



HAL
open science

Estimation in zero-inflated binomial regression with missing covariates

Alpha Oumar Diallo, Aliou Diop, Jean-François Dupuy

► **To cite this version:**

Alpha Oumar Diallo, Aliou Diop, Jean-François Dupuy. Estimation in zero-inflated binomial regression with missing covariates. *Statistics*, 2019, 53 (4), pp.839-865. 10.1080/02331888.2019.1619741 . hal-01585220

HAL Id: hal-01585220

<https://hal.science/hal-01585220>

Submitted on 11 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimation in zero-inflated binomial regression with missing covariates

Alpha Oumar DIALLO^{a,b}, Aliou DIOP^a, Jean-François DUPUY^b

^a*LERSTAD, CEA-MITIC, Gaston Berger University, Saint Louis, Senegal.*

^b*IRMAR-INSA, Rennes, France.*

Abstract

The zero-inflated binomial (ZIB) regression model was recently proposed to account for excess zeros in binomial regression. Since then, the model has been applied in various domains, such as dental epidemiology and health economics. In practice, it often arises that some covariates involved in ZIB regression have missing values. Assuming that the missingness probability can be estimated parametrically, we propose an inverse-probability-weighted estimator of the parameters of a ZIB model with missing-at-random covariates. Consistency and asymptotic normality of the proposed estimator are established. A consistent estimator of the asymptotic variance-covariance matrix is also provided. The finite-sample behavior of the estimator is assessed via simulations.

Keywords: Asymptotics, count data, excess of zeros, inverse-probability-weighting

1. Introduction

Count data with excess zeros arise in many disciplines, such as agriculture, economics, epidemiology, industry, insurance, terrorism study, traffic safety research. . . Excess of zeros refers to the situation where the number of observed zeros is larger than predicted by standard models for count data. Zero-inflated regression models, which are obtained by mixing a degenerate distribution at zero with a standard count regression model (such as Poisson, negative binomial or binomial) have been developed to analyze such data. For example, the zero-inflated Poisson (ZIP) regression model was proposed by Lambert (1992) and further developed by Dietz and Böhning (2000), Lim *et al.* (2014) and Monod (2014), among many others. Recent variants of ZIP regression include random-effects ZIP models (Hall, 2000; Min and Agresti, 2005), semi-varying coefficient ZIP models (Zhao *et al.*, 2015) and semi-parametric ZIP models (Lam *et al.*, 2006; Feng and Zhu, 2011). The zero-inflated negative binomial (ZINB) regression model was proposed by Ridout *et al.* (2001), see also Moghimbegi *et al.* (2008), Mwalili *et al.* (2008), Garay *et al.* (2011). When counts have an upper bound, ZIP and ZINB regression models are no longer appropriate. Hall (2000) and Vieira *et*

Email addresses: Alpha-Oumar.Diallo1@insa-rennes.fr (Alpha Oumar DIALLO), aliou.diop@ugb.edu.sn (Aliou DIOP), Jean-Francois.Dupuy@insa-rennes.fr (Jean-François DUPUY)

al. (2000) thus introduced the zero-inflated binomial (ZIB) regression model, see also Diop *et al.* (2016). ZIB regression was recently used in dental caries epidemiology (Gilthorpe *et al.*, 2009; Matranga *et al.*, 2013) and in health economics (Diallo *et al.*, 2017a). A zero-inflated model for multinomial counts (or ZIM model) was recently proposed by Diallo *et al.* (2017b).

In addition, missing data arise in a wide variety of disciplines. In the past decades, there has been an enormous literature on estimation in regression models with missing covariates, including missing covariates in linear models, generalized linear models, generalized linear mixed models, survival models. . . Despite this interest, only a few papers have focused on missing covariates in zero-inflated regression models. Chen and Fu (2011) develop a model selection criterion for zero-inflated regression models with missing covariates. Lukusa *et al.* (2016) consider estimation in ZIP regression with missing-at-random covariates. Motivated by this work, we investigate estimation in ZIB regression when some covariate values are missing for some of the sample individuals.

One simple approach to estimation with missing data is the so-called complete-case method, which consists in removing all incomplete cases from the statistical analysis. However, this method usually produces asymptotically biased estimators (unless data are missing completely at random, which is rarely the case in practice). Therefore, in this paper, we propose to rely on the alternative and more sophisticated inverse-probability-weighted approach, which has not been investigated yet in ZIB regression with missing covariates. Inverse-probability-weighting (IPW) is a general estimation method under missing data. It was originally proposed by Horvitz and Thompson (1952) and further developed by Zhao and Lipsitz (1992). The basic idea of IPW is to correct for missing data by giving extra weight to subjects with fully observed data. This idea has already proved useful in a variety of models, such as the logistic regression model (Hsieh *et al.*, 2010), proportional hazards regression model (Qi *et al.*, 2005) and single-index model (Li and Hu, 2016), for example.

The rest of the paper is organized as follows. In Section 2, we provide a brief review of ZIB regression, including model formulation and maximum likelihood estimation without missing data. Then, we introduce a IPW estimator of the parameters of a ZIB regression model with missing-at-random covariates. In Section 3, we establish consistency and asymptotic normality of the proposed estimator. A consistent estimator of its asymptotic variance-covariance matrix is provided. Section 4 reports results of a simulation study. A discussion and some perspectives are provided in Section 5.

2. ZIB regression with missing covariates

We first provide a brief review of ZIB regression and maximum likelihood estimation in ZIB model with complete data.

2.1. A brief review of ZIB regression

Let Z_i denote the random count of interest for individual i , $i = 1, \dots, n$. Individuals are assumed to be independent. The ZIB distribution is a mixture of a degenerate distribution

at zero and a binomial distribution. It is given as follows:

$$Z_i \sim \begin{cases} 0 & \text{with probability } p_i, \\ \mathcal{B}(m_i, \pi_i) & \text{with probability } 1 - p_i, \end{cases} \quad (2.1)$$

where p_i is a mixing probability for the accommodation of extra zeros and $\mathcal{B}(m, \pi)$ denotes the binomial distribution with size m and event probability π . The ZIB distribution reduces to a standard binomial distribution when $p_i = 0$.

In ZIB regression, the mixing probabilities p_i and event probabilities π_i are usually modeled via logistic regressions: $\text{logit}(p_i) = \gamma^\top \mathbf{W}_i$ and $\text{logit}(\pi_i) = \beta^\top \mathbf{X}_i$, where $\mathbf{X}_i = (1, X_{i2}, \dots, X_{ip})^\top$ and $\mathbf{W}_i = (1, W_{i2}, \dots, W_{iq})^\top$ are random vectors of predictors or covariates (both categorical and continuous covariates are allowed) and \top denotes the transpose operator. Vectors \mathbf{X}_i and \mathbf{W}_i may either have some common components or be distinct (note that some caution is required in the special case where $m_i = 1$ for all $i = 1, \dots, n$, see Diop *et al.*, 2011). Here, β and γ are respectively p and q -dimensional vectors of unknown regression parameters to be estimated.

Let $\{(Z_i, \mathbf{X}_i, \mathbf{W}_i), i = 1, \dots, n\}$ be a sample of independent observations and $\psi = (\beta^\top, \gamma^\top)^\top$ denote the whole unknown k -dimensional ($k := p + q$) parameter. The log-likelihood function $\ell\ell_n(\psi)$ based on the observed sample is:

$$\begin{aligned} \ell\ell_n(\psi) &= \sum_{i=1}^n \left\{ J_i \log \left(e^{\gamma^\top \mathbf{W}_i} + (1 + e^{\beta^\top \mathbf{X}_i})^{-m_i} \right) - \log \left(1 + e^{\gamma^\top \mathbf{W}_i} \right) \right. \\ &\quad \left. + (1 - J_i) \left[Z_i \beta^\top \mathbf{X}_i - m_i \log \left(1 + e^{\beta^\top \mathbf{X}_i} \right) \right] \right\}, \\ &:= \sum_{i=1}^n \ell_i(\psi), \end{aligned}$$

where $J_i := 1_{\{Z_i=0\}}$ (see Hall, 2000). The maximum likelihood estimator (MLE) $\hat{\psi}_n := (\hat{\beta}_n^\top, \hat{\gamma}_n^\top)^\top$ of ψ is obtained by solving the score equation $U_n(\psi) = 0$, where

$$U_n(\psi) = \frac{1}{\sqrt{n}} \frac{\partial \ell\ell_n(\psi)}{\partial \psi} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial \ell_i(\psi)}{\partial \psi} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{\ell}_i(\psi). \quad (2.2)$$

This estimating equation can be solved by using the expectation-maximization (EM) algorithm (Hall, 2000) or by a direct maximization of $\ell\ell_n(\psi)$ (Diallo *et al.*, 2017a). The MLE $\hat{\psi}_n$ is a consistent and asymptotically normal estimator of the true ψ , see Diallo *et al.* (2017a).

The next section describes the problem and the proposed estimator.

2.2. ZIB regression with missing covariates: the proposed estimator

In this work, we assume that some components of \mathbf{X}_i may be missing for some individuals. Decompose \mathbf{X}_i as $\mathbf{X}_i = (\mathbf{X}_i^{(obs), \top}, \mathbf{X}_i^{(miss), \top})^\top$, where $\mathbf{X}_i^{(obs)}$ and $\mathbf{X}_i^{(miss)}$ contain the observed and missing components of \mathbf{X}_i respectively (we assume that the same components of \mathbf{X}_i may be missing for all individuals). Let δ_i be a dummy variable indicating whether \mathbf{X}_i is fully

observed ($\delta_i = 1$) or not ($\delta_i = 0$). Finally, let $\mathbf{S}_i := (Z_i, \mathbf{X}_i^{(obs)\top}, \mathbf{W}_i^\top)^\top$ denote the vector of variables that are always observed on each individual. Then $\{Z_i, \mathbf{X}_i, \mathbf{W}_i\} = \{\mathbf{S}_i, \mathbf{X}_i^{(miss)}\}$. Let d denote the dimension of \mathbf{S}_i .

We assume that $\mathbf{X}_i^{(miss)}$ is missing at random (MAR, see Rubin, 1976): the probability that some components of \mathbf{X}_i are missing depends only on the observed variables. The MAR assumption can be expressed in terms of the missingness (or selection) probability $\mathbb{P}(\delta_i = 1 | \mathbf{S}_i, \mathbf{X}_i^{(miss)})$, as:

$$\mathbb{P}(\delta_i = 1 | \mathbf{S}_i, \mathbf{X}_i^{(miss)}) = \mathbb{P}(\delta_i = 1 | \mathbf{S}_i).$$

Under missing data, we propose to estimate ψ in ZIB model (2.1) by using the IPW method. Originally proposed by Horvitz and Thompson (1952), IPW has recently been used for estimating various regression models with missing or mismeasured covariates. Basic idea is to inversely weight the observed data by the selection probability $\mathbb{P}(\delta_i = 1 | \mathbf{S}_i)$, so as to reduce the bias due to incomplete cases deletion.

Recall that without missing data, ψ in model (2.1) can be estimated by solving the score equation (2.2). Under missing data, we propose to estimate ψ by solving the following estimating equation, derived from (2.2) by weighting individuals with fully observed data by the inverse of their selection probability:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{\mathbb{P}(\delta_i = 1 | \mathbf{S}_i)} \dot{\ell}_i(\psi) = 0. \quad (2.3)$$

In practice, selection probabilities $\mathbb{P}(\delta_i = 1 | \mathbf{S}_i)$ are usually unknown and need to be estimated. Several estimation procedures can be used. In this work, we consider the case where the $\mathbb{P}(\delta_i = 1 | \mathbf{S}_i)$, $i = 1, \dots, n$, can be estimated parametrically. In this case, logistic regression is the most frequently used option. Let $r_i(\alpha) := \mathbb{P}(\delta_i = 1 | \mathbf{S}_i)$ be defined as:

$$r_i(\alpha) = \frac{\exp(\alpha^\top \mathbf{S}_i)}{1 + \exp(\alpha^\top \mathbf{S}_i)}, \quad (2.4)$$

where α is a d -dimensional vector of unknown regression parameters. We need to estimate α before solving the weighted score equation (2.3). Maximum likelihood estimation can be used for that purpose. The MLE $\hat{\alpha}_n = \arg \max_{\alpha} \prod_{i=1}^n \{r_i(\alpha)^{\delta_i} (1 - r_i(\alpha))^{1 - \delta_i}\}$ in model (2.4) is known to be consistent and asymptotically Gaussian (Gouriéroux and Monfort, 1981). Once $\hat{\alpha}_n$ is available, one can estimate ψ by solving the estimated weighted score equation $U_{w,n}(\psi, \hat{\alpha}_n) = 0$, where

$$U_{w,n}(\psi, \hat{\alpha}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{r_i(\hat{\alpha}_n)} \dot{\ell}_i(\psi) = 0. \quad (2.5)$$

In what follows, the resulting estimator of ψ will be denoted by $\hat{\psi}_n$. Asymptotic properties of this estimator are established in Section 3. First, we need to introduce some further notations.

2.3. Some further notations

Define first the $(k \times d)$, $(k \times k)$ and $(d \times d)$ matrices

$$B(\psi, \alpha) = \lim_{n \rightarrow \infty} \left(-\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\delta_i \frac{1 - r_i(\alpha)}{r_i(\alpha)} \dot{\ell}_i(\psi) \mathbf{S}_i^\top \right] \right),$$

$$J(\psi, \alpha) = \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\delta_i \frac{\dot{\ell}_i(\psi) \dot{\ell}_i(\psi)^\top}{r_i^2(\alpha)} \right] \right),$$

and

$$\Sigma(\alpha) = \mathbb{E} [\mathbf{S}_i \mathbf{S}_i^\top r_i(\alpha) (1 - r_i(\alpha))]. \quad (2.6)$$

For every $i = 1, \dots, n$, let $U_i = (U_{i1}, \dots, U_{ik})^\top$ denote the k -dimensional column vector $U_i := B(\psi_0, \alpha_0) \Sigma(\alpha_0)^{-1} \mathbf{S}_i$. Then, define the $(p \times n)$, $(q \times n)$ and $(k \times n)$ matrices

$$\mathbb{X} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_{12} & X_{22} & \cdots & X_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1p} & X_{2p} & \cdots & X_{np} \end{pmatrix}, \quad \mathbb{W} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ W_{12} & W_{22} & \cdots & W_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ W_{1q} & W_{2q} & \cdots & W_{nq} \end{pmatrix},$$

and

$$\mathbb{U} = \begin{pmatrix} U_{11} & U_{21} & \cdots & U_{n1} \\ U_{12} & U_{22} & \cdots & U_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ U_{1k} & U_{2k} & \cdots & U_{nk} \end{pmatrix} = \begin{bmatrix} \mathbb{U}_1 \\ \mathbb{U}_2 \end{bmatrix},$$

where \mathbb{U}_1 is the $(p \times n)$ sub-matrix of \mathbb{U} consisting of the first p rows of \mathbb{U} and \mathbb{U}_2 is the $(q \times n)$ sub-matrix of \mathbb{U} consisting of the last q rows of \mathbb{U} . Let \mathbb{V} be the $(k \times 3n)$ block-matrix defined as

$$\mathbb{V} = \begin{bmatrix} \mathbb{X} & 0_{p,n} & \mathbb{U}_1 \\ 0_{q,n} & \mathbb{W} & \mathbb{U}_2 \end{bmatrix},$$

and $C(\psi, \alpha) = (C_j(\psi, \alpha))_{1 \leq j \leq 3n}$ be the $3n$ -dimensional column vector defined by

$$C(\psi, \alpha) = \left(\frac{\delta_1}{r_1(\alpha)} A_1(\psi), \dots, \frac{\delta_n}{r_n(\alpha)} A_n(\psi), \frac{\delta_1}{r_1(\alpha)} B_1(\psi), \dots, \frac{\delta_n}{r_n(\alpha)} B_n(\psi), \delta_1 - r_1(\alpha), \dots, \delta_n - r_n(\alpha) \right)^\top,$$

where $0_{a,b}$ denotes the $(a \times b)$ matrix whose components are all equal to zero and for every $i = 1, \dots, n$,

$$A_i(\psi) = -J_i \frac{m_i e^{\beta^\top \mathbf{X}_i}}{e^{\gamma^\top \mathbf{W}_i} (h_i(\beta))^{m_i+1} + h_i(\beta)} + (1 - J_i) \left(Z_i - \frac{m_i e^{\beta^\top \mathbf{X}_i}}{h_i(\beta)} \right) \quad (2.7)$$

and

$$B_i(\psi) = \frac{J_i e^{\gamma^\top \mathbf{W}_i} (h_i(\beta))^{m_i}}{e^{\gamma^\top \mathbf{W}_i} (h_i(\beta))^{m_i} + 1} - \frac{e^{\gamma^\top \mathbf{W}_i}}{1 + e^{\gamma^\top \mathbf{W}_i}}, \quad (2.8)$$

with $h_i(\beta) := 1 + e^{\beta^\top \mathbf{X}_i}$.

If $A = (A_{ij})_{1 \leq i \leq a, 1 \leq j \leq b}$ is a $(a \times b)$ matrix, $A_{\bullet j}$ will denote its j -th column ($j = 1, \dots, b$) that is, $A_{\bullet j} = (A_{1j}, \dots, A_{aj})^\top$.

Finally, under model (2.4), it is known that the MLE $\hat{\alpha}_n$ verifies

$$\sqrt{n}(\hat{\alpha}_n - \alpha_0) = \Sigma(\alpha_0)^{-1} M_n(\alpha_0) + o_{\mathbb{P}}(1), \quad (2.9)$$

where $\Sigma(\alpha_0)$ is given by (2.6) and $M_n(\alpha) = n^{-1/2} \sum_{i=1}^n \mathbf{S}_i(\delta_i - r_i(\alpha))$.

In the next section, we establish rigorously the consistency and asymptotic normality of the proposed IPW-MLE $\hat{\psi}_n$.

3. Asymptotic results

We first state some regularity conditions that will be needed for proving our asymptotic results.

3.1. Regularity conditions and consistency

C1 Covariates are bounded, that is, there exists a finite positive constant c_1 such that $|X_{ij}| \leq c_1$ and $|W_{i\ell}| \leq c_1$ for every $i = 1, \dots, n$, $j = 2, \dots, p$ and $\ell = 2, \dots, q$. For every $i = 1, \dots, n$, $j = 2, \dots, p$ and $\ell = 2, \dots, q$, $\text{var}[X_{ij}] > 0$ and $\text{var}[W_{i\ell}] > 0$. For every $i = 1, \dots, n$, the X_{ij} ($j = 1, \dots, p$) are linearly independent and the $W_{i\ell}$ ($\ell = 1, \dots, q$) are linearly independent.

C2 The true parameter value $\psi_0 := (\beta_0^\top, \gamma_0^\top)^\top$ lies in the interior of some known compact set of $\mathbb{R}^p \times \mathbb{R}^q$. The true α_0 belongs to the interior of some known compact set of \mathbb{R}^d .

C3 As $n \rightarrow \infty$, $n^{-1} \sum_{i=1}^n \mathbb{E} \left[\frac{\delta_i}{r_i(\alpha_0)} \frac{\partial^2 \ell_i(\psi)}{\partial \psi \partial \psi^\top} \right]$ converges to some invertible matrix $A(\psi, \alpha_0)$ and the smallest eigenvalue λ_n of $\mathbb{V}\mathbb{V}^\top$ tends to $+\infty$.

C4 For every $i = 1, \dots, n$, we have $m_i \in \{2, \dots, M\}$ for some finite integer value M .

In what follows, the space \mathbb{R}^k of k -dimensional (column) vectors will be provided with the Euclidean norm $\|\cdot\|_2$ and the space of $(k \times k)$ real matrices will be provided with the norm $\|A\|_2 := \max_{\|x\|_2=1} \|Ax\|_2$ (for notations simplicity, we will use $\|\cdot\|$ for both norms).

We first prove consistency of $\hat{\psi}_n$:

Theorem 3.1. *Assume that conditions C1-C4 hold. Then, as $n \rightarrow \infty$, $\hat{\psi}_n$ converges in probability to ψ_0 .*

Proof of Theorem 3.1. To prove consistency of $\hat{\psi}_n$, we verify the conditions of the inverse function theorem of Foutz (1977). These conditions are proved in a series of technical lemmas.

Lemma 3.2. $\partial U_{w,n}(\psi, \hat{\alpha}_n)/\partial \psi^\top$ exists and is continuous in an open neighborhood of ψ_0 .

Proof of Lemma 3.2. The $\ell_i(\psi), i = 1, \dots, n$ are twice differentiable with respect to ψ . Continuity of $\frac{\partial^2 \ell_i(\psi)}{\partial \psi \partial \psi^\top}$ is straightforward and is omitted. \square

Lemma 3.3. As $n \rightarrow \infty$, $n^{-1/2}U_{w,n}(\psi_0, \hat{\alpha}_n)$ converges in probability to 0.

Proof of Lemma 3.3. Decompose $n^{-1/2}U_{w,n}(\psi_0, \hat{\alpha}_n)$ as:

$$n^{-1/2}U_{w,n}(\psi_0, \hat{\alpha}_n) = \left(n^{-1/2}U_{w,n}(\psi_0, \hat{\alpha}_n) - n^{-1/2}U_{w,n}(\psi_0, \alpha_0) \right) + n^{-1/2}U_{w,n}(\psi_0, \alpha_0), \quad (3.10)$$

and consider the first term on the right-hand side of this decomposition. We have:

$$\begin{aligned} n^{-1/2}U_{w,n}(\psi_0, \hat{\alpha}_n) - n^{-1/2}U_{w,n}(\psi_0, \alpha_0) &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{r_i(\hat{\alpha}_n)} - \frac{1}{r_i(\alpha_0)} \right) \dot{\ell}_i(\psi_0), \\ &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{e^{\hat{\alpha}_n^\top \mathbf{S}_i}} - \frac{1}{e^{\alpha_0^\top \mathbf{S}_i}} \right) \dot{\ell}_i(\psi_0). \end{aligned}$$

By a Taylor expansion of $1/e^{\hat{\alpha}_n^\top \mathbf{S}_i}$ around α_0 ,

$$n^{-1/2}U_{w,n}(\psi_0, \hat{\alpha}_n) - n^{-1/2}U_{w,n}(\psi_0, \alpha_0) = \frac{1}{n} \sum_{i=1}^n \delta_i (\alpha_0 - \hat{\alpha}_n)^\top \mathbf{S}_i \frac{1}{e^{\alpha_*^\top \mathbf{S}_i}} \dot{\ell}_i(\psi_0),$$

where α_* is on the line segment between $\hat{\alpha}_n$ and α_0 . Then, we have:

$$\begin{aligned} \|n^{-1/2}U_{w,n}(\psi_0, \hat{\alpha}_n) - n^{-1/2}U_{w,n}(\psi_0, \alpha_0)\| &\leq \frac{1}{n} \sum_{i=1}^n \left\| \delta_i (\alpha_0 - \hat{\alpha}_n)^\top \mathbf{S}_i \frac{1}{e^{\alpha_*^\top \mathbf{S}_i}} \dot{\ell}_i(\psi_0) \right\|, \\ &\leq \frac{1}{n} \sum_{i=1}^n \left| \delta_i (\alpha_0 - \hat{\alpha}_n)^\top \mathbf{S}_i \frac{1}{e^{\alpha_*^\top \mathbf{S}_i}} \right| \left\| \dot{\ell}_i(\psi_0) \right\|, \\ &\leq \frac{1}{n} \sum_{i=1}^n \|\alpha_0 - \hat{\alpha}_n\| \|\mathbf{S}_i\| \frac{1}{e^{\alpha_*^\top \mathbf{S}_i}} \left\| \dot{\ell}_i(\psi_0) \right\|, \end{aligned}$$

where the second to third line comes from Cauchy-Schwarz inequality. Now, straightforward calculations show that

$$\dot{\ell}_i(\psi_0) = (\mathbf{X}_i^\top, 0_q^\top)^\top \cdot A_i(\psi_0) + (0_p^\top, \mathbf{W}_i^\top)^\top \cdot B_i(\psi_0),$$

where $0_p := 0_{p,1}$ is the p -dimensional column vector having all its components equal to 0 and $A_i(\psi_0), B_i(\psi_0)$ are given by (2.7) and (2.8) respectively. Thus we have:

$$\|\dot{\ell}_i(\psi_0)\| \leq \|(\mathbf{X}_i^\top, 0_q^\top)^\top\| \cdot |A_i(\psi_0)| + \|(0_p^\top, \mathbf{W}_i^\top)^\top\| \cdot |B_i(\psi_0)|.$$

Under conditions C1, C2 and C4, it is easy to see that $|A_i(\psi_0)|$ and $|B_i(\psi_0)|$ are bounded above. Thus, there exists a finite constant c_2 such that $n^{-1} \sum_{i=1}^n \|\dot{\ell}_i(\psi_0)\| \leq c_2$. Note that $\|\mathbf{S}_i\|$ and $1/e^{\alpha_*^\top \mathbf{S}_i}$ are also bounded, by conditions C1 and C2. Therefore, there exists some finite constant c_3 such that $\|n^{-1/2}U_{w,n}(\psi_0, \hat{\alpha}_n) - n^{-1/2}U_{w,n}(\psi_0, \alpha_0)\| \leq c_3\|\alpha_0 - \hat{\alpha}_n\|$. Finally, the convergence of $\hat{\alpha}_n$ to α_0 imply that $\|n^{-1/2}U_{w,n}(\psi_0, \hat{\alpha}_n) - n^{-1/2}U_{w,n}(\psi_0, \alpha_0)\|$ converges in probability to 0 as $n \rightarrow \infty$.

Next, consider the term $n^{-1/2}U_{w,n}(\psi_0, \alpha_0)$ in decomposition (3.10). Some simple algebra yields:

$$n^{-1/2}U_{w,n}(\psi_0, \alpha_0) = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{r_i(\alpha_0)} X_{i1} A_i(\psi_0) \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{r_i(\alpha_0)} X_{ip} A_i(\psi_0) \\ \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{r_i(\alpha_0)} W_{i1} B_i(\psi_0) \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{r_i(\alpha_0)} W_{iq} B_i(\psi_0) \end{pmatrix}.$$

We prove that $n^{-1/2}U_{w,n}(\psi_0, \alpha_0)$ converges in probability to 0 as $n \rightarrow \infty$. To see this, note first that for every $i = 1, \dots, n$ and $\ell = 1, \dots, q$:

$$\begin{aligned} \mathbb{E} \left[\frac{\delta_i}{r_i(\alpha_0)} W_{i\ell} B_i(\psi_0) \right] &= \mathbb{E} \left[\mathbb{E} \left[\frac{\delta_i}{r_i(\alpha_0)} W_{i\ell} B_i(\psi_0) \middle| \mathbf{S}_i \right] \right], \\ &= \mathbb{E} \left[\frac{1}{r_i(\alpha_0)} W_{i\ell} \mathbb{E} [\delta_i B_i(\psi_0) | \mathbf{S}_i] \right]. \end{aligned}$$

Given \mathbf{S}_i , $B_i(\psi_0)$ is a function of $\mathbf{X}_i^{(miss)}$ only. Thus, by the MAR assumption, $B_i(\psi_0)$ and δ_i are independent given \mathbf{S}_i . It follows that:

$$\begin{aligned} \mathbb{E} \left[\frac{\delta_i}{r_i(\alpha_0)} W_{i\ell} B_i(\psi_0) \right] &= \mathbb{E} \left[\frac{1}{r_i(\alpha_0)} W_{i\ell} \mathbb{E} [\delta_i | \mathbf{S}_i] \mathbb{E} [B_i(\psi_0) | \mathbf{S}_i] \right], \\ &= \mathbb{E} [W_{i\ell} \mathbb{E} [B_i(\psi_0) | \mathbf{S}_i]], \\ &= \mathbb{E} [W_{i\ell} B_i(\psi_0)]. \end{aligned}$$

Diallo *et al.* (2017a) proved that $\mathbb{E} [W_{i\ell} B_i(\psi_0)] = 0$ for every $i = 1, \dots, n$ and $\ell = 1, \dots, q$. Thus, $\mathbb{E} \left[\frac{\delta_i}{r_i(\alpha_0)} W_{i\ell} B_i(\psi_0) \right] = 0$ for every $i = 1, \dots, n$ and $\ell = 1, \dots, q$. Similarly, for every $i = 1, \dots, n$ and $j = 1, \dots, p$, we have:

$$\mathbb{E} \left[\frac{\delta_i}{r_i(\alpha_0)} X_{ij} A_i(\psi_0) \right] = \mathbb{E} \left[\mathbb{E} \left[\frac{\delta_i}{r_i(\alpha_0)} X_{ij} A_i(\psi_0) \middle| \mathbf{S}_i \right] \right].$$

Two cases should be considered, namely: *i*) X_{ij} is a component of $\mathbf{X}_i^{(miss)}$ and *ii*) X_{ij} is a component of $\mathbf{X}_i^{(obs)}$. In case *i*), we have:

$$\mathbb{E} \left[\mathbb{E} \left[\frac{\delta_i}{r_i(\alpha_0)} X_{ij} A_i(\psi_0) \middle| \mathbf{S}_i \right] \right] = \mathbb{E} \left[\frac{1}{r_i(\alpha_0)} \mathbb{E} [\delta_i X_{ij} A_i(\psi_0) | \mathbf{S}_i] \right].$$

Given \mathbf{S}_i , $X_{ij}A_i(\psi_0)$ is a function of $\mathbf{X}_i^{(miss)}$ only. Thus, by the MAR assumption,

$$\begin{aligned}\mathbb{E}\left[\frac{1}{r_i(\alpha_0)}\mathbb{E}[\delta_i X_{ij}A_i(\psi_0)|\mathbf{S}_i]\right] &= \mathbb{E}\left[\frac{1}{r_i(\alpha_0)}\mathbb{E}[\delta_i|\mathbf{S}_i]\mathbb{E}[X_{ij}A_i(\psi_0)|\mathbf{S}_i]\right], \\ &= \mathbb{E}[X_{ij}A_i(\psi_0)].\end{aligned}$$

Diallo *et al.* (2017a) proved that $\mathbb{E}[X_{ij}A_i(\psi_0)] = 0$ for every $i = 1, \dots, n$ and $j = 1, \dots, p$. Therefore, $\mathbb{E}\left[\frac{\delta_i}{r_i(\alpha_0)}X_{ij}A_i(\psi_0)\right] = 0$. In case *ii*),

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}\left[\frac{\delta_i}{r_i(\alpha_0)}X_{ij}A_i(\psi_0)\middle|\mathbf{S}_i\right]\right] &= \mathbb{E}\left[\frac{1}{r_i(\alpha_0)}X_{ij}\mathbb{E}[\delta_i A_i(\psi_0)|\mathbf{S}_i]\right], \\ &= \mathbb{E}\left[\frac{1}{r_i(\alpha_0)}X_{ij}\mathbb{E}[\delta_i|\mathbf{S}_i]\mathbb{E}[A_i(\psi_0)|\mathbf{S}_i]\right], \\ &= \mathbb{E}[X_{ij}A_i(\psi_0)], \\ &= 0,\end{aligned}$$

where the first to second line comes from the fact that under MAR, $A_i(\psi_0)$ and δ_i are independent given \mathbf{S}_i . Finally, in case *ii*), we also have: $\mathbb{E}\left[\frac{\delta_i}{r_i(\alpha_0)}X_{ij}A_i(\psi_0)\right] = 0$.

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

$$\text{var}\left(\frac{\delta_i}{r_i(\alpha_0)}W_{i\ell}B_i(\psi_0)\right) \leq \mathbb{E}\left[\frac{\delta_i}{r_i^2(\alpha_0)}W_{i\ell}^2B_i^2(\psi_0)\right].$$

By C1, C2, C4, there exists finite constants c_4 and c_5 such that $1/r_i^2(\alpha_0) \leq c_4$ and $B_i^2(\psi_0) \leq c_5$ for every $i = 1, \dots, n$. Therefore,

$$\text{var}\left(\frac{\delta_i}{r_i(\alpha_0)}W_{i\ell}B_i(\psi_0)\right) \leq c_6 := c_1^2 c_4 c_5.$$

It follows that

$$\sum_{i=1}^{\infty} \frac{\text{var}\left(\frac{\delta_i}{r_i(\alpha_0)}W_{i\ell}B_i(\psi_0)\right)}{i^2} \leq c_6 \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty.$$

One can easily show that a similar result holds for $\text{var}\left(\frac{\delta_i}{r_i(\alpha_0)}X_{ij}A_i(\psi_0)\right)$. By Kolmogorov's law of large numbers (see for example Jiang (2010), Theorem 6.7),

$$\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\delta_i}{r_i(\alpha_0)}W_{i\ell}B_i(\psi_0) - \mathbb{E}\left[\frac{\delta_i}{r_i(\alpha_0)}W_{i\ell}B_i(\psi_0)\right] \right\} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{r_i(\alpha_0)}W_{i\ell}B_i(\psi_0), \quad \ell = 1, \dots, q$$

and

$$\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\delta_i}{r_i(\alpha_0)}X_{ij}A_i(\psi_0) - \mathbb{E}\left[\frac{\delta_i}{r_i(\alpha_0)}X_{ij}A_i(\psi_0)\right] \right\} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{r_i(\alpha_0)}X_{ij}A_i(\psi_0), \quad j = 1, \dots, p$$

converge in probability to 0 as $n \rightarrow \infty$ and thus, $n^{-1/2}U_{w,n}(\psi_0, \alpha_0)$ converges in probability to 0. This implies that $n^{-1/2}U_{w,n}(\psi_0, \hat{\alpha}_n)$ converges to 0, which concludes the proof. \square

Lemma 3.4. *As $n \rightarrow \infty$, $n^{-1/2} \partial U_{w,n}(\psi, \hat{\alpha}_n) / \partial \psi^\top$ converges in probability to a fixed function $A(\psi, \alpha_0)$, uniformly in an open neighborhood of ψ_0 .*

Proof of Lemma 3.4. Let $\tilde{U}_{w,n}(\psi, \alpha) := n^{-1/2} \partial U_{w,n}(\psi, \alpha) / \partial \psi^\top$ and \mathcal{V}_{ψ_0} be an open neighborhood of ψ_0 . Let $\psi \in \mathcal{V}_{\psi_0}$. Using similar arguments as in proof of Lemma 3.3, we have:

$$\|\tilde{U}_{w,n}(\psi, \hat{\alpha}_n) - \tilde{U}_{w,n}(\psi, \alpha_0)\| \leq \frac{1}{n} \sum_{i=1}^n \|\alpha_0 - \hat{\alpha}_n\| \|\mathbf{S}_i\| \frac{1}{e^{\alpha_*^\top \mathbf{S}_i}} \|\ddot{\ell}_i(\psi)\|,$$

where α_* is on the line segment between $\hat{\alpha}_n$ and α_0 . From this, one easily proves that $\|\tilde{U}_{w,n}(\psi, \hat{\alpha}_n) - \tilde{U}_{w,n}(\psi, \alpha_0)\|$ converges in probability to 0 as $n \rightarrow \infty$. Details are omitted. Now, consider the (ℓ, j) -th element of $\tilde{U}_{w,n}(\psi, \alpha_0)$, namely:

$$\left(\tilde{U}_{w,n}(\psi, \alpha_0)\right)_{(\ell, j)} = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\delta_i}{r_i(\alpha_0)} \frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j} \right\}.$$

We have:

$$\left(\tilde{U}_{w,n}(\psi, \alpha_0)\right)_{(\ell, j)} = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\delta_i}{r_i(\alpha_0)} \frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j} - \mathbb{E} \left[\frac{\delta_i}{r_i(\alpha_0)} \frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j} \right] \right\} + \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\frac{\delta_i}{r_i(\alpha_0)} \frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j} \right].$$

Now,

$$\begin{aligned} \text{var} \left(\frac{\delta_i}{r_i(\alpha_0)} \frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j} \right) &\leq \mathbb{E} \left(\frac{\delta_i}{r_i^2(\alpha_0)} \left\{ \frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j} \right\}^2 \right), \\ &\leq c_4 \mathbb{E} \left(\left\{ \frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j} \right\}^2 \right). \end{aligned}$$

We prove that $\text{var} \left(\frac{\delta_i}{r_i(\alpha_0)} \frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j} \right)$ is bounded. Some tedious albeit easy calculations show that $\frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j}$ is the (ℓ, j) -th element of the $(k \times k)$ matrix $(-\mathbf{V}_i \mathbf{U}_i(\psi) \mathbf{V}_i^\top)$, where \mathbf{V}_i is the $(k \times 2)$ matrix defined as

$$\mathbf{V}_i = \begin{pmatrix} \mathbf{X}_i & 0_p \\ 0_q & \mathbf{W}_i \end{pmatrix}$$

and

$$\mathbf{U}_i(\psi) = \begin{pmatrix} \mathbf{U}_{i,1}(\psi) & \mathbf{U}_{i,2}(\psi) \\ \mathbf{U}_{i,2}(\psi) & \mathbf{U}_{i,3}(\psi) \end{pmatrix}$$

is the (2×2) symmetric matrix defined by

$$\begin{aligned} \mathbf{U}_{i,1}(\psi) &= \frac{J_i m_i e^{\beta^\top \mathbf{X}_i}}{(k_i(\psi))^2} \left(k_i(\psi) - e^{\beta^\top \mathbf{X}_i} \left[e^{\gamma^\top \mathbf{W}_i} (m_i + 1) (h_i(\beta))^{m_i} + 1 \right] \right) + \frac{m_i (1 - J_i) e^{\beta^\top \mathbf{X}_i}}{(h_i(\beta))^2}, \\ \mathbf{U}_{i,2}(\psi) &= - \frac{J_i m_i e^{\beta^\top \mathbf{X}_i + \gamma^\top \mathbf{W}_i} (h_i(\beta))^{m_i + 1}}{(k_i(\psi))^2}, \\ \mathbf{U}_{i,3}(\psi) &= \frac{J_i e^{\gamma^\top \mathbf{W}_i} (h_i(\beta))^{m_i + 1}}{(k_i(\psi))^2} \left(e^{\gamma^\top \mathbf{W}_i} (h_i(\beta))^{m_i + 1} - k_i(\psi) \right) + \frac{e^{\gamma^\top \mathbf{W}_i}}{(1 + e^{\gamma^\top \mathbf{W}_i})^2}, \end{aligned}$$

with $k_i(\psi) := e^{\gamma^\top \mathbf{W}_i} (h_i(\beta))^{m_i+1} + h_i(\beta)$, $i = 1, \dots, n$. Using these notations, it is easy to see that

$$\begin{aligned} \frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j} &= - (\mathbf{V}_{i,(\ell,1)} \mathbf{U}_{i,1}(\psi) + \mathbf{V}_{i,(\ell,2)} \mathbf{U}_{i,2}(\psi)) \mathbf{V}_{i,(j,1)} \\ &\quad - (\mathbf{V}_{i,(\ell,1)} \mathbf{U}_{i,2}(\psi) + \mathbf{V}_{i,(\ell,2)} \mathbf{U}_{i,3}(\psi)) \mathbf{V}_{i,(j,2)}, \end{aligned} \quad (3.11)$$

where $\mathbf{V}_{i,(a,b)}$ denotes the (a,b) -th element of matrix \mathbf{V}_i . For a given row ℓ ($\ell = 1, \dots, k$), exactly one of $\mathbf{V}_{i,(\ell,1)}$ and $\mathbf{V}_{i,(\ell,2)}$ must be equal to 0 (this is straightforward from the expression of \mathbf{V}_i). Suppose for example that $\mathbf{V}_{i,(\ell,1)} = 0$ and $\mathbf{V}_{i,(j,2)} = 0$ (other combinations of null and non-null values among $(\mathbf{V}_{i,(\ell,1)}, \mathbf{V}_{i,(\ell,2)})$ and $(\mathbf{V}_{i,(j,1)}, \mathbf{V}_{i,(j,2)})$ can be treated similarly). Then (3.11) reduces to:

$$\frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j} = -\mathbf{V}_{i,(\ell,2)} \mathbf{U}_{i,2}(\psi) \mathbf{V}_{i,(j,1)}.$$

Let $M_{\mathbf{X}} := \max_{\beta, \mathbf{X}} e^{\beta^\top \mathbf{X}}$ and $M_{\mathbf{W}} := \max_{\gamma, \mathbf{W}} e^{\gamma^\top \mathbf{W}}$. Under conditions C1, C2 and C4, we have:

$$|\mathbf{U}_{i,2}(\psi)| \leq M^* := M \cdot M_{\mathbf{X}} \cdot M_{\mathbf{W}} \cdot (1 + M_{\mathbf{X}})^{M+1} < \infty,$$

which implies

$$\mathbb{E} \left(\left\{ \frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j} \right\}^2 \right) \leq c_1^4 M^{*2},$$

and finally,

$$\text{var} \left(\frac{\delta_i}{r_i(\alpha_0)} \frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j} \right) \leq c_4 c_1^4 M^{*2} < \infty.$$

It follows that

$$\sum_{i=1}^{\infty} \frac{\text{var} \left(\frac{\delta_i}{r_i(\alpha_0)} \frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j} \right)}{i^2} \leq c_4 c_1^4 M^{*2} \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty.$$

Therefore, Kolmogorov's law of large numbers implies that

$$\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\delta_i}{r_i(\alpha_0)} \frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j} - \mathbb{E} \left[\frac{\delta_i}{r_i(\alpha_0)} \frac{\partial^2 \ell_i(\psi)}{\partial \psi_\ell \partial \psi_j} \right] \right\}$$

converges in probability to 0 as $n \rightarrow \infty$ and by condition C3, $(\tilde{U}_{w,n}(\psi, \alpha_0))_{(\ell,j)}$ converges in probability to the (ℓ, j) -th element of the matrix $A(\psi, \alpha_0)$. Finally, $\tilde{U}_{w,n}(\psi, \hat{\alpha}_n)$ converges in probability to $A(\psi, \alpha_0)$. Under conditions C1, C2 and C4, the derivative of $\tilde{U}_{w,n}(\psi, \hat{\alpha}_n)$ with respect to ψ is bounded, for every n . Therefore, the sequence $(\tilde{U}_{w,n}(\psi, \hat{\alpha}_n))_n$ is equicontinuous. It follows that the convergence of $\tilde{U}_{w,n}(\psi, \hat{\alpha}_n)$ to $A(\psi, \alpha_0)$ is uniform on \mathcal{V}_{ψ_0} . \square

Having now verified the conditions of Foutz (1977) inverse function theorem, we conclude that $\hat{\psi}_n$ converges in probability to ψ_0 . \square

3.2. Asymptotic normality

Our second main result asserts that the IPW-MLE $\hat{\psi}_n$ is asymptotically Gaussian.

Theorem 3.5. *Assume that conditions C1-C4 hold. Then $\sqrt{n}(\hat{\psi}_n - \psi_0)$ is asymptotically normally distributed with mean zero and covariance matrix Δ , where*

$$\Delta := A(\psi_0, \alpha_0)^{-1} \{J(\psi_0, \alpha_0) - B(\psi_0, \alpha_0)\Sigma(\alpha_0)^{-1}B(\psi_0, \alpha_0)^\top\} [A(\psi_0, \alpha_0)^{-1}]^\top.$$

Proof of Theorem 3.5. A Taylor series expansion of $U_{w,n}(\hat{\psi}_n, \hat{\alpha}_n)$ at (ψ_0, α_0) yields

$$0 = U_{w,n}(\hat{\psi}_n, \hat{\alpha}_n) = U_{w,n}(\psi_0, \alpha_0) + \frac{\partial U_{w,n}(\psi_0, \alpha_0)}{\partial \psi^\top} (\hat{\psi}_n - \psi_0) + \frac{\partial U_{w,n}(\psi_0, \alpha_0)}{\partial \alpha^\top} (\hat{\alpha}_n - \alpha_0) + o_{\mathbb{P}}(1).$$

Let $\check{U}_{w,n}(\psi, \alpha) := n^{-1/2} \partial U_{w,n}(\psi, \alpha) / \partial \alpha^\top$. Then we have:

$$0 = U_{w,n}(\psi_0, \alpha_0) + \check{U}_{w,n}(\psi_0, \alpha_0) \sqrt{n}(\hat{\psi}_n - \psi_0) + \check{U}_{w,n}(\psi_0, \alpha_0) \sqrt{n}(\hat{\alpha}_n - \alpha_0) + o_{\mathbb{P}}(1). \quad (3.12)$$

Now, straightforward calculations yield

$$\check{U}_{w,n}(\psi, \alpha) = -\frac{1}{n} \sum_{i=1}^n \delta_i \frac{1 - r_i(\alpha)}{r_i(\alpha)} \dot{\ell}_i(\psi) \mathbf{S}_i^\top,$$

and it can be proved that $\check{U}_{w,n}(\psi_0, \alpha_0)$ converges in probability to $B(\psi_0, \alpha_0)$ (arguments are similar to those in proof of Lemma 3.4 and are thus omitted). Combining this with (2.9), we can re-express (3.12) as:

$$0 = U_{w,n}(\psi_0, \alpha_0) + \check{U}_{w,n}(\psi_0, \alpha_0) \sqrt{n}(\hat{\psi}_n - \psi_0) + B(\psi_0, \alpha_0) \Sigma(\alpha_0)^{-1} M_n(\alpha_0) + o_{\mathbb{P}}(1),$$

and it follows that:

$$\sqrt{n}(\hat{\psi}_n - \psi_0) = -\check{U}_{w,n}(\psi_0, \alpha_0)^{-1} (U_{w,n}(\psi_0, \alpha_0) + B(\psi_0, \alpha_0) \Sigma(\alpha_0)^{-1} M_n(\alpha_0)) + o_{\mathbb{P}}(1).$$

Using notations introduced in Section 2.3, we finally obtain:

$$\begin{aligned} \sqrt{n}(\hat{\psi}_n - \psi_0) &= -\check{U}_{w,n}(\psi_0, \alpha_0)^{-1} n^{-1/2} \mathbb{V}C(\psi_0, \alpha_0) + o_{\mathbb{P}}(1), \\ &= -\check{U}_{w,n}(\psi_0, \alpha_0)^{-1} \sum_{j=1}^{3n} \mathbb{V}_{\bullet j} C_{j,n}(\psi_0, \alpha_0) + o_{\mathbb{P}}(1), \end{aligned}$$

where $C_{j,n}(\psi_0, \alpha_0) = n^{-1/2} C_j(\psi_0, \alpha_0)$. Let $\mathbb{C}_n^2 = \text{var}(U_{w,n}(\psi_0, \alpha_0) + B(\psi_0, \alpha_0) \Sigma(\alpha_0)^{-1} M_n(\alpha_0))$. Then, by Eicker (1966), the random linear form $\mathbb{C}_n^{-1} \sum_{j=1}^{3n} \mathbb{V}_{\bullet j} C_{j,n}(\psi_0, \alpha_0)$ converges in distribution to the k -dimensional standard Gaussian distribution if the following conditions are satisfied: a) $\max_{1 \leq j \leq 3n} \mathbb{V}_{\bullet j}^\top (\mathbb{V}\mathbb{V}^\top)^{-1} \mathbb{V}_{\bullet j} \rightarrow 0$ as $n \rightarrow \infty$, b) $\sup_{1 \leq j \leq 3n} \mathbb{E}[C_{j,n}^2(\psi_0, \alpha_0) \mathbf{1}_{\{|C_{j,n}(\psi_0, \alpha_0)| > c\}}] \rightarrow 0$ as $c \rightarrow \infty$, c) $\inf_{1 \leq j \leq 3n} \mathbb{E}[C_{j,n}^2(\psi_0, \alpha_0)] > 0$. Note first that

$$0 < \max_{1 \leq j \leq 3n} \mathbb{V}_{\bullet j}^\top (\mathbb{V}\mathbb{V}^\top)^{-1} \mathbb{V}_{\bullet j} \leq \max_{1 \leq j \leq 3n} \|\mathbb{V}_{\bullet j}\|^2 \|(\mathbb{V}\mathbb{V}^\top)^{-1}\| = \max_{1 \leq j \leq 3n} \|\mathbb{V}_{\bullet j}\|^2 / \lambda_n.$$

Since $\|\mathbb{V}_{\bullet j}\|$ is bounded, condition C3 implies that condition a) is satisfied. Condition b) follows by noting that $C_{j,n}(\psi_0, \alpha_0)$ (for $j = 1, \dots, 3n$) are bounded under C1, C2, C4. Finally, under C1, C2 and C4, we have $\mathbb{E}[C_{j,n}^2(\psi_0, \alpha_0)] > 0$ for every $j = 1, \dots, 3n$.

Moreover,

$$\begin{aligned} \mathbb{C}_n^2 &= \text{var}(U_{w,n}(\psi_0, \alpha_0)) + \text{var}(B(\psi_0, \alpha_0)\Sigma(\alpha_0)^{-1}M_n(\alpha_0)) \\ &\quad + 2\text{cov}(U_{w,n}(\psi_0, \alpha_0), B(\psi_0, \alpha_0)\Sigma(\alpha_0)^{-1}M_n(\alpha_0)). \end{aligned}$$

Straightforward calculations yield: $\text{var}(U_{w,n}(\psi_0, \alpha_0)) = \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\delta_i \frac{\dot{\ell}_i(\psi_0)\dot{\ell}_i(\psi_0)^\top}{r_i^2(\alpha_0)} \right]$, $\text{var}(M_n(\alpha_0)) = \Sigma(\alpha_0)$ and

$$\text{cov}(U_{w,n}(\psi_0, \alpha_0), B(\psi_0, \alpha_0)\Sigma(\alpha_0)^{-1}M_n(\alpha_0)) = \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[\dot{\ell}_i(\psi_0)\delta_i \frac{(1-r_i(\alpha_0))}{r_i(\alpha_0)} \mathbf{S}_i^\top \right] \Sigma(\alpha_0)^{-1}B(\psi_0, \alpha_0)^\top.$$

Hence, \mathbb{C}_n^2 converges to $J(\psi_0, \alpha_0) - B(\psi_0, \alpha_0)\Sigma(\alpha_0)^{-1}B(\psi_0, \alpha_0)^\top$. It follows that $\sum_{j=1}^{3n} \mathbb{V}_{\bullet j} C_{j,n}(\psi_0, \alpha_0)$ converges in distribution to a k -dimensional Gaussian vector with mean zero and variance $J(\psi_0, \alpha_0) - B(\psi_0, \alpha_0)\Sigma(\alpha_0)^{-1}B(\psi_0, \alpha_0)^\top$. Finally, using Lemma 3.4, condition C3 and Slutsky's theorem, $\sqrt{n}(\hat{\psi}_n - \psi_0)$ converges in distribution to a mean-zero Gaussian vector with variance Δ , where Δ is defined in Theorem 3.5. \square

Remark. A consistent estimator of Δ is given by

$$\hat{\Delta}_n := A_n(\hat{\psi}_n, \hat{\alpha}_n)^{-1} \{J_n(\hat{\psi}_n, \hat{\alpha}_n) - B_n(\hat{\psi}_n, \hat{\alpha}_n)\Sigma_n(\hat{\alpha}_n)^{-1}B_n(\hat{\psi}_n, \hat{\alpha}_n)^\top\} \left[A_n(\hat{\psi}_n, \hat{\alpha}_n)^{-1} \right]^\top \quad (3.13)$$

where

$$\begin{aligned} A_n(\psi, \alpha) &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{r_i(\alpha)} \frac{\partial^2 \ell_i(\psi)}{\partial \psi \partial \psi^\top}, \\ B_n(\psi, \alpha) &= -\frac{1}{n} \sum_{i=1}^n \delta_i \frac{1-r_i(\alpha)}{r_i(\alpha)} \dot{\ell}_i(\psi) \mathbf{S}_i^\top, \\ J_n(\psi, \alpha) &= \frac{1}{n} \sum_{i=1}^n \delta_i \frac{\dot{\ell}_i(\psi)\dot{\ell}_i(\psi)^\top}{r_i^2(\alpha)}, \\ \Sigma_n(\alpha) &= \frac{1}{n} \sum_{i=1}^n \mathbf{S}_i \mathbf{S}_i^\top r_i(\alpha)(1-r_i(\alpha)). \end{aligned}$$

The proof proceeds along the same lines as proof of Lemma 3.4 and is therefore omitted.

4. Simulation study

In this section, we investigate the finite-sample performances of the IPW estimator under various conditions.

4.1. Simulation design

The following ZIB regression model is used to simulate data:

$$\text{logit}(\pi_i) = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6}$$

and

$$\text{logit}(p_i) = \gamma_1 W_{i1} + \gamma_2 W_{i2} + \gamma_3 W_{i3} + \gamma_4 W_{i4},$$

where $X_{i1} = W_{i1} = 1$ and the X_{i2}, \dots, X_{i6} and W_{i4} are independently drawn from normal $\mathcal{N}(0, 1)$, uniform $\mathcal{U}(2, 5)$, normal $\mathcal{N}(1, 1.5)$, exponential $\mathcal{E}(1)$, binomial $\mathcal{B}(1, 0.3)$ and normal $\mathcal{N}(-1, 1)$ distributions respectively. Linear predictors in $\text{logit}(\pi_i)$ and $\text{logit}(p_i)$ are allowed to share common terms by letting $W_{i2} = X_{i2}$ and $W_{i3} = X_{i6}$. The regression parameter β is chosen as $\beta = (-0.3, 1.2, 0.5, -0.75, -1, 0.8)^\top$. The regression parameter γ is chosen as:

- case 1: $\gamma = (-0.55, -0.7, -1, 0.45)^\top$
- case 2: $\gamma = (0.25, -0.4, 0.8, 0.45)^\top$

We consider the following sample sizes, $n = 500, 1000$. The numbers m_i are allowed to vary across subjects, with $m_i \in \{4, 8, 10, 15\}$. Let $n_j = \text{card}\{i : m_i = j\}$, for $j = 4, 8, 10, 15$. For $n = 500$, we let $(n_4, n_8, n_{10}, n_{15}) = (125, 125, 125, 125)$ and for $n = 1000$, we let $(n_4, n_8, n_{10}, n_{15}) = (250, 250, 250, 250)$.

Using these values, in case 1 (respectively case 2), the average percentage of zero-inflation in the simulated data sets is 25% (respectively 50%). Missingness indicators δ_i are simulated from a logistic regression model with selection probability $r_i(\alpha) := \mathbb{P}(\delta_i = 1 | \mathbf{S}_i) = \text{logit}^{-1}(\alpha^\top \mathbf{S}_i)$, with $\mathbf{S}_i := (1, Z_i, X_{i2}, W_{i4})$. The regression parameter α is chosen to yield average missingness proportions in the simulated samples successively equal to 0.2 and 0.4. Finally, for each combination of the simulation design parameters (sample size, proportions of zero-inflation and missing data), we simulate $N = 1000$ samples and we calculate the IPW estimate $\hat{\psi}_n$. Simulations are carried out using the statistical software R. We use the package `maxLik` (Henningsen and Toomet, 2011) to solve the estimating equation (2.5).

4.2. Results

For each configuration [sample size \times zero-inflation proportion \times proportion of missing data] of the simulation parameters, we calculate the average absolute relative bias (as a percentage) of the estimates $\hat{\beta}_{j,n}$ and $\hat{\gamma}_{k,n}$ over the N simulated samples (for example, the absolute relative bias of $\hat{\beta}_{j,n}$ is obtained as $N^{-1} \sum_{t=1}^N |(\hat{\beta}_{j,n}^{(t)} - \beta_j)/\beta_j| \times 100$, where $\hat{\beta}_{j,n}^{(t)}$ denotes the IPW estimate of β_j in the t -th simulated sample). We also obtain the average standard error SE (calculated from (3.13)), empirical standard deviation (SD) and root mean square error (RMSE) for each estimator $\hat{\beta}_{j,n}$ ($j = 1, \dots, 6$) and $\hat{\gamma}_{k,n}$ ($k = 1, \dots, 4$). Finally, we provide the empirical coverage probability (CP) of 95%-level confidence intervals for the β_j and γ_k . Results are given in Table 1 (case 1, $n = 500$), Table 2 (case 1, $n = 1000$), Table 3 (case 2, $n = 500$), Table 4 (case 2, $n = 1000$). For purpose of comparison, we also

provide results for the MLE that would be obtained if there were no missing covariates. This estimator solves the score equation $U_n(\psi) = 0$ given by (2.2) (in what follows, we refer this estimator to as the "full data" - or FD - estimator). The complete-case estimator of ψ can be obtained by solving the score equation (2.2), based on complete-cases only. However, this estimator is generally highly biased. For example, in our simulations, its relative bias can reach up to 200%, resulting in very low coverage probabilities. Therefore, we do not provide results for this estimate.

From these results, we observe, as expected, that the bias, SE, SD and RMSE of the IPW estimator decrease as the sample size increases and the proportion of individuals with missing covariates decreases. Moreover, the bias of the IPW estimator stays moderate and its empirical coverage probabilities are close to the nominal confidence level, even when the sample size is moderate ($n = 500$). As may also be expected, for a given proportion of missing data, we observe that the IPW estimator of the β_{js} (respectively γ_{ks}) performs better when the zero-inflation proportion decreases (respectively increases). The FD estimator obviously performs better than the IPW estimator, but FD analysis cannot be performed when missing data are present. Overall, these numerical results indicate the good performance of the IPW method for estimating a ZIB regression model under missing data.

Finally, in order to assess the quality of the Gaussian approximation stated in Theorem 3.5, we provide normal Q-Q plots of the estimates. See figures 1 and 2 for $n = 500$ in case 2 and a fraction of missing data equal to 0.4 and figures 3 and 4 for $n = 1000$ in case 2 and a fraction of missing data equal to 0.2 (plots for the other simulated scenarios yield similar observations and are thus omitted). From these figures, it appears that the Gaussian approximation of the distribution of the IPW estimator is reasonably satisfied, even when the sample size is moderate and the proportion of individuals with missing covariates is as high as 0.4.

average fraction of missing data		$\widehat{\beta}_n$						$\widehat{\gamma}_n$				
		$\widehat{\beta}_{1,n}$	$\widehat{\beta}_{2,n}$	$\widehat{\beta}_{3,n}$	$\widehat{\beta}_{4,n}$	$\widehat{\beta}_{5,n}$	$\widehat{\beta}_{6,n}$	$\widehat{\gamma}_{1,n}$	$\widehat{\gamma}_{2,n}$	$\widehat{\gamma}_{3,n}$	$\widehat{\gamma}_{4,n}$	
FD		rel. bias	3.4726	0.2880	0.1150	0.3161	0.5336	0.8380	1.8986	0.9012	1.9091	1.5690
		SD	0.1944	0.0611	0.0519	0.0361	0.0575	0.0964	0.1785	0.1509	0.3360	0.1361
		SE	0.1979	0.0591	0.0536	0.0370	0.0592	0.0981	0.1854	0.1513	0.3264	0.1372
		RMSE	0.2775	0.0850	0.0746	0.0517	0.0827	0.1376	0.2575	0.2138	0.4687	0.1933
		CP	0.9629	0.9459	0.9569	0.9609	0.9619	0.9559	0.9609	0.9559	0.9429	0.9599
IPW	0.2	rel. bias	2.7202	0.4861	0.0590	0.4251	0.6387	0.6311	3.6159	0.9854	4.0350	2.9869
		SD	0.2103	0.0649	0.0554	0.0386	0.0643	0.1047	0.2135	0.1965	0.4653	0.1865
		SE	0.2150	0.0640	0.0582	0.0408	0.0666	0.1057	0.2167	0.1925	0.4466	0.1799
		RMSE	0.3008	0.0913	0.0804	0.0563	0.0927	0.1488	0.3048	0.2751	0.6460	0.2594
		CP	0.9620	0.9550	0.9610	0.9650	0.9610	0.9570	0.9700	0.9560	0.9520	0.9490
IPW	0.4	rel. bias	0.4867	0.3730	0.7212	0.6061	0.7341	0.8895	3.4735	2.5081	5.9316	3.5603
		SD	0.2491	0.0817	0.0664	0.0462	0.0759	0.1234	0.2453	0.2391	0.5275	0.2178
		SE	0.2537	0.0802	0.0684	0.0476	0.0760	0.1261	0.2569	0.2450	0.5602	0.2249
		RMSE	0.3555	0.1146	0.0954	0.0665	0.1077	0.1765	0.3556	0.3427	0.7716	0.3134
		CP	0.9479	0.9550	0.9530	0.9550	0.9479	0.9550	0.9690	0.9700	0.9690	0.9670

Table 1: Simulation results (case 1, $n = 500$). SD: empirical standard deviation. SE: average standard error. RMSE: empirical root mean square error. CP: empirical coverage probability of 95%-level confidence intervals.

average fraction of missing data		$\widehat{\beta}_n$						$\widehat{\gamma}_n$				
		$\widehat{\beta}_{1,n}$	$\widehat{\beta}_{2,n}$	$\widehat{\beta}_{3,n}$	$\widehat{\beta}_{4,n}$	$\widehat{\beta}_{5,n}$	$\widehat{\beta}_{6,n}$	$\widehat{\gamma}_{1,n}$	$\widehat{\gamma}_{2,n}$	$\widehat{\gamma}_{3,n}$	$\widehat{\gamma}_{4,n}$	
FD	rel. bias	0.6276	0.0786	0.1061	0.2259	0.0748	0.4154	1.1894	0.2426	0.7370	0.2883	
	SD	0.1346	0.0416	0.0370	0.0242	0.0423	0.0708	0.1279	0.1040	0.2341	0.0954	
	SE	0.1389	0.0415	0.0376	0.0260	0.0414	0.0690	0.1300	0.1057	0.2270	0.0958	
	RMSE	0.1933	0.0588	0.0528	0.0356	0.0592	0.0989	0.1824	0.1482	0.3261	0.1352	
	CP	0.9570	0.9510	0.9570	0.9640	0.9449	0.9479	0.9560	0.9499	0.9540	0.9560	
	IPW	0.2	rel. bias	0.2160	0.1263	0.1224	0.2970	0.0850	0.3158	1.4117	0.6978	2.5321
		SD	0.1460	0.0446	0.0399	0.0259	0.0465	0.0771	0.1472	0.1323	0.3032	0.1267
		SE	0.1505	0.0448	0.0408	0.0284	0.0475	0.0748	0.1511	0.1357	0.3050	0.1269
		RMSE	0.2096	0.0632	0.0570	0.0385	0.0665	0.1074	0.2110	0.1895	0.4308	0.1794
		CP	0.9590	0.9450	0.9540	0.9670	0.9500	0.9500	0.9610	0.9600	0.9590	0.9550
IPW	0.4	rel. bias	0.7322	0.3466	0.2252	0.2982	0.5650	0.2990	0.7791	0.9170	3.8227	2.8852
		SD	0.1712	0.0537	0.0462	0.0330	0.0512	0.0868	0.1665	0.1647	0.3638	0.1514
		SE	0.1774	0.0556	0.0480	0.0334	0.0532	0.0875	0.1728	0.1625	0.3537	0.1490
		RMSE	0.2465	0.0774	0.0666	0.0470	0.0741	0.1233	0.2400	0.2314	0.5087	0.2128
		CP	0.9620	0.9560	0.9600	0.9520	0.9610	0.9580	0.9590	0.9500	0.9540	0.9440

Table 2: Simulation results (case 1, $n = 1000$). SD: empirical standard deviation. SE: average standard error. RMSE: empirical root mean square error. CP: empirical coverage probability of 95%-level confidence intervals.

average fraction of missing data		$\widehat{\beta}_n$						$\widehat{\gamma}_n$				
		$\widehat{\beta}_{1,n}$	$\widehat{\beta}_{2,n}$	$\widehat{\beta}_{3,n}$	$\widehat{\beta}_{4,n}$	$\widehat{\beta}_{5,n}$	$\widehat{\beta}_{6,n}$	$\widehat{\gamma}_{1,n}$	$\widehat{\gamma}_{2,n}$	$\widehat{\gamma}_{3,n}$	$\widehat{\gamma}_{4,n}$	
FD	rel. bias	0.8028	0.7069	0.5104	0.7715	0.9714	0.0022	0.0209	0.0103	0.2573	0.6092	
	SD	0.2427	0.0749	0.0677	0.0461	0.0736	0.1352	0.1711	0.1162	0.2346	0.1118	
	SE	0.2475	0.0738	0.0675	0.0469	0.0759	0.1410	0.1669	0.1164	0.2293	0.1110	
	RMSE	0.3466	0.1054	0.0957	0.0660	0.1061	0.1953	0.2389	0.1645	0.3280	0.1575	
	CP	0.9559	0.9349	0.9529	0.9529	0.9499	0.9549	0.9499	0.9529	0.9449	0.9439	
IPW	0.2	rel. bias	1.0135	0.7771	0.5658	0.8772	1.0211	0.1457	1.2474	0.3254	0.2989	0.8908
		SD	0.2524	0.0787	0.0705	0.0477	0.0754	0.1386	0.1833	0.1277	0.2522	0.1204
		SE	0.2563	0.0771	0.0698	0.0487	0.0788	0.1453	0.1833	0.1289	0.2551	0.1207
		RMSE	0.3596	0.1106	0.0992	0.0684	0.1095	0.2007	0.2592	0.1814	0.3586	0.1705
		CP	0.9500	0.9310	0.9510	0.9450	0.9550	0.9580	0.9440	0.9600	0.9550	0.9530
IPW	0.4	rel. bias	1.3807	0.9649	0.5492	1.0489	0.9995	0.5640	1.6315	4.0769	2.0604	5.0802
		SD	0.2771	0.0872	0.0773	0.0545	0.0938	0.1549	0.2181	0.1845	0.3650	0.1812
		SE	0.2888	0.0854	0.0789	0.0559	0.0928	0.1631	0.2182	0.1805	0.3557	0.1778
		RMSE	0.4001	0.1226	0.1105	0.0784	0.1323	0.2249	0.3085	0.2586	0.5099	0.2548
		CP	0.9640	0.9460	0.9590	0.9550	0.9460	0.9610	0.9510	0.9470	0.9440	0.9540

Table 3: Simulation results (case 2, $n = 500$). SD: empirical standard deviation. SE: average standard error. RMSE: empirical root mean square error. CP: empirical coverage probability of 95%-level confidence intervals.

average fraction of missing data		$\widehat{\beta}_n$						$\widehat{\gamma}_n$				
		$\widehat{\beta}_{1,n}$	$\widehat{\beta}_{2,n}$	$\widehat{\beta}_{3,n}$	$\widehat{\beta}_{4,n}$	$\widehat{\beta}_{5,n}$	$\widehat{\beta}_{6,n}$	$\widehat{\gamma}_{1,n}$	$\widehat{\gamma}_{2,n}$	$\widehat{\gamma}_{3,n}$	$\widehat{\gamma}_{4,n}$	
FD		rel. bias	0.8391	0.1674	0.0338	0.4053	0.1128	0.0499	1.9813	1.7535	1.1359	0.9103
		SD	0.1688	0.0518	0.0461	0.0316	0.0552	0.0969	0.1194	0.0837	0.1668	0.0801
		SE	0.1731	0.0514	0.0472	0.0328	0.0528	0.0985	0.1175	0.0818	0.1615	0.0780
		RMSE	0.2417	0.0730	0.0660	0.0456	0.0764	0.1382	0.1676	0.1172	0.2323	0.1119
		CP	0.9549	0.9529	0.9599	0.9619	0.9379	0.9579	0.9449	0.9339	0.9409	0.9409
IPW	0.2	rel. bias	0.7979	0.1926	0.0279	0.4838	0.1741	0.1785	1.2374	1.6033	1.6918	0.9563
		SD	0.1738	0.0534	0.0474	0.0322	0.0566	0.1004	0.1272	0.0911	0.1811	0.0866
		SE	0.1814	0.0552	0.0495	0.0343	0.0556	0.1031	0.1272	0.0893	0.1823	0.0858
		RMSE	0.2512	0.0768	0.0686	0.0472	0.0793	0.1439	0.1799	0.1277	0.2572	0.1220
		CP	0.9570	0.9520	0.9570	0.9630	0.9540	0.9640	0.9550	0.9440	0.9520	0.9490
IPW	0.4	rel. bias	2.6218	0.2938	0.2344	0.5926	0.1376	0.4597	4.6748	3.3595	0.6924	3.6510
		SD	0.1964	0.0584	0.0537	0.0374	0.0666	0.1105	0.1536	0.1246	0.2435	0.1308
		SE	0.2005	0.0596	0.0547	0.0391	0.0647	0.1133	0.1516	0.1251	0.2482	0.1262
		RMSE	0.2807	0.0835	0.0766	0.0542	0.0928	0.1583	0.2161	0.1770	0.3476	0.1824
		CP	0.9540	0.9480	0.9570	0.9600	0.9470	0.9560	0.9500	0.9480	0.9540	0.9420

Table 4: Simulation results (case 2, $n = 1000$). SD: empirical standard deviation. SE: average standard error. RMSE: empirical root mean square error. CP: empirical coverage probability of 95%-level confidence intervals.

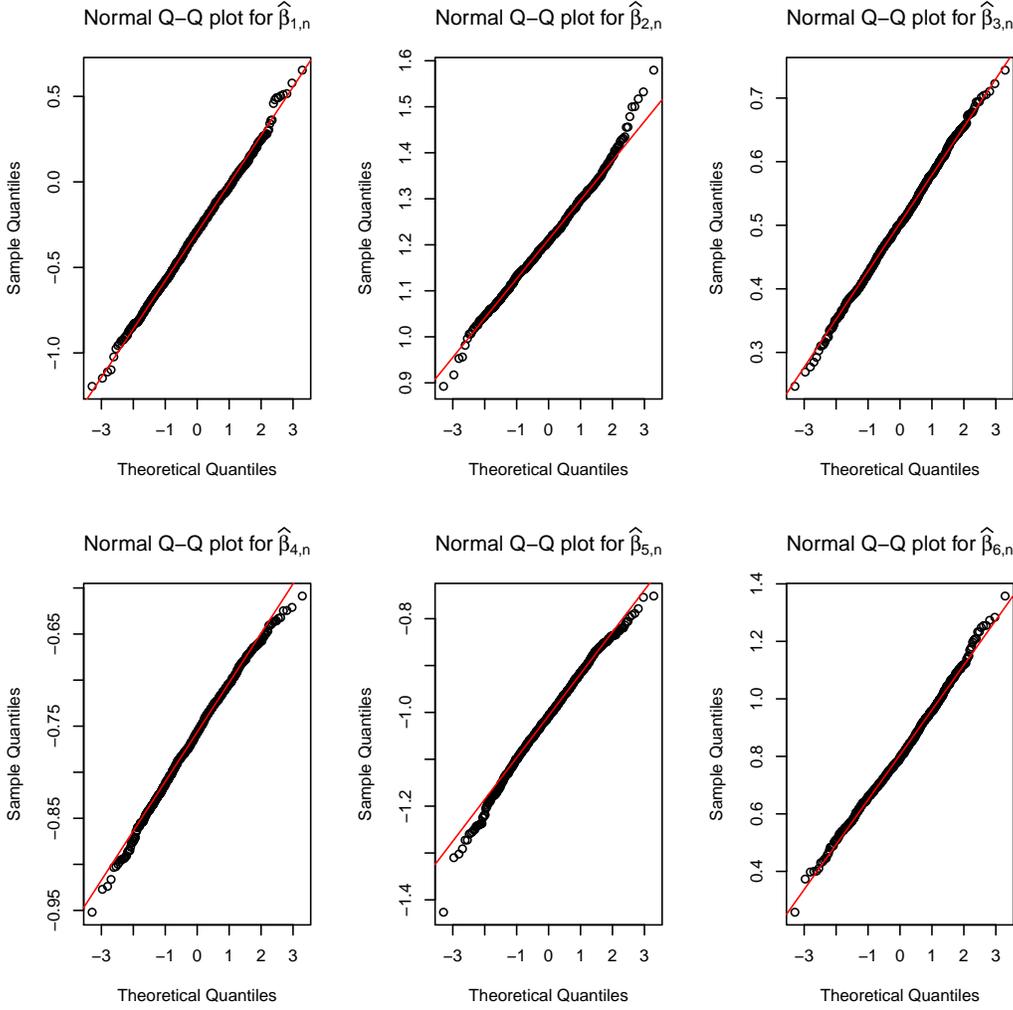


Figure 1: Normal Q-Q plots for $\hat{\beta}_{1,n}, \dots, \hat{\beta}_{6,n}$ with $n = 500$ (case 2) and a fraction of missing data equal to 0.4.

5. Discussion

Zero-inflated binomial (ZIB) regression is now commonly used for investigating count data with excess of zeros; see, for example, Gilthorpe *et al.* (2009), Matranga *et al.* (2013), Diallo *et al.* (2017a). In this paper, we extend the scope of ZIB regression by considering the situation where some covariates are missing at random. In this setting, we propose an inverse-probability-weighted-type estimator by assuming that the missingness probabilities can be modeled parametrically. Consistency and asymptotic normality of the proposed estimator are established and a constant variance estimator is constructed. Our simulation study suggests that the IPW estimator performs well under a wide range of conditions.

Now, several issues deserve attention. First, the proposed estimator is valid if the parametric model for missingness probabilities $\mathbb{P}(\delta_i = 1|\mathbf{S}_i)$ is correctly specified. Misspecifying this model may lead to a biased IPW estimator. Several solutions to this issue might be investigated. For example, one may consider semi- or nonparametric estimation of the missingness probabilities. An alternative approach relies on the so-called augmented IPW method, which is robust to a misspecification of the selection probabilities. Some additional work is now needed to investigate the relative merits of these approaches.

Robustness of these various estimation methods to a violation of the MAR assumption also constitutes a topic of great interest in view of applications.

Another stimulating topic for future work is as follows. In this paper, we consider missing covariates in the basic ZIB regression model (2.1). The same issue could be investigated in various generalizations of ZIB regression (such as the random-effects ZIB model proposed by Hall (2000), or semi-parametric ZIB models).

As a conclusion, the present work constitutes a promising first step towards the analysis of zero-inflated binomial counts with missing data. Further research is now needed to extend this contribution to more complex ZIB models and to more sophisticated and robust estimation methods.

Acknowledgements

Authors acknowledge financial support from the "Service de Coopération et d'Action Culturelle" of the French Embassy in Senegal and logistical support from Campus France (French national agency for the promotion of higher education, international student services, and international mobility). Authors also acknowledge grants from CEA-MITIC (an African Center of Excellence in Mathematics, Informatics) and ICT, implemented by Gaston Berger University (Senegal).

References

- Chen, X.-D., Fu, Y.-Z., 2011. Model selection for zero-inflated regression with missing covariates. *Computational Statistics & Data Analysis* 55(1), 765-773.
- Diallo, A. O, Diop, A., Dupuy, J.-F., 2017a. Asymptotic properties of the maximum likelihood estimator in zero-inflated binomial regression. *Communications in Statistics - Theory and Methods*, to appear, doi 10.1080/03610926.2016.1222437.
- Diallo, A. O, Diop, A., Dupuy, J.-F., 2017b. Analysis of multinomial counts with joint zero-inflation, with an application to health economics. Submitted.
- Dietz, E., Böhning, D., 2000. On estimation of the Poisson parameter in zero-modified Poisson models. *Computational Statistics & Data Analysis* 34(4), 441-459.
- Diop, A., Diop, A., Dupuy, J.-F., 2011. Maximum likelihood estimation in the logistic regression model with a cure fraction. *Electronic Journal of Statistics* 5, 460-483.

- Diop, A., Diop, A., Dupuy, J.-F., 2016. Simulation-based inference in a zero-inflated Bernoulli regression model. *Communications in Statistics - Simulation and Computation*, 45(10), 3597-3614.
- Eicker, F., 1966. A multivariate central limit theorem for random linear vector forms. *The Annals of Mathematical Statistics* 37(6), 1825-1828.
- Feng, J., Zhu, Z., 2011. Semiparametric analysis of longitudinal zero-inflated count data. *Journal of Multivariate Analysis* 102, 61-72.
- Foutz, R. V., 1977. On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association* 72, 147-148.
- Garay, A. M., Hashimoto, E. M., Ortega, E. M. M., Lachos, V. H., 2011. On estimation and influence diagnostics for zero-inflated negative binomial regression models. *Computational Statistics & Data Analysis* 55(3), 1304-1318.
- Gilthorpe, M. S., Frydenberg, M., Cheng, Y., Baelum, V., 2009. Modelling count data with excessive zeros: The need for class prediction in zero-inflated models and the issue of data generation in choosing between zero-inflated and generic mixture models for dental caries data. *Statistics in Medicine* 28, 3539-3553.
- Gouriéroux, C., Monfort, A., 1981. Asymptotic properties of the maximum likelihood estimator in dichotomous logit models. *Journal of Econometrics* 102, 17, 83-97.
- Hall, D. B., 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics* 56(4), 1030-1039.
- Henningsen, A., Toomet, O., 2011. maxLik: A package for maximum likelihood estimation in R. *Computational Statistics* 26(3), 443-458.
- Horvitz, D.G., Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47, 663-685.
- Hsieh, S. H., Lee, S. M., Shen, P. S., 2010. Logistic regression analysis of randomized response data with missing covariates. *Journal of Statistical Planning and Inference* 140(4), 927-940.
- Jiang, J., 2010. *Large Sample Techniques for Statistics*. Springer, New York.
- Lam, K. F., Xue, H., Cheung, Y. B., 2006. Semiparametric analysis of zero-inflated count data. *Biometrics* 62(4), 996-1003.
- Lambert, D., 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing.
- Li, T., Hu, Y., 2010. Inverse probability weighted estimators for single-index models with missing covariates. *Communications in Statistics. Theory and Methods* 45(5), 1199-1214.

- Lim, H. K., Li, W. K., Yu, P. L. H., 2006. Zero-inflated Poisson regression mixture model. *Computational Statistics & Data Analysis* 71, 151-158.
- Lukusa, T. M., Lee, S.-M., Li, C.-S., 2016. Semiparametric estimation of a zero-inflated Poisson regression model with missing covariates. *Metrika* 79(4), 457-483.
- Matranga, D., Firenze, A., Vullo, A., 2013. Can bayesian models play a role in dental caries epidemiology? Evidence from an application to the BELCAP data set. *Community Dentistry and Oral Epidemiology* 41(5), 473-480.
- Min, Y., Agresti, A., 2005. Random effect models for repeated measures of zero-inflated count data. *Statistical Modelling* 5(1), 1-19.
- Moghimbeigi, A., Eshraghian, M. R., Mohammad, K., McArdle, B., 2008. Multilevel zero-inflated negative binomial regression modeling for over-dispersed count data with extra zeros. *Journal of Applied Statistics* 35(9), 1193-1202.
- Monod, A., 2014. Random effects modeling and the zero-inflated Poisson distribution. *Communications in Statistics. Theory and Methods* 43 (4), 664-680. *Technometrics* 34, 1-14.
- Mwalili, S. M., Lesaffre, E., Declerck, D., 2008. The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research. *Statistical Methods in Medical Research* 17(2), 123-139.
- Qi, L., Wang, C. Y., Prentice, R. L., 2005. Weighted estimators for proportional hazards regression with missing covariates. *Journal of the American Statistical Association* 100(472), 1250-1263.
- R Core Team, 2013. R Foundation for Statistical Computing. Vienna, Austria <http://www.R-project.org/>.
- Ridout, M., Hinde, J., Demetrio, C. G. B., 2001. A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics* 57(1), 219-223.
- Rubin, D. B., 1976. Inference and missing data. *Biometrika* 63(3), 581-592.
- Vieira, A. M. C., Hinde, J. P., Demetrio, C. G. B., 2000. Zero-inflated proportion data models applied to a biological control assay. *Journal of Applied Statistics* 27(3), 373-389.
- Zhao, L. P., Lipsitz, S., 1992. Designs and analysis of two-stage studies. *Statistics in Medicine* 11(6), 769-782.
- Zhao, W., Zhang, R., Liu, J., Lv, Y., 2015. Semi varying coefficient zero-inflated generalized Poisson regression model. *Communications in Statistics. Theory and Methods* 44(1), 171-185.

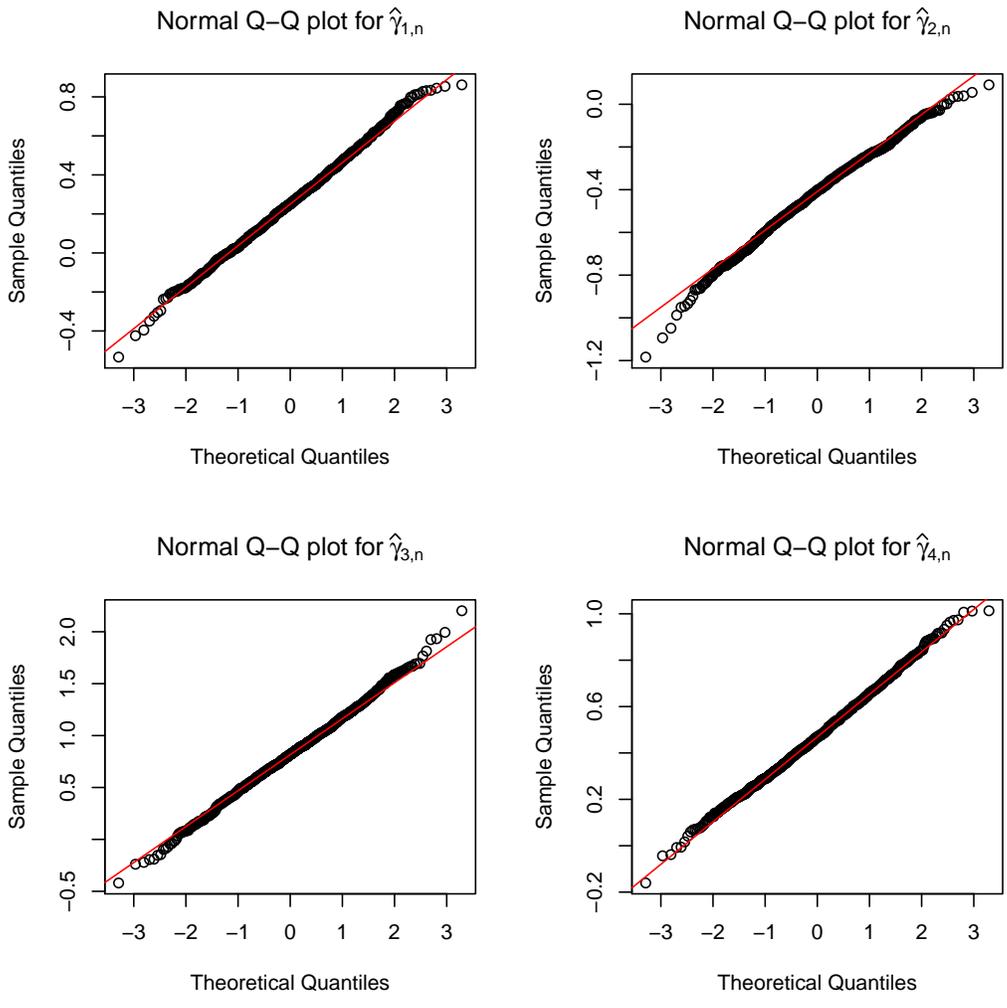


Figure 2: Normal Q-Q plots for $\hat{\gamma}_{1,n}, \dots, \hat{\gamma}_{4,n}$ with $n = 500$ (case 2) and a fraction of missing data equal to 0.4.

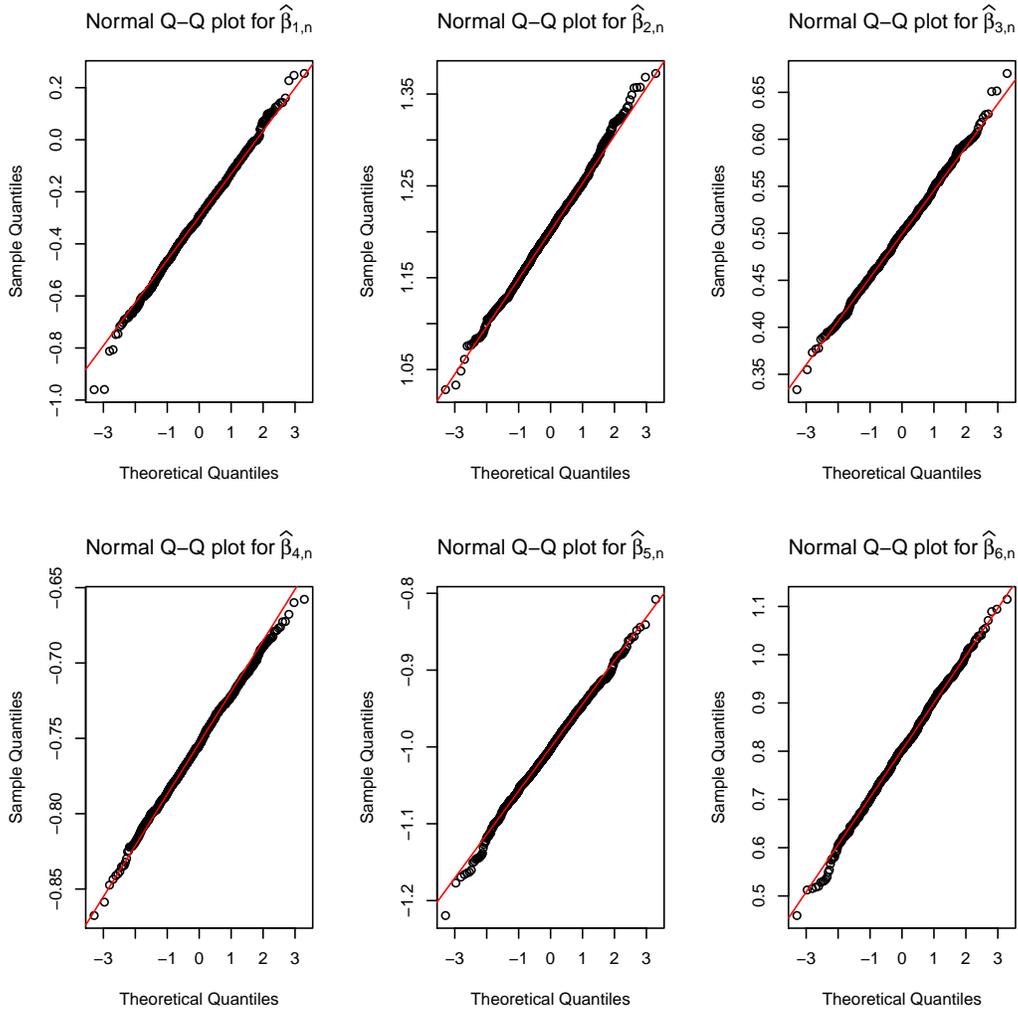


Figure 3: Normal Q-Q plots for $\hat{\beta}_{1,n}, \dots, \hat{\beta}_{6,n}$ with $n = 1000$ (case 2) and a fraction of missing data equal to 0.2.

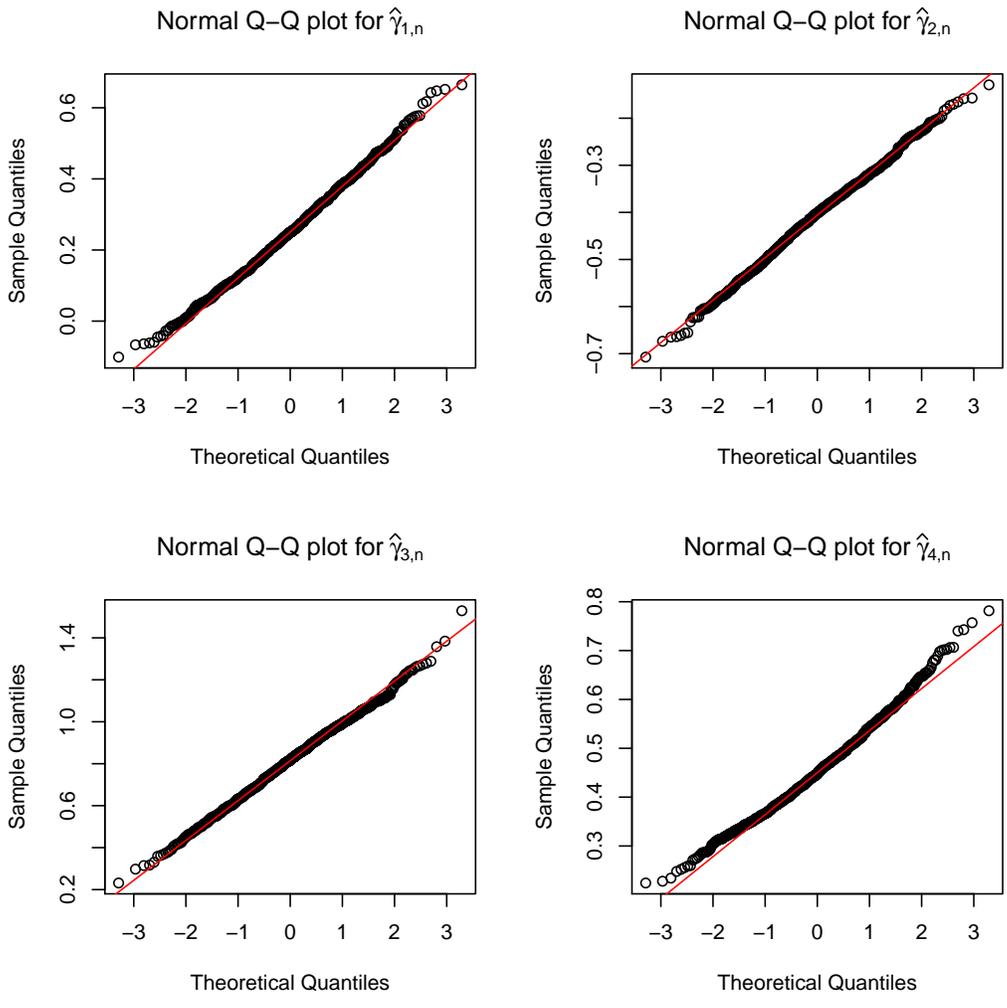


Figure 4: Normal Q-Q plots for $\hat{\gamma}_{1,n}, \dots, \hat{\gamma}_{4,n}$ with $n = 1000$ (case 2) and a fraction of missing data equal to 0.2.