2 l_1(\theta)}{\partial \theta \partial \theta^T} \right] = \Sigma(\theta)$.

Theorem 1. *Assume that conditions (H1), (H2), and (H4) hold. Then $\hat{\theta}_n^F$ converges in probability to θ_0 , as $n \rightarrow \infty$ and $\sqrt{n}(\hat{\theta}_n^F - \theta_0)$ has an asymptotic normal distribution with mean zero and covariance matrix Δ_F , with $\Delta_F := \Sigma(\theta_0)^{-1} Q_F(\theta_0) [\Sigma(\theta_0)^{-1}]^T$, where $Q_F(\theta) = \mathbb{E} \left[\dot{l}_1(\theta) \dot{l}_1(\theta)^T \right]$.*

Since the inverse of the Fisher information matrix is the variance of the score function, we can have $\Sigma(\theta_0) = Q_F(\theta_0)$. Finally $\Delta_F = \Sigma(\theta_0)^{-1}$.

Theorem 2. *Assume that conditions (H1), (H2), and (H4) hold. Then $\hat{\theta}_n^{ws}$ converges in probability to θ_0 , as $n \rightarrow \infty$ and $\sqrt{n}(\hat{\theta}_n^{ws} - \theta_0)$ has an asymptotic normal distribution with mean zero and covariance matrix Δ_{ws} , with $\Delta_{ws} := \Sigma(\theta_0)^{-1} \{ \Omega_3(\theta_0, \pi) - [\Omega_4(\theta_0, \pi) - \Omega_5(\theta_0, \pi)] [\Sigma(\theta_0)^{-1}]^T \}$, where $\Omega_3(\theta_0, \pi) = \mathbb{E} \left[\frac{\dot{l}_i(\theta_0) \dot{l}_i(\theta_0)^T}{\pi(Y_i, \mathbf{S}_i^D, \mathbf{S}'_i{}^D)} \right]$, $\Omega_4(\theta_0, \pi) = \mathbb{E} \left[\frac{\dot{l}_i^*(\theta_0) \dot{l}_i^*(\theta_0)^T}{\pi(Y_i, \mathbf{S}_i^D, \mathbf{S}'_i{}^D)} \right]$, $\Omega_5(\theta_0, \pi) = \mathbb{E} \left[\dot{l}_i^*(\theta_0) \dot{l}_i^*(\theta_0)^T \right]$, and $\dot{l}_i^*(\theta_0) = \mathbb{E} \left[\dot{l}_i(\theta_0) | Y_i, \mathbf{S}_i^D, \mathbf{S}'_i{}^D \right]$.*

4. SIMULATIONS STUDY

In this section, we study the performances under various conditions of the following estimators:

- $\hat{\theta}_n^F$ the maximum likelihood estimator obtained by solving the equation $U_{F,n}(\theta) = 0$ where $U_{F,n}(\theta)$ is defined in 2.3.
- $\hat{\theta}_n^{wsk}$ the SIPWK estimator obtained by solving the Eq. (3.1).
- $\hat{\theta}_n^{ws}$ the SIPW estimator obtained by solving the Eq. (3.2).

In this numerical study, we consider samples of size $n = 2000$ and 1000 .

$$\begin{aligned} \text{logit}(\psi_i) &= \gamma_1 Z_{i1} + \gamma_2 Z_{i2} + \gamma_3 Z_{i3} + \gamma_4 Z_{i4}, \\ \log(\nu_i) &= \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i3}, \end{aligned} \tag{4.1}$$

where $X_{i1} = Z_{i1} = 1$, $Z_{i2} = X_{i2}$, and $Z_{i2}, Z_{i3}, Z_{i4}, X_{i3}$, follows, respectively, the Gaussian distribution $N(0, 1.7)$, Poisson distribution $P(0.5)$, exponential distribution $E(1)$, and binomial distribution $B(1, 0.5)$. The regression parameter α is chosen as follows $\alpha = (1.2, 0.2, -0.7)^T$. The regression parameter γ is chosen as follows

- case 1: $\gamma = (-1, 0.4, 0.3, 0.45)^T$,
- case 2: $\gamma = (-1, 0.62, 0.3, 0.8)^T$.

In case 1 (respectively case 2), the average percentage of zero inflation in this simulation is 41% (respectively 65%). In the variable Z_{i4} , we assume that the data are missing. The average fraction of missing data (AFMD) in the simulated samples is equal to 15 and 30%. We used a multiplicative kernel (the Dirac discrete kernel for discrete variables and the Gaussian kernel for the continuous variable) for the kernel-based weighting estimator of an MZIP model. Finally, for each configuration (sample size, proportions of zero inflation and missing data), we simulate $N = 1000$ samples and calculate $\hat{\theta}_n^{ws}$ and $\hat{\theta}_n^{wsk}$. We use the statistical software R.3.5.2 to perform our simulations and the maxlik package (see Henningsen et al. [19]) to solve Eqs. (2.3), (3.1). We compute the bias of the estimates $\hat{\gamma}_{j,n}$ and $\hat{\alpha}_{k,n}$. We obtain the bias, the standard deviation (SD) and the mean square error (RMSE) for each estimator $\hat{\gamma}_{j,n} (j = 1, \dots, 4)$ and $\hat{\alpha}_{k,n} (k = 1, \dots, 3)$. For comparison purposes, we also provide the results that would be obtained if there were no missing covariates. In this case, the MLE is obtained by solving the score equation (2.2) (FD estimator). In Table 1, we present the results for $n = 500$, 41% (top) and 65% (bottom) zero inflation and mean missing data 15 and 30%. Table 2, we present the results for $n = 1000$, 41% (top) and 65% (bottom) zero inflation and mean missing data 15 and 30%. Table 3 provides the results for $n = 2000$, 41% (top) and 65% (bottom) zero inflation and the average missing data 15 and 30%. The Tables 1–3 show that both methods perform well, as the results obtained with both methods are close to the base case. The results also show that the bias and RMSE of the proposed method are generally better than the bias and RMSE of the SIPWK method. Let us now examine the performance of the proposed estimator. The results in Tables 1–3 show that the bias, standard deviation, and RMSE decrease as the sample size increases and the proportion of individuals with missing covariates decreases. Furthermore, the bias remains reasonable even with 30% missing data. The estimator $\hat{\theta}_n^F$ is obviously better than $\hat{\theta}_n^{wsk}$ and $\hat{\theta}_n^{ws}$, but FD is only possible in the absence of missing data.

Table 1. Simulation results for $n = 500$, zero inflation: 41% (top) and 65% (bottom)

AFMD			$\hat{\gamma}_n$				$\hat{\alpha}_n$		
			$\hat{\gamma}_{1,n}$	$\hat{\gamma}_{2,n}$	$\hat{\gamma}_{3,n}$	$\hat{\gamma}_{4,n}$	$\hat{\alpha}_{1,n}$	$\hat{\alpha}_{2,n}$	$\hat{\alpha}_{3,n}$
FD		Bias	-0.0133	0.0038	0.0019	0.0068	-0.0026	-0.0005	-0.0003
		SD	0.1379	0.0925	0.0747	0.0524	0.0527	0.0479	0.0657
		RMSE	0.1385	0.0926	0.0747	0.0528	0.0527	0.0479	0.0657
SIPWK	0.15	Bias	-0.0029	0.0120	0.0057	0.0058	-0.0109	-0.0082	-0.0031
		SD	0.1422	0.1008	0.0803	0.0576	0.0555	0.0525	0.0710
		RMSE	0.1422	0.1015	0.0805	0.0578	0.0565	0.0531	0.0711
SIPW	0.15	Bias	-0.0142	0.0046	0.0011	0.0075	-0.0026	-0.0007	-0.0017
		SD	0.1502	0.1012	0.0817	0.0573	0.0577	0.0524	0.0730
		RMSE	0.1508	0.1013	0.0817	0.0577	0.0577	0.0524	0.0730
SIPWK	0.30	Bias	0.0097	0.0207	0.0110	0.0057	-0.0226	-0.0157	-0.0061
		SD	0.1521	0.1080	0.0860	0.0657	0.0581	0.0567	0.0754
		RMSE	0.1523	0.1099	0.0867	0.0659	0.0623	0.0588	0.0756
SIPW	0.30	Bias	-0.0149	0.0062	0.0008	0.0087	-0.0048	-0.0004	-0.0005
		SD	0.1702	0.1107	0.0898	0.0645	0.0635	0.0572	0.0787
		RMSE	0.1707	0.1108	0.0897	0.0651	0.0636	0.0572	0.0787
FD		Bias	-0.0202	0.0020	0.0054	0.0083	0.0008	0.0021	0.0009
		SD	0.1362	0.0898	0.0675	0.0530	0.0568	0.0544	0.0656
		RMSE	0.1376	0.0898	0.0677	0.0536	0.0567	0.0544	0.0656
SIPWK	0.15	Bias	-0.0040	0.0144	0.0104	0.0079	-0.0123	-0.0085	-0.0034
		SD	0.1433	0.0972	0.0747	0.0583	0.0594	0.0582	0.0704
		RMSE	0.1433	0.0983	0.0754	0.0588	0.0606	0.0588	0.0704
SIPW	0.15	Bias	-0.0194	0.0030	0.0061	0.0092	0.0003	0.0015	-0.0019
		SD	0.1503	0.0974	0.0751	0.0586	0.0623	0.0583	0.0716
		RMSE	0.1515	0.0974	0.0753	0.059	0.0623	0.0583	0.0716

Table 1. (Contd.)

AFMD			$\hat{\gamma}_n$				$\hat{\alpha}_n$		
			$\hat{\gamma}_{1,n}$	$\hat{\gamma}_{2,n}$	$\hat{\gamma}_{3,n}$	$\hat{\gamma}_{4,n}$	$\hat{\alpha}_{1,n}$	$\hat{\alpha}_{2,n}$	$\hat{\alpha}_{3,n}$
SIPWK	0.30	Bias	0.0179	0.0289	0.0154	0.0086	-0.0294	-0.0213	-0.0068
		SD	0.1523	0.1077	0.0833	0.0662	0.0636	0.0646	0.0776
		RMSE	0.1532	0.1115	0.0847	0.0667	0.0700	0.0680	0.0779
SIPW	0.30	Bias	-0.0204	0.0034	0.0059	0.0111	0.0003	0.0015	-0.0024
		SD	0.1664	0.1084	0.0851	0.0665	0.0694	0.0645	0.0794
		RMSE	0.1675	0.1084	0.0853	0.0674	0.0693	0.0645	0.0794

5. APPLICATION

In this section, we describe an application of the MZIP model to NMES1988 data obtained from the National Medical Expenditure Survey (NMES) conducted in 1987–1988. We analyze the variable ofnp (number of consultations with a non-physician health professional in a practice) by the MZIP. The proportion of zero in the observations of this variable is equal to 0.6818. This very high proportion suggest a situation of inflation of zeros. For each of the individuals i ($i = 1 \dots n = 4406$) of the sample, let Y_i denote the number of consultations a non-physician health professional in a practice.

- ψ_i represents the probability that patient i will give up in such a way systematic to consult a non-physician professional.
- ν_i represents the average number of consultations with a health professional not doctor, for a patient i .

To model the marginal mean and zero-inflation parameters ν_i and ψ_i defined in (2.2), where Z_i and X_i are the set of covariates, we proceeded as follows. First, we fitted an MZIP regression model incorporating all the covariates available in (2.2), i.e., taking $X_i = Z_i$ for each i . Next, Wald tests were used to select the relevant covariates in the sub-models (2.2). Through this procedure, we identify three significant predictors included in ν_i (chronic, gender, school) and six significant predictors included in ψ_i (chronic, medicaid, age, income, gender, school). The significant covariates are gender (1 for female, 0 for male), age (in years, divided by 10), school (number of years of education), income (in 10 000 dollars), chronic diseases (cancer, arthritis, diabetes...), and medicaid (a binary variable indicating whether the individual is covered by medicaid or not). The covariate age (in years, divided by 10) was discretized before applying the proposed method. We therefore model ψ_i and ν_i as follows:

$$\text{logit}(\psi_i) = \gamma_1\text{inter} + \gamma_2\text{chronic} + \gamma_3\text{medicaid} + \gamma_4\text{age} + \gamma_5\text{income} + \gamma_6\text{gender} + \gamma_7\text{school}, \quad (5.1)$$

$$\log(\nu_i) = \alpha_1\text{inter} + \alpha_2\text{chronic} + \alpha_3\text{gender} + \alpha_4\text{school}. \quad (5.2)$$

We simulated 15% (moderate) and 30% (high) proportions of missing data in the “income” variable, respectively. Indeed, among the covariates, the “income” variable is the most likely to have missing data, as it is more sensitive and confidential information. Respondents are often reluctant to disclose their income, which can lead to higher rates of missing data for this variable. According to Mishra et al. [29], National Health and Nutrition Examination Survey, the rate of missing data in the “income” variable is often high, reaching or exceeding 15%. Tables 4 and 5 show the estimation results for the case with no missing data (FD) and 15% missing data, followed by the case with no missing data (FD) and 30% missing data, respectively. We can say that the proposed method is robust because when the percentage of missing data increases, the covariates remain significant and the coefficients keep the same signs as in the reference case (FD). We can state that the variables of Medicaid status and

Table 2. Simulation results for $n = 1000$, zero inflation: 41% (top) and 65% (bottom)

AFMD			$\hat{\gamma}_n$				$\hat{\alpha}_n$		
			$\hat{\gamma}_{1,n}$	$\hat{\gamma}_{2,n}$	$\hat{\gamma}_{3,n}$	$\hat{\gamma}_{4,n}$	$\hat{\alpha}_{1,n}$	$\hat{\alpha}_{2,n}$	$\hat{\alpha}_{3,n}$
FD		Bias	-0.0230	0.0075	0.0066	0.0053	-0.0053	0.0001	-0.0047
		SD	0.2479	0.0839	0.1005	0.0754	0.0784	0.0470	0.1025
		RMSE	0.2488	0.0842	0.1006	0.075	0.0785	0.0470	0.1025
SIPWK	0.15	Bias	-0.0421	0.0307	0.0123	0.0066	-0.0101	-0.0131	-0.0059
		SD	0.2646	0.0877	0.1130	0.0828	0.0800	0.0492	0.1110
		RMSE	0.2678	0.0928	0.1137	0.0830	0.0806	0.0509	0.1111
SIPW	0.15	Bias	-0.0355	0.0137	0.0091	0.0067	-0.0047	-0.0023	-0.0055
		SD	0.2685	0.0894	0.1109	0.0822	0.0842	0.0498	0.1098
		RMSE	0.2707	0.0904	0.1112	0.0824	0.0843	0.0499	0.1099
SIPWK	0.30	Bias	-0.0661	0.0587	0.0165	0.0090	-0.0195	-0.0263	-0.0075
		SD	0.2963	0.0965	0.1292	0.0978	0.0847	0.0519	0.1224
		RMSE	0.3034	0.1129	0.1302	0.0982	0.0869	0.0582	0.1226
SIPW	0.30	Bias	-0.0407	0.0166	0.0094	0.0075	-0.0083	-0.0015	-0.0053
		SD	0.3148	0.1042	0.1262	0.0963	0.0941	0.0540	0.1221
		RMSE	0.3173	0.1055	0.1265	0.0966	0.0945	0.0540	0.1221
FD		Bias	-0.0143	0.0057	0.0058	0.0097	-0.0089	0.0025	-0.0036
		SD	0.1836	0.0904	0.0845	0.0748	0.0974	0.0664	0.1072
		RMSE	0.1840	0.0905	0.0847	0.0754	0.0978	0.0664	0.1072
SIPWK	0.15	Bias	-0.0126	0.0352	0.0121	0.0163	-0.0309	-0.0178	-0.0068
		SD	0.1986	0.0974	0.0981	0.0869	0.1018	0.0711	0.1186
		RMSE	0.1989	0.1035	0.0988	0.0884	0.1063	0.0733	0.1187
SIPW	0.15	Bias	-0.0207	0.0074	0.0076	0.0135	-0.0098	0.0034	-0.0041
		SD	0.2033	0.1007	0.0972	0.084	0.1048	0.0731	0.1177
		RMSE	0.2043	0.1009	0.0974	0.0855	0.1052	0.0731	0.1177

gender are identified as the most influential factors in the decision to never use consultations with a non-physician health care professional. Medicaid recipients are more likely to forego a non-physician health care professional during an office visit. One explanation is that patients covered by Medicaid can

Table 2. (Contd.)

AFMD			$\hat{\gamma}_n$				$\hat{\alpha}_n$		
			$\hat{\gamma}_{1,n}$	$\hat{\gamma}_{2,n}$	$\hat{\gamma}_{3,n}$	$\hat{\gamma}_{4,n}$	$\hat{\alpha}_{1,n}$	$\hat{\alpha}_{2,n}$	$\hat{\alpha}_{3,n}$
SIPWK	0.30	Bias	-0.0082	0.0696	0.0224	0.0217	-0.0618	-0.0401	-0.0077
		SD	0.2219	0.1088	0.1147	0.1010	0.1112	0.0777	0.1366
		RMSE	0.2219	0.1291	0.1168	0.1032	0.1272	0.0875	0.1368
SIPW	0.30	Bias	-0.0266	0.0089	0.0110	0.0163	-0.0127	0.0045	-0.0012
		SD	0.2314	0.1143	0.1110	0.0973	0.1209	0.0818	0.1337
		RMSE	0.2328	0.1145	0.1115	0.0986	0.1215	0.0819	0.1337

limit their consultations to those that are necessary, i.e., not see a doctor, given that Medicaid is health insurance for the less well-off.

The probability of never using a doctor decreases with chronic, income, school, and age. The probability of never using a non-physician health care professional in a medical office decreases with the level of education because better-informed patients may tend to diversify their use of care. This probability decreases as health status worsens (in part because patients with worsening health status tend to favor visits to health professionals). This probability decreases with income because patients with higher incomes prefer to visit a health care professional.

The number of chronic illnesses and the level of education are the variables that most influence the average number consultations with non-physician healthcare professionals because patients with chronic conditions and those with higher levels of education visit regularly.

6. CONCLUSIONS

In this article, we have proposed a method for estimating the parameters of the MZIP model with MAR covariates. We compare the performance of this estimator with that of the kernel-assisted weighted estimator. The analysis of the numerical results concludes that the proposed $\hat{\theta}_n^{ws}$ estimator and the $\hat{\theta}_n^{wsk}$ estimator has a good performance. However, the simulation results suggest that the proposed method is more efficient than the kernel-assisted weighting method. The proposed SIPW estimator was used to analyze data from the U.S. public health economics NMES1988. The results of this analysis confirm the robustness of the proposed SIPW estimator.

In this paper, we assume that our data are MAR. But the missing data model is not monotonic in many practical situations. Adapting this approach to non-monotonic missing data in MZIP regression deserves further research.

Appendix A

PROOFS OF ASYMPTOTIC RESULTS

6.1. Proof of Theorem 1

We prove consistency of $\hat{\theta}_n^F$ by checking the conditions of the inverse function theorem of Foutz [13]. These conditions are proved in a series of technical lemmas.

Lemma 1. *As $n \rightarrow \infty$, $n^{-1/2}U_{F,n}(\theta_0)$ converges in probability to 0.*

Table 3. Simulation results for $n = 2000$, zero inflation: 41% (top) and 65% (bottom)

AFMD			$\hat{\gamma}_n$				$\hat{\alpha}_n$		
			$\hat{\gamma}_{1,n}$	$\hat{\gamma}_{2,n}$	$\hat{\gamma}_{3,n}$	$\hat{\gamma}_{4,n}$	$\hat{\alpha}_{1,n}$	$\hat{\alpha}_{2,n}$	$\hat{\alpha}_{3,n}$
FD	0.15	Bias	-0.0039	0.0006	0.0009	0.0005	0.0001	-0.0002	4e-04
		SD	0.0688	0.0259	0.0343	0.0231	0.0270	0.0154	0.0311
		RMSE	0.0689	0.0259	0.0343	0.0231	0.0269	0.0153	0.0311
SIPWK	0.15	Bias	0.0031	0.0046	0.0014	-0.0002	-0.0047	-0.0039	4e-04
		SD	0.0712	0.0275	0.0373	0.0255	0.0280	0.0162	0.0342
		RMSE	0.0712	0.0279	0.0373	0.0255	0.0283	0.0167	0.0342
SIPW	0.15	Bias	-0.0054	0.0009	0.0009	0.0009	0.0004	-0.0005	9e-04
		SD	0.0751	0.0279	0.0378	0.0254	0.0294	0.0165	0.0343
		RMSE	0.0752	0.0279	0.0378	0.0254	0.0294	0.0165	0.0343
SIPWK	0.30	Bias	-0.0434	0.0357	0.0090	0.0069	-0.0070	-0.0189	-0.0038
		SD	0.1963	0.0674	0.0829	0.0605	0.0633	0.0369	0.0850
		RMSE	0.2010	0.0762	0.0833	0.0608	0.0637	0.0415	0.0850
SIPW	0.30	Bias	-0.0238	0.0040	0.0058	0.0066	-0.0034	0.0014	0.0003
		SD	0.2082	0.0722	0.0833	0.0591	0.0686	0.0383	0.0848
		RMSE	0.2095	0.0723	0.0834	0.0594	0.0687	0.0383	0.0847
FD	0.15	Bias	-0.0120	0.0051	0.0006	0.0049	-0.0024	-0.0011	-0.0024
		SD	0.1316	0.0639	0.0584	0.0465	0.0699	0.0474	0.0735
		RMSE	0.1321	0.0640	0.0584	0.0468	0.0699	0.0474	0.0735
SIPWK	0.15	Bias	-0.0089	0.0269	0.0047	0.0072	-0.0166	-0.0184	-0.0050
		SD	0.1404	0.0677	0.0666	0.0532	0.0727	0.0503	0.0825
		RMSE	0.1406	0.0728	0.0667	0.0537	0.0746	0.0535	0.0826
SIPW	0.15	Bias	-0.0175	0.0064	0.0024	0.0065	-0.0011	-0.0017	-0.0027
		SD	0.1461	0.0708	0.0653	0.0526	0.0756	0.0525	0.0818
		RMSE	0.1470	0.0711	0.0653	0.0530	0.0756	0.0525	0.0818

Table 3. (Contd.)

AFMD			$\hat{\gamma}_n$				$\hat{\alpha}_n$		
			$\hat{\gamma}_{1,n}$	$\hat{\gamma}_{2,n}$	$\hat{\gamma}_{3,n}$	$\hat{\gamma}_{4,n}$	$\hat{\alpha}_{1,n}$	$\hat{\alpha}_{2,n}$	$\hat{\alpha}_{3,n}$
SIPWK	0.30	Bias	-0.0063	0.0539	0.0080	0.0115	-0.0366	-0.0377	-0.0098
		SD	0.1548	0.0727	0.0781	0.0626	0.0781	0.0534	0.0934
		RMSE	0.1549	0.0905	0.0784	0.0636	0.0862	0.0654	0.0939
SIPW	0.30	Bias	-0.0223	0.0082	0.0026	0.0086	-0.0018	-0.0013	-0.0043
		SD	0.1619	0.0774	0.0747	0.0613	0.0825	0.0571	0.0922
		RMSE	0.1633	0.0778	0.0747	0.0619	0.0825	0.0571	0.0922

Table 4. Analysis of health care data with 15% missing data

Variable		FD			SIPW		
		$\hat{\theta}_{F,n}$	S.E.	<i>P</i> -value	$\hat{\theta}_{W_s}$	S.E.	<i>P</i> -value
Intercept	$\hat{\gamma}_1$	2.3063	0.2732	2e-16	2.2755	0.2543	2e-16
Chronic	$\hat{\gamma}_2$	-0.1018	0.0224	5.58e-06	-0.0967	0.0237	4.44e-05
Medicaid	$\hat{\gamma}_3$	0.2980	0.0598	6.41e-07	0.3473	0.0623	2.49e-08
Age	$\hat{\gamma}_4$	-0.1650	0.0279	3.58e-09	-0.1692	0.0286	3.35e-09
Income	$\hat{\gamma}_5$	-0.0221	0.0066	0.000949	-0.0240	0.0075	0.0015
Gender	$\hat{\gamma}_6$	0.2836	0.0683	3.34e-05	0.2957	0.0709	3.08e-05
School	$\hat{\gamma}_7$	-0.0787	0.0089	2e-16	-0.0730	0.0090	5.59e-16
Intercept	$\hat{\alpha}_1$	-0.3475	0.1154	0.002602	-0.4896	0.1195	4.22e-05
Chronic	$\hat{\alpha}_2$	0.1043	0.0166	3.54e-10	0.1036	0.0180	8.91e-09
Gender	$\hat{\alpha}_3$	-0.1426	0.0513	0.005459	-0.1387	0.0546	0.0111
School	$\hat{\alpha}_4$	0.0782	0.0067	2e-16	0.0823	0.0070	2e-16

Proof of Lemma 1. Decompose $n^{-1/2}U_{F,n}(\theta_0)$ as, for every $i = 1, \dots, n$, we have

$$n^{-1/2}U_{F,n}(\theta_0) = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n Z_{i1} B_i(\theta_0) \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n Z_{iq} B_i(\theta_0) \\ \frac{1}{n} \sum_{i=1}^n X_{i1} A_i(\theta_0) \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n X_{ip} A_i(\theta_0) \end{pmatrix}.$$

Table 5. Analysis of health care data with 30% missing data

Variable		FD			SIPW		
		$\hat{\theta}_{F,n}$	S.E.	<i>P</i> -value	$\hat{\theta}_{W_s}$	S.E.	<i>P</i> -value
Intercept	$\hat{\gamma}_1$	2.3063	0.2732	2e-16	2.5796	0.3164	3.60e-16
Chronic	$\hat{\gamma}_2$	-0.1018	0.0224	5.58e-06	-0.1158	0.0261	9.75e-06
Medicaid	$\hat{\gamma}_3$	0.2980	0.0598	6.41e-07	0.3153	0.0717	1.10e-05
Age	$\hat{\gamma}_4$	-0.1650	0.0279	3.58e-09	-0.1755	0.0326	7.57e-08
Income	$\hat{\gamma}_5$	-0.0221	0.0066	0.000949	-0.0140	0.0073	0.05381
Gender	$\hat{\gamma}_6$	0.2836	0.0683	3.34e-05	0.2647	0.0745	0.00038
School	$\hat{\gamma}_7$	-0.0787	0.0089	2e-16	-0.0817	0.0103	2.52e-15
Intercept	$\hat{\alpha}_1$	-0.3475	0.1154	0.002602	-0.6187	0.1356	5.12e-06
Chronic	$\hat{\alpha}_2$	0.1043	0.0166	3.54e-10	0.1113	0.0200	2.70e-08
Gender	$\hat{\alpha}_3$	-0.1426	0.0513	0.005459	-0.1387	0.0586	0.01790
School	$\hat{\alpha}_4$	0.0782	0.0067	2e-16	0.0832	0.0081	2e-16

For $i = 1, \dots, n$ and $l = 1, \dots, q$;

$$\mathbb{E}[Z_{il}B_i(\theta_0)] = \mathbb{E}[\mathbb{E}[Z_{il}B_i(\theta_0)|\mathbf{X}_i, \mathbf{Z}_i]] = \mathbb{E}[Z_{il}\mathbb{E}[B_i(\theta_0)|\mathbf{X}_i, \mathbf{Z}_i]].$$

We have

$$\mathbb{E}[B_i(\theta_0)|\mathbf{X}_i, \mathbf{Z}_i] = \frac{\mathbb{E}(J_i|\mathbf{X}_i, \mathbf{Z}_i)e^{\mathbf{Z}_i^T \gamma_0} (e^{h_i(\theta_0)} - e^{\mathbf{X}_i^T \alpha_0})}{e^{\mathbf{Z}_i^T \gamma_0 + h_i(\theta_0)} + 1} + \frac{e^{\mathbf{Z}_i^T \gamma_0} (\mathbb{E}(Y_i|\mathbf{X}_i, \mathbf{Z}_i) - 1)}{1 + e^{\mathbf{Z}_i^T \gamma_0}} - [1 - \mathbb{E}(J_i|\mathbf{X}_i, \mathbf{Z}_i)] e^{\mathbf{X}_i^T \alpha_0 + \mathbf{Z}_i^T \gamma_0}.$$

Now, we have

$\mathbb{E}(J_i|\mathbf{X}_i, \mathbf{Z}_i) = \mathbb{P}(Y_i = 0|\mathbf{X}_i, \mathbf{Z}_i)$, $\mathbb{E}(Y_i|\mathbf{X}_i, \mathbf{Z}_i) = \nu_i$ and $\mathbb{E}(1 - J_i|\mathbf{X}_i, \mathbf{Z}_i) = \mathbb{P}(Y_i > 0|\mathbf{X}_i, \mathbf{Z}_i)$. It follows that $\mathbb{E}[Z_{il}B_i(\theta_0)] = 0$.

Using similarly arguments we prove that, for every $i = 1, \dots, n$ and $j = 1, \dots, p$, $\mathbb{E}[X_{ij}A_i(\theta_0)] = 0$.

Now, for every $i = 1, \dots, n$ and $l = 1, \dots, q$, we have

$$\text{var}(Z_{il}B_i(\theta_0)) \leq \mathbb{E}(Z_{il}^2 B_i^2(\theta_0)).$$

By **H3**, we have $\mathbb{E}(Z_{il}^2 B_i^2(\theta_0)) < \infty$.

Using similar arguments, we prove $\text{var}(X_{ij}A_i(\theta_0)) < \infty$ for every $i = 1, \dots, n$ and $j = 1, \dots, p$.

Thus, by the weak law of large numbers, $n^{-1/2}U_{F,n}(\theta_0)$ converges in probability to 0, which concludes the proof.

Lemma 2. As $n \rightarrow \infty$, $n^{-1/2} \frac{\partial U_{F,n}(\theta)}{\partial \theta^T}$ converges in probability to a fixed function $-\Sigma(\theta)$, uniformly in an open neighbourhood of θ_0 .

Proof of Lemma 2: Let $\tilde{U}_{F,n}(\theta) := n^{-1/2} \frac{\partial U_{F,n}(\theta)}{\partial \theta^T}$, and ν_{θ_0} be an open neighbourhood of θ_0 . Let $\theta \in \nu_{\theta_0}$.

By the weak law of large numbers and **H3**, $\tilde{U}_{F,n}(\theta) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\partial^2 l_i(\theta)}{\partial \theta \partial \theta^T} \right\}$ converges in probability to the matrix $-\Sigma(\theta)$ as $n \rightarrow \infty$, where $\Sigma(\theta) = \mathbb{E} \left[-\frac{\partial^2 l_1(\theta)}{\partial \theta \partial \theta^T} \right]$.

By conditions **H4**, we prove that the convergence of $\tilde{U}_{F,n}(\theta)$ to $-\Sigma(\theta)$ is uniform on ν_{θ_0} .

The conditions inverse function theorem of Foutz [13] are verified. Finally $\hat{\theta}_n$ converges in probability to θ_0 .

Now, we prove that $\hat{\theta}_n^F$ is asymptotically Gaussian. To do this, it follows by a Taylor's expansion of $U_{F,n}(\hat{\theta}_{F,n})$ at θ_0 yields

$$0 = U_{F,n}(\theta_0) + \tilde{U}_{F,n}(\theta_0)\sqrt{n}(\hat{\theta}_n^F - \theta_0) + o_p(1)$$

By calculations $\text{var}(U_{F,n}(\theta_0)) = \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left(\dot{l}_i(\theta_0)\dot{l}_i(\theta_0)^T \right) = Q_F(\theta_0)$.

Finally, by Lemma 2 and Slutsky's theorem, $\sqrt{n}(\hat{\theta}_n^F - \theta_0)$ converges in distribution to the Gaussian vector of mean zero and variance Δ_F , where Δ_F is defined in Theorem 1.

Appendix B

6.2. Proof of Theorem 2

We prove consistency of $\hat{\theta}_n^{ws}$ by checking the conditions of the inverse function theorem of Foutz [13]. These conditions are proved in a series of technical lemmas.

Lemma 3. As $n \rightarrow \infty$, $n^{-1/2}U_{w,n}(\theta_0, \hat{\pi})$ converges in probability to 0.

Proof of Lemma 3. We decompose $n^{-1/2}U_{w,n}(\theta_0, \hat{\pi})$ as

$$n^{-1/2}U_{ws,n}(\theta_0, \hat{\pi}) = (n^{-1/2}U_{ws,n}(\theta_0, \hat{\pi}) - n^{-1/2}U_{ws,n}(\theta_0, \pi)) + n^{-1/2}U_{ws,n}(\theta_0, \pi). \tag{6.1}$$

Considering the first term of this decomposition.

Let $\mathbf{S}'_i = (\mathbf{S}_i^D, \mathbf{S}'_i^D)$ and $G_n(\theta_0, \pi) = n^{-1/2}U_{ws,n}(\theta_0, \hat{\pi}) - n^{-1/2}U_{ws,n}(\theta_0, \pi)$, we have

$$\begin{aligned} G_n(\theta_0, \pi) &= \frac{1}{n} \sum_{i=1}^n \Delta_i \left(\frac{1}{\hat{\pi}(Y_i, \mathbf{S}'_i)} - \frac{1}{\pi(Y_i, \mathbf{S}'_i)} \right) \dot{l}_i(\theta_0), \\ &= \frac{1}{n} \sum_{i=1}^n \Delta_i \left[\frac{\hat{\pi}(Y_i, \mathbf{S}'_i) - \pi(Y_i, \mathbf{S}'_i)}{\pi^2(Y_i, \mathbf{S}'_i)} + O_P((\hat{\pi}(Y_i, \mathbf{S}'_i) - \pi(Y_i, \mathbf{S}'_i))^2) \right] \dot{l}_i(\theta_0), \\ &= \frac{1}{n} \sum_{i=1}^n \Delta_i \left[\frac{\frac{\sum_{k=1}^n \Delta_k I(Y_k=y, \mathbf{S}'_k=s')}{\sum_{i=1}^n I(Y_i=y, \mathbf{S}'_i=s')} - \pi(Y_i, \mathbf{S}'_i)}{\pi^2(Y_i, \mathbf{S}'_i)} \right] \dot{l}_i(\theta_0) + o_p^* \left(\frac{1}{\sqrt{n}} \right), \\ &= \frac{1}{n} \sum_{i=1}^n \Delta_i \left[\frac{\frac{1}{n} \sum_{i=1}^n [\Delta_k - \pi(Y_i, \mathbf{S}'_i)] I(Y_k = Y_i, \mathbf{S}'_k = \mathbf{S}'_i)}{\pi^2(Y_i, \mathbf{S}'_i) P(Y = Y_i, \mathbf{S}' = \mathbf{S}'_i)} + O_p \left(\frac{1}{n} \right) \right] \\ &\quad \times \dot{l}_i(\theta_0) + o_p^* \left(\frac{1}{\sqrt{n}} \right), \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \left[\frac{[\Delta_i - \pi(Y_i, \mathbf{S}'_i)] [\Delta_k - \pi(Y_i, \mathbf{S}'_i)] I(Y_k = Y_i, \mathbf{S}'_k = \mathbf{S}'_i)}{\pi^2(Y_i, \mathbf{S}'_i) P(Y = Y_i, \mathbf{S}' = \mathbf{S}'_i)} \right] \dot{l}_i(\theta_0) \\ &\quad + \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \left[\frac{[\Delta_k - \pi(Y_i, \mathbf{S}'_i)] I(Y_k = Y_i, \mathbf{S}'_k = \mathbf{S}'_i)}{\pi(Y_i, \mathbf{S}'_i) P(Y = Y_i, \mathbf{S}' = \mathbf{S}'_i)} \right] \dot{l}_i(\theta_0) + o_p^* \left(\frac{1}{\sqrt{n}} \right), \end{aligned}$$

where $o_p^*(a_n)$ denotes a matrix whose components are uniformly $o_p(a_n)$. By the weak law of large numbers we have

$$\frac{1}{n} \sum_{k=1}^n \left[\frac{[\Delta_k - \pi(Y_i, \mathbf{S}'_i)] I(Y_k = Y_i, \mathbf{S}'_k = \mathbf{S}'_i)}{\pi(Y_i, \mathbf{S}'_i) P(Y = Y_i, \mathbf{S}' = \mathbf{S}'_i)} \right]$$

converges in probability to 0 as $n \rightarrow \infty$.

Using conditions **H3**, we prove that $\dot{l}_i(\theta_0)$ is finite a.s. Finally, by Slutsky's theorem

$$G_n(\theta_0, \pi) = \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \left[\frac{[\Delta_i - \pi(Y_i, \mathbf{S}'_i)] [\Delta_k - \pi(Y_i, \mathbf{S}'_i)] I(Y_k = Y_i, \mathbf{S}'_k = \mathbf{S}'_i)}{\pi_i^2(Y_i, \mathbf{S}'_i) P(Y = Y_i, \mathbf{S}' = \mathbf{S}'_i)} \right] \\ \times \dot{l}_i(\theta_0) + \frac{1}{n^2} \sum_{i=1}^n \sum_{k=1}^n \left[\frac{[\Delta_k - \pi(Y_i, \mathbf{S}'_i)] I(Y_k = Y_i, \mathbf{S}'_k = \mathbf{S}'_i)}{\pi(Y_i, \mathbf{S}'_i) P(Y = Y_i, \mathbf{S}' = \mathbf{S}'_i)} \right] \dot{l}_i(\theta_0)$$

converges in probability to 0 as $n \rightarrow \infty$.

Next, consider the term $n^{-1/2}U_{ws,n}(\theta_0, \pi(Y_i, \mathbf{S}'_i))$ in decomposition (6.1).

We show that $n^{-1/2}U_{ws,n}(\theta_0, \pi(Y_i, \mathbf{S}'_i))$ converges in probability to 0 as $n \rightarrow \infty$.

For every $i = 1, \dots, n$, we have

$$n^{-1/2}U_{ws,n}(\theta_0, \pi(Y_i, \mathbf{S}'_i)) = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n \frac{\Delta_i}{\pi(Y_i, \mathbf{S}'_i)} Z_{i1} B_i(\theta_0) \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n \frac{\Delta_i}{\pi(Y_i, \mathbf{S}'_i)} Z_{iq} B_i(\theta_0) \\ \frac{1}{n} \sum_{i=1}^n \frac{\Delta_i}{\pi(Y_i, \mathbf{S}'_i)} X_{i1} A_i(\theta_0) \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n \frac{\Delta_i}{\pi(Y_i, \mathbf{S}'_i)} X_{ip} A_i(\theta_0) \end{pmatrix}.$$

For $i = 1, \dots, n$ and $l = 1, \dots, q$;

$$\mathbb{E} \left[\frac{\Delta_i}{\pi(Y_i, \mathbf{S}'_i)} Z_{il} B_i(\theta_0) \right] = \mathbb{E} \left[\mathbb{E} \left[\frac{\Delta_i}{\pi(Y_i, \mathbf{S}'_i)} Z_{i1} B_i(\theta_0) | Y_i, \mathbf{S}'_i \right] \right].$$

Two cases should be considered, namely: (i) Z_{il} is a component of Z^{obs} and (ii) Z_{il} is a component of Z^{miss} . In case (i), we have

$$\mathbb{E} \left[\mathbb{E} \left[\frac{\Delta_i}{\pi(Y_i, \mathbf{S}'_i)} Z_{il} B_i(\theta_0) | Y_i, \mathbf{S}'_i \right] \right] = \mathbb{E} \left[\frac{1}{\pi_i(Y_i, \mathbf{S}'_i)} Z_{il} \mathbb{E}[\Delta_i B_i(\theta_0) | Y_i, \mathbf{S}'_i] \right].$$

Given $\mathbf{V}_i = (Y_i, \mathbf{S}'_i)$, $Z_{il} B_i(\theta_0)$ is a function of $(\mathbf{X}^{\text{miss}}, \mathbf{Z}^{\text{miss}})$ only. Thus, by the MAR assumption, $B_i(\theta_0)$ and Δ_i are independent

$$\mathbb{E} \left[\frac{1}{\pi_i(Y_i, \mathbf{S}'_i)} Z_{il} \mathbb{E}[\Delta_i B_i(\theta_0) | \mathbf{V}_i] \right] = \mathbb{E} \left[\frac{1}{\pi_i(Y_i, \mathbf{S}'_i)} Z_{il} \mathbb{E}[\Delta_i | \mathbf{V}_i] \mathbb{E}[B_i(\theta_0) | \mathbf{V}_i] \right], \\ = \mathbb{E} [Z_{il} \mathbb{E}[B_i(\theta_0) | \mathbf{V}_i]], \\ = \mathbb{E} [Z_{il} B_i(\theta_0)], \\ = 0.$$

In case (ii),

$$\mathbb{E} \left[\mathbb{E} \left[\frac{\Delta_i}{\pi_i(Y_i, \mathbf{S}'_i)} Z_{i1} B_i(\theta_0) | \mathbf{V}_i \right] \right] = \mathbb{E} \left[\frac{1}{\pi_i(Y_i, \mathbf{S}'_i)} \mathbb{E}[\Delta_i Z_{il} B_i(\theta_0) | \mathbf{V}_i] \right].$$

Given \mathbf{V}_i , $Z_{il} B_i(\theta_0)$ is a function of $(\mathbf{X}^{\text{miss}}, \mathbf{Z}^{\text{miss}})$ only. Thus, by the MAR assumption, $B_i(\theta_0)$ and Δ_i are independent

$$\mathbb{E} \left[\frac{1}{\pi_i(Y_i, \mathbf{S}'_i)} \mathbb{E}[\Delta_i Z_{il} B_i(\theta_0) | \mathbf{V}_i] \right] = \mathbb{E} \left[\frac{1}{\pi_i(Y_i, \mathbf{S}'_i)} \mathbb{E}[\Delta_i | \mathbf{V}_i] \mathbb{E}[Z_{il} B_i(\theta_0) | \mathbf{V}_i] \right],$$

$$= \mathbb{E}[Z_{il}B_i(\theta_0)],$$

$$= 0.$$

It follows that $\mathbb{E}[\frac{\Delta_i}{\pi_i(Y_i, \mathbf{S}'_i)} Z_{il}B_i(\theta_0)] = 0$.

Using similar arguments, we prove that $\mathbb{E}[\frac{\Delta_i}{\pi_i(Y_i, \mathbf{S}'_i)} X_{ij}A_i(\theta_0)] = 0$.

Now, for every $i = 1, \dots, n$ and $l = 1, \dots, q$, we have

$$\text{var} \left(\frac{\Delta_i}{\pi_i(Y_i, \mathbf{S}'_i)} Z_{il}B_i(\theta_0) \right) \leq \mathbb{E} \left(\frac{\Delta_i}{\pi_i^2(Y_i, \mathbf{S}'_i)} Z_{il}^2 B_i^2(\theta_0) \right).$$

By **H3**, we have $\mathbb{E} \left(\frac{\Delta_i}{\pi_i^2(Y_i, \mathbf{S}'_i)} Z_{il}^2 B_i^2(\theta_0) \right) < \infty$.

Using similar arguments, we prove

$$\text{var} \left(\frac{\Delta_i}{\pi_i(Y_i, \mathbf{S}'_i)} X_{ij}A_i(\theta_0) \right) < \infty \text{ for every } i = 1, \dots, n \text{ and } j = 1, \dots, p.$$

Thus, by the weak law of large numbers, $n^{-1/2}U_{ws,n}(\theta_0, \pi(Y_i, \mathbf{S}'_i))$ converges in probability to 0 as $n \rightarrow \infty$.

Finally $n^{-1/2}U_{w,n}(\theta_0, \hat{\pi}(Y_i, \mathbf{S}'_i))$ converges to 0, which concludes the proof.

Lemma 4. As $n \rightarrow \infty$, $n^{-1/2} \frac{\partial U_{ws,n}(\theta, \hat{\pi})}{\partial \theta^T}$ converges in probability to a fixed function $-\Sigma(\theta)$, uniformly in a neighbourhood of θ_0 .

Proof of Lemma 4. Let $\bar{U}_{ws,n}(\theta, \hat{\pi}) := n^{-1/2} \frac{\partial U_{ws,n}(\theta, \pi)}{\partial \theta^T}$ and $\ddot{l}_i(\theta) = \frac{\partial^2 l_i(\theta)}{\partial \theta \partial \theta^T}$. We have

$$\bar{U}_{ws,n}(\theta, \hat{\pi}) = [\bar{U}_{ws,n}(\theta, \hat{\pi}) - \bar{U}_{ws,n}(\theta, \pi)] + \bar{U}_{ws,n}(\theta, \pi).$$

Using similiary argument in Lemma 4, we have $\bar{U}_{ws,n}(\theta, \hat{\pi}) - \bar{U}_{ws,n}(\theta, \pi)$ converges in probability to 0. By the weak law of large numbers, and **H3**

$$\bar{U}_{ws,n}(\theta, \pi) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\Delta_i}{\pi(Y_i, \mathbf{S}'_i)} \ddot{l}_i(\theta) \right\}$$

converges in probability to the matrix $-\Sigma(\theta)$ as $n \rightarrow \infty$.

By **H5**, we prove that the convergence of $\bar{U}_{ws,n}(\theta, \hat{\pi})$ to $-\Sigma(\theta)$ is uniform.

The conditions inverse function Foutz [13] are verified. Finally $\hat{\theta}_n^{ws}$ converges in probability to θ_0 .

Now, we prove that θ_n^{ws} is asymptotically Gaussian.

It follows by a Taylor's expansion of $U_{ws,n}(\hat{\theta}_n^{ws}, \hat{\pi})$ at $(\theta_0, \hat{\pi})$ yields

$$0 = U_{ws,n}(\hat{\theta}_n^{ws}, \hat{\pi}) = U_{ws,n}(\theta_0, \hat{\pi}) + \bar{U}_{ws,n}(\theta_0, \hat{\pi})\sqrt{n}(\hat{\theta}_n^{ws} - \theta_0) + o_p(1),$$

therefore

$$\sqrt{n}(\hat{\theta}_n^{ws} - \theta_0) = -[\bar{U}_{ws,n}(\theta_0, \hat{\pi})]^{-1} U_{ws,n}(\theta_0, \hat{\pi}) + o_p(1),$$

thus

$$\sqrt{n}(\hat{\theta}_n^{ws} - \theta_0) = \Sigma^{-1}(\theta_0)U_{ws,n}(\theta_0, \hat{\pi}) + [-\bar{U}_{ws,n}^{-1}(\theta_0, \hat{\pi}) - \Sigma^{-1}(\theta_0)] U_{ws,n}(\theta_0, \hat{\pi}) + o_p(1).$$

By calculations,

$$\begin{aligned} \text{Var} [U_{ws,n}(\theta_0, \hat{\pi})] &= \text{Var} \{ U_{ws,n}(\theta_0, \pi) + [U_{ws,n}(\theta_0, \hat{\pi}) - U_{ws,n}(\theta_0, \pi)] \}, \\ &= \text{Var} [U_{ws,n}(\theta_0, \pi)] + \text{Var} [U_{ws,n}(\theta_0, \hat{\pi}) - U_{ws,n}(\theta_0, \pi)] \\ &\quad + 2\text{Cov} [U_{ws,n}(\theta_0, \pi), U_{ws,n}(\theta_0, \hat{\pi}) - U_{ws,n}(\theta_0, \pi)], \end{aligned}$$

$$\text{Var} [U_{ws,n}(\theta_0, \pi)] = \mathbb{E} \left[\frac{\ddot{l}_i(\theta_0, \pi)\ddot{l}_i(\theta_0, \pi)^T}{\pi(Y_i, \mathbf{S}'_i)} \right] = \Omega_3(\theta_0, \pi).$$

Let $H(\theta_0, \pi) = U_{ws,n}(\theta_0, \hat{\pi}) - U_{ws,n}(\theta_0, \pi)$

$$\begin{aligned} H(\theta_0, \pi) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Delta_i \left(\frac{1}{\hat{\pi}(Y_i, \mathbf{S}'_i)} - \frac{1}{\pi(Y_i, \mathbf{S}'_i)} \right) \dot{l}_i(\theta_0), \\ &= -\frac{1}{\sqrt{n^3}} \sum_{i=1}^n \sum_{k=1}^n \left[\frac{[\Delta_i - \pi(Y_i, \mathbf{S}'_i)] [\Delta_k - \pi(Y_i, \mathbf{S}'_i)] I(Y_k = Y_i, \mathbf{S}'_k = \mathbf{S}'_i)}{\pi^2(Y_i, \mathbf{S}'_i) P(Y = Y_i, \mathbf{S}' = \mathbf{S}'_i)} \right] \dot{l}_i(\theta_0) \\ &\quad - \frac{1}{\sqrt{n^3}} \sum_{i=1}^n \sum_{k=1}^n \left[\frac{[\Delta_k - \pi(Y_i, \mathbf{S}'_i)] I(Y_k = Y_i, \mathbf{S}'_k = \mathbf{S}'_i)}{\pi(Y_i, \mathbf{S}'_i) P(Y = Y_i, \mathbf{S}' = \mathbf{S}'_i)} \right] \dot{l}_i(\theta_0) + o_p(1), \\ &= -\mathbf{Q}_{1n} - \mathbf{Q}_{2n} + o_p(1), \end{aligned}$$

where $o_p(a_n)$ denotes a column vector whose components are uniformly $o_p(a_n)$.

$$\begin{aligned} \mathbf{Q}_{1n} &= \frac{1}{\sqrt{n^3}} \sum_{i=1}^n \sum_{k=1}^n \left[\frac{[\Delta_i - \pi(Y_i, \mathbf{S}'_i)] [\Delta_k - \pi(Y_i, \mathbf{S}'_i)] I(Y_k = Y_i, \mathbf{S}'_k = \mathbf{S}'_i)}{\pi^2(Y_i, \mathbf{S}'_i) P(Y = Y_i, \mathbf{S}' = \mathbf{S}'_i)} \right] \dot{l}_i(\theta_0). \\ \mathbf{Q}_{2n} &= \frac{1}{\sqrt{n^3}} \sum_{i=1}^n \sum_{k=1}^n \left[\frac{[\Delta_k - \pi(Y_i, \mathbf{S}'_i)] I(Y_k = Y_i, \mathbf{S}'_k = \mathbf{S}'_i)}{\pi(Y_i, \mathbf{S}'_i) P(Y = Y_i, \mathbf{S}' = \mathbf{S}'_i)} \right] \dot{l}_i(\theta_0). \end{aligned}$$

Let

$$P_{ik} = \left[\frac{[\Delta_i - \pi(Y_i, \mathbf{S}'_i)] [\Delta_k - \pi(Y_i, \mathbf{S}'_i)] I(Y_k = Y_i, \mathbf{S}'_k = \mathbf{S}'_i)}{\pi^2(Y_i, \mathbf{S}'_i) P(Y = Y_i, \mathbf{S}' = \mathbf{S}'_i)} \right] \dot{l}_i(\theta_0).$$

In order to show that $\mathbf{Q}_{1n} = \mathbf{O}_p(1/\sqrt{n})$, $\mathbb{E}(\mathbf{Q}_{1n}) = \mathbf{O}_p(1/\sqrt{n})$ and $\text{Var}(\mathbf{Q}_{1n}) = \mathbf{O}_p^*(1/n)$ where $\mathbf{O}^*(a_n)$ and $\mathbf{O}(a_n)$ denote a matrix and column vector whose components are uniformly $\mathbf{O}(a_n)$. It first can be shown that

$$\mathbb{E}[P_{ik}] = \mathbb{E}[\mathbb{E}(P_{ik} | Y_k = Y_i, \mathbf{S}'_k = \mathbf{S}'_i)] = \begin{cases} 0 & \text{if } i \neq k \\ \mathbb{E} \left[\frac{[1 - \pi(Y_i, \mathbf{S}'_i)] \dot{l}_i(\theta_0)}{\pi_i(Y_i, \mathbf{S}'_i)} \right] & \text{if } i = k, \end{cases}$$

and then

$$\begin{aligned} \mathbb{E} \left[\frac{[1 - \pi(Y_i, \mathbf{S}'_i)] \dot{l}_i(\theta_0)}{\pi(Y_i, \mathbf{S}'_i)} \right] &= \mathbb{E} \left\{ \mathbb{E} \left[\frac{[1 - \pi(Y_i, \mathbf{S}'_i)] \dot{l}_i(\theta_0)}{\pi(Y_i, \mathbf{S}'_i)} \mid Y_i, \mathbf{S}'_i \right] \right\}, \\ &= \mathbb{E} \left[\frac{[1 - \pi(Y_i, \mathbf{S}'_i)] \dot{l}_i^*(\theta_0)}{\pi(Y_i, \mathbf{S}'_i)} \right]. \end{aligned}$$

Thus, we have

$$\mathbb{E}(\mathbf{Q}_{1n}) = \frac{1}{n^{3/2}} \sum_{k=1}^n \sum_{i=1}^n \mathbb{E}[P_{ik}] = \frac{1}{n^{3/2}} \sum_{i=1}^n \mathbb{E} \left[\frac{[1 - \pi(Y_i, \mathbf{S}'_i)] \dot{l}_i^*(\theta_0)}{\pi(Y_i, \mathbf{S}'_i)} \right] = \mathbf{O} \left(\frac{1}{\sqrt{n}} \right).$$

We have

$$\begin{aligned} \text{Cov}(P_{ij}, P_{kl}) &= \mathbb{E} \left\{ \left[\frac{[\Delta_i - \pi(Y_i, \mathbf{S}'_i)] [\Delta_j - \pi(Y_j, \mathbf{S}'_j)] I(Y_j = Y_i, \mathbf{S}'_j = \mathbf{S}'_i)}{\pi^2(Y_i, \mathbf{S}'_i) P(Y = Y_i, \mathbf{S}' = \mathbf{S}'_i)} \right] \dot{l}_i(\theta_0) \right. \\ &\quad \times \left. \left[\frac{[\Delta_k - \pi(Y_k, \mathbf{S}'_k)] [\Delta_l - \pi(Y_l, \mathbf{S}'_l)] I(Y_l = Y_k, \mathbf{S}'_l = \mathbf{S}'_k)}{\pi^2(Y_k, \mathbf{S}'_k) P(Y = Y_k, \mathbf{S}' = \mathbf{S}'_k)} \right] \dot{l}_k(\theta_0)^T \right\}, \\ \text{Cov}(P_{ij}, P_{lk}) &= \begin{cases} 0 & \text{if } k \neq i, j \text{ and } l \neq i, j \\ 0 & \text{if } k \neq i, j \text{ and } l = i \text{ or } j \\ \mathbb{E} \left[\frac{[1 - \pi(Y_i, \mathbf{S}'_i)]^2 \dot{l}_i^*(\theta_0) \dot{l}_i^*(\theta_0)^T}{\pi^2(Y_i, \mathbf{S}'_i)} \right] & \text{if } i = l \text{ and } j = k \end{cases} \end{aligned}$$

and

$$\begin{aligned} \text{Var}(\mathbf{Q}_{1n}) &= \frac{1}{n^3} \left\{ \sum_{i,j} \text{Var}(P_{ij}) + \sum_{i,j} \left[\sum_{l=j, k \neq j} \text{Cov}(P_{ij}, P_{lk}) \right. \right. \\ &\quad \left. \left. + \sum_{k=j, l \neq j} \text{Cov}(P_{ij}, P_{lk}) + \sum_{k \neq j, l \neq j} \text{Cov}(P_{ij}, P_{lk}) \right] \right\}, \\ &= \frac{1}{n^3} \left\{ n \mathbb{E} \left\{ \frac{\dot{l}_i^*(\theta_0) \dot{l}_i^*(\theta_0)^T}{\pi^4(Y_i, \mathbf{S}'_i)} \left[\pi(Y_i, \mathbf{S}'_i) - 4\pi^4(Y_i, \mathbf{S}'_i) \right. \right. \right. \\ &\quad \left. \left. \left. + 6\pi^3(Y_i, \mathbf{S}'_i) - 3\pi^4(Y_i, \mathbf{S}'_i) \right] \right\} \right. \\ &\quad \left. + n(n-1) \mathbb{E} \left\{ \frac{\dot{l}_i^*(\theta_0) \dot{l}_i^*(\theta_0)^T}{\pi^4(Y_i, \mathbf{S}'_i)} [1 - \pi(Y_i, \mathbf{S}'_i)]^2 \right\} \right. \\ &\quad \left. + n(n-1) \mathbb{E} \left\{ \frac{\dot{l}_i^*(\theta_0) \dot{l}_i^*(\theta_0)^T}{\pi^4(Y_i, \mathbf{S}'_i)} [1 - \pi(Y_i, \mathbf{S}'_i)]^2 \right\} \right\}, \\ &= O^*\left(\frac{1}{n^2}\right) + O^*\left(\frac{1}{n}\right) + O^*\left(\frac{1}{n}\right) \\ &= O^*\left(\frac{1}{n}\right). \end{aligned}$$

Therefore, $\mathbf{Q}_{1n} = O_p\left(\frac{1}{\sqrt{n}}\right)$, \mathbf{Q}_{2n} can be expressed as follows:

$$\begin{aligned} \mathbf{Q}_{2n} &= \frac{1}{\sqrt{n^3}} \sum_{i=1}^n \sum_{k=1}^n \left[\frac{[\Delta_k - \pi(Y_i, \mathbf{S}'_i)] I(Y_k = Y_i, \mathbf{S}'_k = \mathbf{S}'_i)}{\pi(Y_i, \mathbf{S}'_i) P(Y = Y_i, \mathbf{S}' = \mathbf{S}'_i)} \right] \dot{l}_i(\theta_0), \\ &= \frac{1}{\sqrt{n}} \sum_{k=1}^n \left[\frac{\Delta_k - \pi(Y_k, \mathbf{S}'_k)}{\pi(Y_k, \mathbf{S}'_k)} \right] \left[\dot{l}_k^*(\theta_0) + \frac{\frac{1}{n} \sum_{i=1}^n I(Y_i = Y_k, \mathbf{S}'_i = \mathbf{S}'_k) \dot{l}_i(\theta_0)}{P(Y = Y_k, \mathbf{S}' = \mathbf{S}'_k)} - \dot{l}_k^*(\theta_0) \right], \\ &= \frac{1}{\sqrt{n}} \sum_{k=1}^n \left[\frac{\Delta_k - \pi(Y_k, \mathbf{S}'_k)}{\pi(Y_k, \mathbf{S}'_k)} \right] \dot{l}_k^*(\theta_0) + \frac{1}{\sqrt{n}} \sum_{k=1}^n \Phi_k \left[\frac{1}{n} \sum_{i=1}^n \Psi_{ik}(\theta_0) \right], \end{aligned}$$

where $\Phi_k = \frac{\Delta_k - \pi(Y_k, \mathbf{S}'_k)}{\pi(Y_k, \mathbf{S}'_k)}$ and

$$\Psi_{ik}(\theta_0) = \frac{I(Y_i = Y_k, \mathbf{S}'_i = \mathbf{S}'_k) \dot{l}_i(\theta_0) - P(Y = Y_k, \mathbf{S}' = \mathbf{S}'_k) \dot{l}_k^*(\theta_0)}{P(Y = Y_k, \mathbf{S}' = \mathbf{S}'_k)}.$$

We have $\mathbb{E}[\Psi_{ik}(\theta_0) | Y_i = Y_k, \mathbf{S}'_i = \mathbf{S}'_k] = 0$ and, hence,

$$\begin{aligned} \mathbb{E} \left[\Phi_k \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_{ik}(\theta_0) \right) \right] &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E} \left[\mathbb{E}(\Phi_k \Psi_{ik}(\theta_0) | Y_i = Y_k, \mathbf{S}'_i = \mathbf{S}'_k) \right], \\ &= 0. \end{aligned}$$

Let $\Psi_{iks}(\theta_0)$ be the s th element of $\Psi_{ik}(\theta_0)$. Then, by Cauchy–Schwarz’s inequality,

$$\mathbb{E} \left[\left| \Phi_k \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_{iks}(\theta_0) \right) \right| \right] \leq [\mathbb{E}(\Phi_k^2)]^{1/2} \left\{ \mathbb{E} \left[\frac{1}{n} \left(\sum_{i=1}^n \Psi_{iks}(\theta_0) \right)^2 \right] \right\}^{1/2}.$$

Because for each element of $\Psi_{ik}(\theta_0)$

$$\mathbb{E} \left[\frac{1}{n} \left(\sum_{i=1}^n \Psi_{iks}(\theta_0) \right)^2 \right] = \frac{1}{n} \left[\mathbb{E} \left(\sum_{i=1}^n \Psi_{iks}^2(\theta_0) \right) + \sum_{i=1}^n \sum_{j=1, j \neq i}^n \mathbb{E}(\Psi_{iks}(\theta_0) \Psi_{jks}(\theta_0)) \right],$$

$$= \mathbb{E}(\Psi_{iks}^2(\theta_0)) < \infty,$$

we can prove $\mathbb{E} \left[\left| \Phi_k \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_{ik}(\theta_0) \right) \right| \right] < \infty$.

By the weak law of large numbers $\frac{1}{n} \sum_{k=1}^n \left\{ \Phi_k \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_{ik}(\theta_0) \right] \right\} = o_p(1)$. Hence, \mathbf{Q}_{2n} can be expressed as $\mathbf{Q}_{2n} = \frac{1}{\sqrt{n}} \sum_{k=1}^n \left[\frac{\Delta_k - \pi(Y_k, \mathbf{S}'_k)}{\pi(Y_k, \mathbf{S}'_k)} \right] i_k^*(\theta_0 + o_p(1))$.

$$\begin{aligned} \text{Var} [U_{ws,n}(\theta_0, \hat{\pi}) - U_{ws,n}(\theta_0, \pi)] &= \mathbb{E} \left\{ \mathbb{E} \left[\frac{[\Delta_i - \pi(Y_i, \mathbf{S}'_i)]^2 [i_i^*(\theta_0) i_i^*(\theta_0)^T]}{\pi^2(Y_i, \mathbf{S}'_i)} \mid Y_i, \mathbf{S}'_i \right] \right\} \\ &\quad + o^*(1), \\ &= \mathbb{E} \left[\frac{[1 - \pi(Y_i, \mathbf{S}'_i)] [i_i^*(\theta_0) i_i^*(\theta_0)^T]}{\pi(Y_i, \mathbf{S}'_i)} \right] + o^*(1) \end{aligned}$$

and let $\Sigma = \text{Cov} [U_{ws,n}(\theta_0, \pi), U_{ws,n}(\theta_0, \hat{\pi}) - U_{ws,n}(\theta_0, \pi)]$, we have

$$\begin{aligned} \Sigma &= -\mathbb{E} \left[\frac{\Delta_i [\Delta_i - \pi(Y_i, \mathbf{S}'_i)] i_i^*(\theta_0) i_i^*(\theta_0)^T}{\pi^2(Y_i, \mathbf{S}'_i)} \right] + o^*(1), \\ &= -\mathbb{E} \left[\frac{[1 - \pi(Y_i, \mathbf{S}'_i)] i_i^*(\theta_0) i_i^*(\theta_0)^T}{\pi(Y_i, \mathbf{S}'_i)} \right] + o^*(1), \end{aligned}$$

where the notation $o^*(a_n)$ denotes a matrix whose components are uniformly $o^*(a_n)$. Finally,

$$\begin{aligned} \text{Var} [U_{ws,n}(\theta_0, \hat{\pi})] &= \mathbb{E} \left[\frac{i_i(\theta_0, \pi) i_i(\theta_0, \pi)^T}{\pi(Y_i, \mathbf{S}'_i)} \right] + \mathbb{E} \left[\frac{[1 - \pi(Y_i, \mathbf{S}'_i)] [i_i^*(\theta_0) i_i^*(\theta_0)^T]}{\pi(Y_i, \mathbf{S}'_i)} \right] \\ &\quad - 2\mathbb{E} \left[\frac{[1 - \pi(Y_i, \mathbf{S}'_i)] i_i^*(\theta_0) i_i^*(\theta_0)^T}{\pi(Y_i, \mathbf{S}'_i)} \right] + o^*(1), \\ &= \Omega_3(\theta_0, \pi) - [\Omega_4(\theta_0, \pi) - \Omega_5(\theta_0, \pi)] + o^*(1). \end{aligned}$$

Thus, by the central limit theorem, we have $U_{ws,n}(\theta_0, \hat{\pi})$ converges in distribution to the Gaussian vector of mean zero and variance $\Omega_3(\theta_0, \pi) - [\Omega_4(\theta_0, \pi) - \Omega_5(\theta_0, \pi)]$. Because $[-\bar{U}_{ws,n}^{-1}(\theta_0, \hat{\pi}) - \Sigma^{-1}(\theta_0)]$ converges in probability to 0, by Slutsky's theorem $[-\bar{U}_{ws,n}^{-1}(\theta_0, \hat{\pi}) - \Sigma^{-1}(\theta_0)] U_{ws,n}(\theta_0, \hat{\pi})$ converges in distribution to 0.

Finally, by Lemma 4 and Slutsky's theorem, $\sqrt{n}(\hat{\theta}_n^{ws} - \theta_0)$ converges in distribution to the Gaussian vector of mean zero and variance

$$\Delta_{ws} := \Sigma(\theta_0)^{-1} \{ \Omega_3(\theta_0, \pi) - [\Omega_4(\theta_0, \pi) - \Omega_5(\theta_0, \pi)] \} [\Sigma(\theta_0)^{-1}]^T.$$

ACKNOWLEDGMENTS

Authors are grateful to referees and editor for their comments and suggestions that led to significant improvements of earlier versions of this article.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. D. Lambert, "Zero-inflated Poisson regression with an application to defects in manufacturing," *Technometrics* **34**, 1–14 (1992).
2. E. Dietz and D. Böhning, "On estimation of the Poisson parameter in zero-modified Poisson models," *Comput. Stat. Data Anal.* **4**, 441–459 (2000).
3. K. K. W. Yau and A. H. Lee, "Zero-inflated Poisson regression with random effects to evaluate an occupational injury prevention programme," *Stat. Med.* **20**, 2907–2920 (2001).
4. A. O. Diallo, A. Diop, and J.-F. Dupuy, "Asymptotic properties of the maximum likelihood estimator in zero-inflated binomial regression," *Commun. Stat. Theory Methods* **46** (20), 9930–9948 (2017).
5. Y. B. Cheung, "Zero-inflated models for regression analysis of count data: a study of growth and development," *Stat. Med.* **21**, 1461–1469 (2002).
6. M. Reilly and M. S. Pepe, "A mean score method for missing and auxiliary covariates data in regression methods," *Biometrika* **82**, 299–314 (1995).
7. A. Diallo, A. Diop, and J.-F. Dupuy, "Estimation in zero-inflated binomial regression with missing covariates," *Statistics* **53** (4), 839–865 (2019).
8. T. M. Lukusa, S.-M. Lee, and C.-S. Li, "Semiparametric estimation of a zero-inflated Poisson regression model with missing covariates," *Metrika* **79** (4), 457–483 (2016).
9. T. M. Lukusa and F. K. Hing Phoa, "A note on the weighting-type estimations of the zero-inflated Poisson regression model with missing data in covariates," *Journal Pre-proof*. (2019).
10. D. G. Horvitz and D. J. Thompson, "A generalization of sampling without replacement from a finite universe," *Current Res. Biostat.* **47**, 663–685 (1952).
11. S. H. Hsieh, S. M. Lee, and P. S. Shen, "Logistic regression analysis of randomized response data with missing covariates," *J. Stat. Plan. Inference* **140**, 927–940 (2010).
12. D. B. Rubin, "Inference and missing data," *Biometrika* **63** (3), 581–592 (1976).
13. R. V. Foutz, "On the unique consistent solution to the likelihood equations," *J. Am. Stat. Assoc.* **72**, 147–148 (1977).
14. D. Böhning, E. Dietz, P. Schlattmann, L. Mendonca, and U. Kirchner, "The zero-inflated Poisson model and the decayed, missing, and filled teeth index in dental epidemiology," *J. R. Stat. Soc. Ser. A* **162**, 195–209 (1999).
15. S. H. Hsieh, S. M. Lee, and P. S. Shen, "Semiparametric analysis of randomized response data with missing covariates in logistic regression," *Comput. Stat. Data Anal.* **53**, 2673–2692 (2009).
16. D. Long, J. S. Preisser, A. H. Herringb, and C. E. Golin, "A marginalized zero-inflated Poisson regression model with overall exposure effects," *Statist. Med.* **33**, 5151–5165 (2014).
17. J. S. Preisser, J. W. Stamm, D. L. Long, and M. E. Kincade, "Review and recommendations for zero-inflated count regression modeling of dental caries indices in epidemiological studies," *Caries Research*. **46** (4), 413–423 (2012).
18. K. H. Benecha, J. S. Preisser, and K. Das, "Marginal Zero-inflated models with missing covariates," *Biometric Journal* (2018).
19. A. Henningsen and O. Toomet, "maxLik: A package for maximum likelihood estimation in R," *Computational Statistics* **26** (3), 443–458 (2011).
20. D. Wang and S. X. Chen, "Empirical likelihood for estimating equations with missing values," *Ann. Stat.* **37**, 490–517 (2000).
21. M. Reilly and M. S. Pepe, "A mean score method for missing and auxiliary covariates data in regression methods," *Biometrika* **82**, 299–314 (2019).
22. E. A. Nadaraya, "On estimating regression," *Theory of probability and its applications* **9**, 141–142 (1964).
23. S. Wang and C. Y. Wang, "A note on kernel assisted estimators in missing covariate regression," *Stat. Probabil. Lett.* **55**, 439–449 (2001).
24. G. S. Watson, "Smooth regression analysis," *Sankhya, Series A* **26**, 359–372 (1964).
25. Herbert A. Sturges, "The Choice of a Class Interval," *Journal of the American Statistical Association* **21**, 153 (1926).
26. G. F. Jenks, *Optimal Data Classification for Choropleth Maps* (Lawrence Kansas, 1977).
27. K. J. G. Kouakou, O. Hili, and J. F. Dupuy, "Estimation in the Zero-Inflated Bivariate Poisson model, with an application to health-care utilization data," *Africa Statistica* **16** (2): 2767–2788 (2021).
28. E. Ali, M. L. Diop, and A. Diop, "Statistical inference in a Zero-Inflated Bell regression model," *Mathematical methods of Statistic* **31**, 91–104 (2022).
29. Suruchi Mishra, Cynthia L. Ogden, and Melissa Dimeler, Dietary Supplement Use in the United States: National Health and Nutrition Examination Survey, National Health Statistics Reports, 2017–March 2020 (2023).

Publisher's Note. Pleiades Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.