Chilean Journal of Statistics
Volume 16, Number 1, June 2025, Pages 36–59
DOI: 10.32372/ChJS.16-01-03

Statistical Modeling and Missing Data
Research Article

Statistical inference for the marginalized zero-inflated Bell regression model with missing covariates

Konan J.G. Kouakou[1,*], Kouakou Mathias Amani[1], Lanciné Bamba[2], Brahima Soro[1,3], and Ouagnina Hili[1]

[1]Laboratoire de Statistique, Probabilités et Recherche Opérationnelle, UMRI-Mathématiques et Sciences du Numérique, Institut National Polytechnique Félix Houphouët-Boigny (INP-HB), Yamoussoukro, Côte d'Ivoire

[2] Unité de Formation et de Recherche des Sciences Fondamentales et Appliquées, Laboratoire de Mathématiques et Informatique (LMI), Université Nangui Abrogoua, Abidjan, Côte d'Ivoire

[3]Département de Mathématiques, Physique et Chimie, Université Peleforo Gon Coulibaly, Korhogo, Côte d'Ivoire

(Received: 12 October 2024 · Accepted in final form: 15 June 2025)

Abstract

Zero-inflated regression models are widely used to relate covariates to count outcomes that exhibit an excess of zeros. In their marginal formulations, these models yield direct inference on the population mean instead of on an unobserved susceptible subpopulation. The recently proposed marginalized zero-inflated Bell (MZIBell) model provides a parsimonious alternative for highly overdispersed counts, yet the available theory assumes that all covariates are fully observed. This article extends the MZIBell framework to settings in which covariates are missing at random. Parameter estimation is carried out through parametric inverse probability weighting, and the resulting estimators are shown to be consistent and asymptotically normal under standard regularity conditions. Monte Carlo experiments confirm good finite-sample performance even when more than one third of the covariate information is absent. An application to physician-visit counts from the Health Insurance Experiment of the Research and Development Corporation demonstrates that the proposed inverse-probability-weighted MZIBell model achieves a superior fit, according to the Akaike and Bayesian information criteria, when compared with competing marginalized Poisson and negative binomial specifications.

Keywords: Asymptotic properties · Count data · Inverse probability weighting · Marginal models · Simulation study

Mathematics Subject Classification: Primary 62F12 · Secondary 62J12, 62F07

*Corresponding author. Email: geoffroy.kouakou@inphb.ci (K.J.G. Kouakou)

ISSN: 0718-7912 (print)/ISSN 0718-7920 (online)

http://www.soche.cl/chjs

1. Introduction

The analysis of survey data in health, insurance, and related fields is often complicated by two concurrent features: (i) count responses that are highly overdispersed and frequently zero-inflated, and (ii) missing values in key covariates.

Poisson regression is inadequate in such settings because it enforces equality between the mean and the variance [1, 2]. To accommodate extra-Poisson variation, several alternatives have been proposed, including the negative binomial and generalized Poisson regressions. Among these, the Bell regression model was advocated in [3, 4] as a parsimonious one-parameter alternative capable of handling overdispersion without an additional scale parameter.

When the variance inflation is partly driven by an excess of zeros, zero-inflated extensions of the Poisson, negative-binomial, and generalized-Poisson models are typically employed. However, the usual latent-class parameterization makes the regression coefficients difficult to interpret at the population level. To address this drawback, marginalized zero-inflated models have been developed —see, for example, the marginalized zero-inflated Poisson model in [5]. More recently, the marginalized zero-inflated Bell (MZIBell) regression model was introduced in [6], which yields population-level (marginal) mean parameters while retaining the flexibility of the Bell distribution for the count component.

In many empirical studies, however, some covariates are only partially observed. Missingness can arise for several reasons —nonresponse, recording errors, etcetera— and is widespread in medicine, insurance, sociology, economics, transportation, and other disciplines [7]. Restricting the analysis to complete cases is simple but can lead to biased inference [8]. Multiple imputation offers an alternative but may be computationally demanding [9]. Other approaches, such as likelihood-based methods [10, 11], require strong modeling assumptions. Despite these developments, the application of MZIBell in the presence of missing covariates has not yet been explored. This article fills that gap.

We extend the MZIBell framework to the case where covariates are missing at random by employing parametric inverse probability weighting (IPW) for estimation [12]. The proposed IPW estimators are shown to be consistent and asymptotically normal [13]. A comprehensive simulation study evaluates their finite-sample performance under various proportions of missingness, and an application to data from the Health Insurance Experiment (1974–1982) of the Research and Development (RAND) Corporation illustrates the practical gains of the proposed MZIBell model over competing marginal zero-inflated specifications.

The remainder of the article is organized as follows. Section 2 first introduces the MZIBell model and the corresponding maximum likelihood (ML) estimation procedure under complete data, and then extends it to handle covariates that are missing at random using IPW. In Section 3, we establishe the asymptotic properties of the proposed estimator under standard regularity conditions. In Section 4, a simulation study is presented, which evaluates the finite-sample performance of the method under various missingness scenarios. Section 5 illustrates the practical application of the methodology using data from the RAND Health Insurance Experiment. In Section 6, we provide the conclusions. Technical derivations are provided in Appendix A, whereas Appendix B provides the codes used in the `R` software for simulations and real data analysis.

2. MZIBell regression with missing covariates

2.1 *MZIBell regression model and notation*

Let (Y_i), for $i = 1, \ldots, n$, be a random sample on the probability space $(\boldsymbol{\omega}, \mathcal{F}, \mathbb{P})$.

For each i, the count variable Y_i is said to follow a Bell distribution with parameter $\theta > 0$ if its probability mass function is given by

$$\mathbb{P}(Y = y) = \exp\left(1 - \exp(\theta)\right) \frac{\theta^y B_y}{y!}, \quad y \in \mathbb{N}_0,$$

where B_y denotes the yth Bell number, which is defined as

$$B_k = \frac{1}{\mathrm{e}} \sum_{d=0}^{\infty} \frac{d^k}{d!}, \quad k \in \mathbb{N}_0,$$

and satisfies $B_0 = 1$, where "e" is the Euler number. Using the exponential generating function stated as

$$\sum_{k=0}^{\infty} \frac{B_k \theta^k}{k!} = \exp(\exp(\theta) - 1),$$

it follows that $\mathbb{E}(Y) = \theta \exp(\theta)$ and $\mathbb{V}\mathrm{ar}(Y) = \theta(1+\theta)\exp(\theta)$, so that the dispersion index $\mathbb{V}\mathrm{ar}(Y)/\mathbb{E}(Y) = 1 + \theta > 1$, for every $\theta > 0$.

Although the Bell distribution is originally parametrized by θ, we reparametrize it in terms of the marginal mean $\mu = \theta \exp(\theta)$, which will serve as the regression parameter in subsequent modeling.

Because regression modeling is usually formulated in terms of the mean, set $\mu = \theta \exp(\theta)$; then $\theta = W(\mu)$, where W is the principal branch of the Lambert W function [14]. Reparametrizing the probability mass function, we obtain

$$\mathbb{P}(Y = y) = \exp\left(1 - \exp(W(\mu))\right) \frac{W(\mu)^y B_y}{y!}, \quad y \in \mathbb{N}_0,$$

with $\mathbb{E}(Y) = \mu$ and $\mathbb{V}\mathrm{ar}(Y) = \mu\left(1 + W(\mu)\right)$, confirming overdispersion relative to the Poisson model.

Now, suppose Y_i arises from a mixture of a degenerate mass at zero and a Bell distribution, that is, the zero-inflated Bell (ZIBell) model is given by

$$\mathbb{P}(Y = k) = \begin{cases} \psi + (1-\psi)\exp\left(1 - \exp(W(\mu))\right), & k = 0; \\ (1-\psi)\exp\left(1 - \exp(W(\mu))\right) \dfrac{W(\mu)^k B_k}{k!}, & k > 0; \end{cases} \tag{1}$$

where $\mu > 0$ and $\psi \in (0, 1)$. From the expression stated in (1), we have that

$$\mathbb{E}(Y) = (1-\psi)\mu, \quad \mathbb{V}\mathrm{ar}(Y) = (1-\psi)\mu\left(1 + W(\mu) + \psi\mu\right).$$

The regression specification is formulated as

$$Y_i \sim \mathrm{ZIBell}(\mu_i, \psi_i), \quad \log(\mu_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}, \quad \mathrm{logit}(\psi_i) = \boldsymbol{z}_i^\top \boldsymbol{\gamma},$$

with vectors of covariate values $\boldsymbol{x}_i = (1, x_{i2}, \ldots, x_{ip})^\top$ and $\boldsymbol{z}_i = (1, x_{i2}, \ldots, x_{iq})^\top$, which may share components. While this latent-class formulation handles extra zeros, it limits its population-level interpretation of covariate effects, motivating the marginal approach developed in the next sections.

To address this limitation and allow for direct interpretation of parameters in terms of the marginal mean, the MZIBell regression model was introduced in [6]. This model retains the flexibility of the Bell distribution for the count component while enabling straightforward population-level inference.

Given that $\mathbb{E}(Y_i \mid X_i = x_i, Z_i = z_i) = \nu_i = \mu_i(1-\psi_i)$, the marginal mean ν_i is modeled directly. Since $\mu_i = \nu_i/(1-\psi_i)$, the probability mass function of an MZIBell distributed random variable is expressed as

$$\mathbb{P}(Y_i = k) = \begin{cases} \psi_i + (1-\psi_i)\exp\left(1 - \exp(W(\nu_i/(1-\psi_i)))\right), & k = 0; \\ (1-\psi_i)\exp\left(1 - \exp(W(\nu_i/(1-\psi_i)))\right)\dfrac{W\left(\nu_i/(1-\psi_i)\right)^k B_k}{k!}, & k > 0. \end{cases} \tag{2}$$

The parameters ν_i and ψ_i presented in (2) are modeled as $\text{logit}(\psi_i) = \boldsymbol{z}_i^\top \boldsymbol{\gamma}$ and $\log(\nu_i) = \boldsymbol{x}_i^\top \boldsymbol{\alpha}$, where $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_q)^\top$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)^\top$. Let $\boldsymbol{\Theta} = (\boldsymbol{\gamma}^\top, \boldsymbol{\alpha}^\top)^\top$.

2.2 *Likelihood function*

For the sample $\{Y_i, i = 1, \ldots, n\}$, define $J_i = \mathbb{1}_{\{Y_i=0\}}$, an indicator for zero responses. Then, the individual log-likelihood function is represented as

$$\begin{aligned} \ell_i(\boldsymbol{\Theta}) = &- \log(1 + \exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) + J_i \log(\exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) + \exp(1 - \exp(W(\xi_i)))) \\ &+ (1 - J_i)1 - \exp(W(\xi_i)) + \log(B_{Y_i}) - \log(Y_i)) \\ &+ (1 - J_i)Y_i(\log(1 + \exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) + \boldsymbol{x}_i^\top \boldsymbol{\alpha} - W(\xi_i)). \end{aligned}$$

Therefore, the log-likelihood function is $\ell_n(\boldsymbol{\Theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\Theta})$. Define the individual contribution as

$$\Psi_i(\boldsymbol{\Theta}) = \begin{pmatrix} \boldsymbol{z}_i B_i(\boldsymbol{\Theta}) \\ \boldsymbol{x}_i A_i(\boldsymbol{\Theta}) \end{pmatrix} \in \mathbb{R}^{q+p},$$

that is, a column vector obtained by stacking the component associated with $\boldsymbol{\gamma}$ on top of the component associated with $\boldsymbol{\alpha}$. The functions $A_i(\boldsymbol{\Theta})$ and $B_i(\boldsymbol{\Theta})$ are stated as

$$\begin{aligned} A_i(\boldsymbol{\Theta}) = &J_i \left(\frac{\exp(\boldsymbol{x}_i^\top \boldsymbol{\alpha})(1 + \exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))}{h_i(\boldsymbol{\Theta}) g_i(\boldsymbol{\Theta})} \right) \\ &+ (1 - J_i) \left(Y_i \left(1 - \frac{\exp(\boldsymbol{x}_i^\top \boldsymbol{\alpha})(1 + \exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))}{g_i(\boldsymbol{\Theta})} \right) - \frac{\exp(\boldsymbol{x}_i^\top \boldsymbol{\alpha})(1 + \exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma}))}{g_i(\boldsymbol{\Theta})} \right), \\ B_i(\boldsymbol{\Theta}) = &- \frac{\exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma})}{1 + \boldsymbol{z}_i^\top \boldsymbol{\gamma})} + J_i \left(\frac{\exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) - \dfrac{\exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma} + \boldsymbol{x}_i^\top \boldsymbol{\alpha})}{g_i(\boldsymbol{\Theta})} \left(h_i(\boldsymbol{\Theta}) - \exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) \right)}{h_i(\boldsymbol{\Theta})} \right) \\ &+ (1 - J_i) \left(- \frac{\exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma} + \boldsymbol{x}_i^\top \boldsymbol{\alpha})}{g_i(\boldsymbol{\Theta})} + Y_i \left(\frac{\exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma})}{1 + \exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma})} \right) \right. \\ &\left. \times \left(1 - \frac{W((1 + \exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) \exp(\boldsymbol{x}_i^\top \boldsymbol{\alpha}))}{g_i(\boldsymbol{\Theta})} \right) \right), \end{aligned}$$

with

$$h_i(\boldsymbol{\Theta}) = \exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma}) + \exp(1 - \exp(W(\xi_i))), g_i(\boldsymbol{\Theta}) = 1 + W(\xi_i), \xi_i = (1 + \exp(\boldsymbol{z}_i^\top \boldsymbol{\gamma})) \exp(\boldsymbol{x}_i^\top \boldsymbol{\alpha}). \tag{3}$$

The score is defined as

$$U_n(\boldsymbol{\Theta}) = \sum_{i=1}^{n} \Psi_i(\boldsymbol{\Theta}). \tag{4}$$

The ML estimator $\widehat{\boldsymbol{\Theta}}_{\mathrm{ML}} = (\widehat{\boldsymbol{\gamma}}_n^\top, \widehat{\boldsymbol{\alpha}}_n^\top)^\top$ is obtained by solving the score equation $U_n(\boldsymbol{\Theta}) = 0$ obtained from the expression given in (4). Consistency and asymptotic normality of $\widehat{\boldsymbol{\Theta}}_{\mathrm{ML}}$ under the complete-data setting were established in [6].

2.3 *Regression model with missing covariates*

Let $\boldsymbol{x}_i = (\boldsymbol{x}_{i_{\mathrm{obs}}}^\top, \boldsymbol{x}_{i_{\mathrm{mis}}}^\top)^\top$ and $\boldsymbol{z}_i = (\boldsymbol{z}_{i_{\mathrm{obs}}}^\top, \boldsymbol{z}_{i_{\mathrm{mis}}}^\top)^\top$ denote, respectively, the observed and missing components of the covariate vectors for subject i, with $i = 1, \ldots, n$. Set $\boldsymbol{O}_i = (Y_i, \boldsymbol{x}_{i_{\mathrm{obs}}}^\top, \boldsymbol{z}_{i_{\mathrm{obs}}}^\top)^\top$, the vector that is always observed. Introduce the response indicator defined as

$$\delta_i = \begin{cases} 1, & \text{if all elements of } (\boldsymbol{x}_i, \boldsymbol{z}_i) \text{ are observed;} \\ 0, & \text{otherwise.} \end{cases}$$

Throughout, we assume a missing-at-random mechanism given by

$$\mathbb{P}(\delta_i = 1 \mid \boldsymbol{O}_i, \boldsymbol{x}_{i_{\mathrm{mis}}}, \boldsymbol{Z}_{i_{\mathrm{mis}}}) = \mathbb{P}(\delta_i = 1 \mid \boldsymbol{O}_i) = \pi(\boldsymbol{O}_i), \quad i = 1, \ldots, n.$$

Under missing-at-random, the IPW estimator of the parameter vector $\boldsymbol{\Theta} = (\boldsymbol{\gamma}^\top, \boldsymbol{\alpha}^\top)^\top$ is defined as the solution of the weighted score equation stated as

$$U_{W,n}(\boldsymbol{\Theta}, \pi) = \sum_{i=1}^{n} \frac{\delta_i}{\pi(\boldsymbol{O}_i)} \Psi_i(\boldsymbol{\Theta}) = \mathbf{0}, \tag{5}$$

where $\Psi_i(\boldsymbol{\Theta})$ is presented in (4).

2.4 *Parametric model for the selection probability*

Denote by $\pi(\boldsymbol{\omega}, \boldsymbol{O}_i)$ a parametric model for the selection probability, with unknown vector $\boldsymbol{\omega} \in \mathbb{R}^d$ and true value $\boldsymbol{\omega}_0$. A convenient choice is the logistic specification $\pi(\boldsymbol{\omega}, \boldsymbol{O}_i) = \mathrm{logit}^{-1}(\boldsymbol{\omega}^\top \boldsymbol{O}_i)$. Its ML estimator is stated as

$$\widehat{\boldsymbol{\omega}}_n = \arg\max_{\boldsymbol{\omega} \in \mathcal{B}} \left\{ \prod_{i=1}^{n} \pi(\boldsymbol{\omega}, \boldsymbol{O}_i)^{\delta_i} (1 - \pi(\boldsymbol{\omega}, \boldsymbol{O}_i))^{1-\delta_i} \right\}.$$

Under mild regularity conditions, we obtain

$$\sqrt{n}(\widehat{\boldsymbol{\omega}}_n - \boldsymbol{\omega}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i - \pi(\boldsymbol{\omega}_0, \boldsymbol{O}_i)}{\pi(\boldsymbol{\omega}_0, \boldsymbol{O}_i)\,(1 - \pi(\boldsymbol{\omega}_0, \boldsymbol{O}_i))} \Omega^{-1}(\boldsymbol{\omega}_0) \dot{\pi}(\boldsymbol{\omega}_0, \boldsymbol{O}_i) + o_{\mathbb{P}}(1), \tag{6}$$

where $o_{\mathbb{P}}(1)$ denotes a random term that converges to zero in probability as $n \to \infty$.

In the expression formulated in (6), we have

$$\dot{\pi}(\boldsymbol{\omega}, \boldsymbol{O}_i) = \frac{\partial \pi(\boldsymbol{\omega}, \boldsymbol{O}_i)}{\partial \boldsymbol{\omega}}, \quad \Omega(\boldsymbol{\omega}) = \mathbb{E}_{\boldsymbol{O}} \left(\frac{\dot{\pi}(\boldsymbol{\omega}, \boldsymbol{O}) \dot{\pi} \boldsymbol{\omega}, \boldsymbol{O})^\top}{\pi(\boldsymbol{\omega}, \boldsymbol{O}) \left(1 - \pi(\boldsymbol{\omega}, \boldsymbol{O})\right)} \right),$$

with $\Omega(\boldsymbol{\omega}_0)$ positive-definite so that $\Omega^{-1}(\boldsymbol{\omega}_0)$ exists.

Replacing $\pi(\boldsymbol{O}_i)$, for $i = 1, \ldots, n$, in the weighted score stated in (5) by its parametric estimate $\pi(\widehat{\boldsymbol{\omega}}_n, \boldsymbol{O}_i)$, we attain at

$$U_{W,n}(\boldsymbol{\Theta}, \widehat{\boldsymbol{\omega}}_n) = \sum_{i=1}^{n} \frac{\delta_i}{\pi(\widehat{\boldsymbol{\omega}}_n, \boldsymbol{O}_i)} \Psi_i(\boldsymbol{\Theta}),$$

with $\Psi_i(\boldsymbol{\Theta})$ defined in (4) and $\pi(\widehat{\boldsymbol{\omega}}_n, \boldsymbol{O}_i) = \text{logit}^{-1}(\widehat{\boldsymbol{\omega}}_n^\top \boldsymbol{O}_i)$. Solving the resulting estimating equation, we reach

$$U_{W,n}(\boldsymbol{\Theta}, \widehat{\boldsymbol{\omega}}_n) = \mathbf{0}, \tag{7}$$

which yields the IPW estimator $\widehat{\boldsymbol{\Theta}}_{W,n} = (\widehat{\boldsymbol{\gamma}}_{W,n}^\top, \widehat{\boldsymbol{\alpha}}_{W,n}^\top)^\top$.

3. Regularity conditions and large-sample properties

3.1 *Assumptions*

Next, we analyze the IPW estimator $\widehat{\boldsymbol{\Theta}}_{W,n}$ defined in (7). Let $\widehat{\boldsymbol{\omega}}_n \in \mathbb{R}^d$ be the ML estimator of the selection–model parameters $\boldsymbol{\omega}_0$ that appear in $\pi(\boldsymbol{\omega}, \boldsymbol{O}) = \mathbb{P}(\delta = 1 \mid \boldsymbol{O})$; see Subsection 2.3 for details on $\boldsymbol{O}$ and δ.

Consistency and asymptotic normality of $(\widehat{\boldsymbol{\Theta}}_{W,n}, \widehat{\boldsymbol{\omega}}_n)$ are established under the following regularity conditions:

- C1 Parameter spaces —The true values satisfy $\boldsymbol{\Theta}_0 \in \text{int}\,\text{Normal} \subset \mathbb{R}^{p+q}$ and $\boldsymbol{\omega}_0 \in \text{int}(\mathcal{B}) \subset \mathbb{R}^d$, where $\mathcal{N}$ and $\mathcal{B}$ are compact neighbourhoods, not depending on n.
- C2 Positive selection probability —For every $\boldsymbol{O}$ in the support of the always-observed data and every $\boldsymbol{\omega} \in \mathcal{B}$, we have $0 < \pi(\boldsymbol{\omega}, \boldsymbol{O}) < 1$.
- C3 Smoothness of the selection model (Lipschitz property) —The map $\boldsymbol{\omega} \mapsto \pi(\boldsymbol{\omega}, \boldsymbol{O})$ is continuously differentiable on $\mathcal{B}$ and Lipschitz in $\boldsymbol{\omega}$ so that $|\pi(\boldsymbol{\omega}, \boldsymbol{O}) - \pi(\widetilde{\boldsymbol{\omega}}, \boldsymbol{O})| \leq \kappa(\boldsymbol{O}) \|\boldsymbol{\omega} - \widetilde{\boldsymbol{\omega}}\|$ and $\mathbb{E}(\kappa(\boldsymbol{O})^2) < \infty$. (The function κ does not depend on $\boldsymbol{\omega}$.)
- C4 Moments of the individual score —With $\Psi_i(\boldsymbol{\Theta})$ being defined in (4), we reach $\mathbb{E}(\|\Psi_i(\boldsymbol{\Theta})\|^2) < \infty$ and $\mathbb{E}(\Psi_i(\boldsymbol{\Theta})\Psi_i(\boldsymbol{\Theta})^\top)$ is positive-definite for every $\boldsymbol{\Theta}$ in a neighbourhood of $\boldsymbol{\Theta}_0$.
- C5 Uniform convergence of the weighted Hessian —In a neighbourhood of $(\boldsymbol{\Theta}_0, \boldsymbol{\omega}_0)$ the first- and second-order partial derivatives of the weighted score given by

$$U_{W,n}(\boldsymbol{\Theta}, \boldsymbol{\omega}) = \sum_{i=1}^{n} \frac{\delta_i \Psi_i(\boldsymbol{\Theta})}{\pi(\boldsymbol{\omega}, \boldsymbol{O}_i)}$$

with respect to $\boldsymbol{\Theta}$ are dominated by an integrable function, and

$$-\frac{1}{n} \left(\frac{\partial U_{W,n}(\boldsymbol{\Theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\Theta}^\top} \right) \xrightarrow{\mathbb{P}} \Lambda(\boldsymbol{\Theta}, \boldsymbol{\omega}),$$

as $n \to \infty$, where $\xrightarrow{\mathbb{P}}$ denotes convergence in probability to and $\Lambda(\boldsymbol{\Theta}_0, \boldsymbol{\omega}_0)$ is positive-definite.

3.2 *Consistency and normality of the ML estimator*

The following theorem stated consistency and normality of the ML estimator.

THEOREM 1. Assume conditions C1—C5. Then, we have $\widehat{\boldsymbol{\Theta}}_{W,n} \xrightarrow{\mathbb{P}} \boldsymbol{\Theta}_0$ and $\sqrt{n}(\widehat{\boldsymbol{\Theta}}_{W,n} - \boldsymbol{\Theta}_0) \xrightarrow{\mathbb{D}} \text{Normal}(\mathbf{0}, \boldsymbol{\Delta}_W)$ and where $\xrightarrow{\mathbb{D}}$ denotes convergence in distribution to, $\boldsymbol{\Delta}_W = \boldsymbol{\Lambda}^{-1}(\Phi - \boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma}^\top)(\boldsymbol{\Lambda}^{-1})^\top$, with

$$
\begin{aligned}
\boldsymbol{\Lambda} &= \Lambda(\boldsymbol{\Theta}_0, \boldsymbol{\omega}_0), \\
\boldsymbol{\Phi} &= \mathbb{E}(\delta\Psi_i(\boldsymbol{\Theta}_0)\Psi_i(\boldsymbol{\Theta}_0)^\top/\pi(\boldsymbol{\omega}_0, \boldsymbol{O}_i)^2), \\
\boldsymbol{\Gamma} &= -\mathbb{E}(\delta\Psi_i(\boldsymbol{\Theta}_0)\dot{\pi}(\boldsymbol{\omega}_0, \boldsymbol{O}_i)^\top/\pi(\boldsymbol{\omega}_0, \boldsymbol{O}_i)^2), \\
\boldsymbol{\Omega} &= \mathbb{E}(\dot{\pi}(\boldsymbol{\omega}_0, \boldsymbol{O}_i)\dot{\pi}(\boldsymbol{\omega}_0, \boldsymbol{O}_i)^\top/(\pi(\boldsymbol{\omega}_0, \boldsymbol{O}_i)(1 - \pi(\boldsymbol{\omega}_0, \boldsymbol{O}_i)))),
\end{aligned}
$$

and $\dot{\pi}(\boldsymbol{\omega}, \boldsymbol{O}) = \partial\pi(\boldsymbol{\omega}, \boldsymbol{O})/\partial\boldsymbol{\omega}$.

Now, a consistent variance estimator is stated. A plug-in estimator of the asymptotic covariance matrix Δ_W defined in Theorem 1 is obtained by replacing population quantities with their sample analogues, defined as $\widehat{\boldsymbol{\Delta}}_{W,n} = \boldsymbol{\Lambda}_n^{-1}(\boldsymbol{\Phi}_n - \boldsymbol{\Gamma}_n\boldsymbol{\Omega}_n^{-1}\boldsymbol{\Gamma}_n^\top)(\boldsymbol{\Lambda}_n^{-1})^\top$, where

$$
\begin{aligned}
\boldsymbol{\Lambda}_n &= -\frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{\pi(\widehat{\boldsymbol{\omega}}_n, \boldsymbol{O}_i)} \frac{\partial^2 \ell_i(\widehat{\boldsymbol{\Theta}}_{W,n})}{\partial\boldsymbol{\Theta}\partial\boldsymbol{\Theta}^\top}, \\
\boldsymbol{\Phi}_n &= \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{\pi(\widehat{\boldsymbol{\omega}}_n, \boldsymbol{O}_i)^2} \Psi_i(\widehat{\boldsymbol{\Theta}}_{W,n})\Psi_i(\widehat{\boldsymbol{\Theta}}_{W,n})^\top, \\
\boldsymbol{\Gamma}_n &= -\frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i}{\pi(\widehat{\boldsymbol{\omega}}_n, \boldsymbol{O}_i)^2} \Psi_i(\widehat{\boldsymbol{\Theta}}_{W,n})\dot{\pi}(\widehat{\boldsymbol{\omega}}_n, \boldsymbol{O}_i)^\top, \\
\boldsymbol{\Omega}_n &= \frac{1}{n}\sum_{i=1}^{n} \frac{\dot{\pi}(\widehat{\boldsymbol{\omega}}_n, \boldsymbol{O}_i)\dot{\pi}(\widehat{\boldsymbol{\omega}}_n, \boldsymbol{O}_i)^\top}{\pi(\widehat{\boldsymbol{\omega}}_n, \boldsymbol{O}_i)\,(1 - \pi(\widehat{\boldsymbol{\omega}}_n, \boldsymbol{O}_i))}.
\end{aligned}
$$

Under conditions C1-C5, we have that $\widehat{\boldsymbol{\Delta}}_{W,n} \xrightarrow{\mathbb{P}} \boldsymbol{\Delta}_W$.

Proof [Theorem 1]
Let

$$
U_{W,n}(\boldsymbol{\Theta}, \boldsymbol{\omega}) = \sum_{i=1}^{n} \frac{\delta_i\Psi_i(\boldsymbol{\Theta})}{\pi(\boldsymbol{\omega}, \boldsymbol{O}_i)}, \quad \bar{U}_{W,n} = \frac{1}{n}U_{W,n}.
$$

We verify the three conditions of the Foutz inverse-function theorem [15]. These conditions concern (i) differentiability of the score, (ii) convergence of the average score to zero, and (iii) nonsingularity of the limit Hessian and are detailed as follows:

- F1 —Differentiability. By construction, $U_{W,n}$ is obtained from the individual log-likelihood contributions $\ell_i(\boldsymbol{\Theta})$ through sums, products and quotients of the smooth functions exp and log, and of the auxiliary terms $h_i(\boldsymbol{\Theta})$ and $g_i(\boldsymbol{\Theta})$ defined in (3). Both $h_i(\boldsymbol{\Theta})$ and $g_i(\boldsymbol{\Theta})$ involve the Lambert function W, which is $\mathcal{C}^\infty$ on $(-\exp(-1), \infty)$ (see Appendix A for the proof). Consequently, $U_{W,n}(\boldsymbol{\Theta}, \boldsymbol{\omega})$ is twice continuously differentiable in $\boldsymbol{\Theta}$, and therefore $\partial U_{W,n}/\partial\boldsymbol{\Theta}^\top$ is continuous in an open neighbourhood of $\boldsymbol{\Theta}_0$.

- F2 —Consistency of the score. Let $\bar{U}_{W,n} = n^{-1}U_{W,n}$ and decompose

$$\bar{U}_{W,n}(\boldsymbol{\Theta}_0, \widehat{\boldsymbol{\omega}}_n) = \underbrace{\left(\bar{U}_{W,n}(\boldsymbol{\Theta}_0, \widehat{\boldsymbol{\omega}}_n) - \bar{U}_{W,n}(\boldsymbol{\Theta}_0, \boldsymbol{\omega}_0)\right)}_{(a)} + \underbrace{\bar{U}_{W,n}(\boldsymbol{\Theta}_0, \boldsymbol{\omega}_0)}_{(b)}.$$

To control (a), using the definition of $U_{W,n}$ and the triangle inequality, we have

$$\|(a)\| \leq \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i \|\Psi_i(\boldsymbol{\Theta}_0)\|}{\pi(\boldsymbol{\omega}_0, \boldsymbol{O}_i)\pi(\widehat{\boldsymbol{\omega}}_n, \boldsymbol{O}_i)} |\pi(\boldsymbol{\omega}_0, \boldsymbol{O}_i) - \pi(\widehat{\boldsymbol{\omega}}_n, \boldsymbol{O}_i)|.$$

Note that conditions C2 (positivity of π) and C4 (finite second moment of Ψ_i) guarantee the existence of a constant $k_1 > 0$ such that

$$\|(a)\| \leq k_1 \left(\frac{1}{n}\sum_{i=1}^{n} |\pi(\boldsymbol{\omega}_0, \boldsymbol{O}_i) - \pi(\widehat{\boldsymbol{\omega}}_n, \boldsymbol{O}_i)|\right).$$

Moreover, condition C3 gives $|\pi(\boldsymbol{\omega}_0, \boldsymbol{O}_i) - \pi(\widehat{\boldsymbol{\omega}}_n, \boldsymbol{O}_i)| \leq \kappa(\boldsymbol{O}_i)\|\widehat{\boldsymbol{\omega}}_n - \boldsymbol{\omega}_0\|$, with $\mathbb{E}[\kappa(\boldsymbol{O}_i)^2] < \infty$. Hence, we get $\|(a)\| \leq k_1\|\widehat{\boldsymbol{\omega}}_n - \boldsymbol{\omega}_0\|(1/n)\sum_{i=1}^{n}\kappa(\boldsymbol{O}_i) = o_{\mathbb{P}}(1)$, because $\widehat{\boldsymbol{\omega}}_n \xrightarrow{\mathbb{P}} \omega_0$ and the sample mean of $\kappa(\boldsymbol{O}_i)$ is bounded in probability. By condition C4 and the law of large numbers, we reach $\bar{U}_{W,n}(\boldsymbol{\Theta}_0, \boldsymbol{\omega}_0) \xrightarrow{\mathbb{P}} \mathbf{0}$. Combining the two parts, we obtain $\bar{U}_{W,n}(\boldsymbol{\Theta}_0, \widehat{\boldsymbol{\omega}}_n) \xrightarrow{\mathbb{P}} \mathbf{0}$.
- F3 —Nonsingular limit of the Hessian. Let

$$H_{W,n}(\boldsymbol{\Theta}, \boldsymbol{\omega}) = -\frac{1}{n}\left(\frac{\partial U_{W,n}(\boldsymbol{\Theta}, \boldsymbol{\omega})}{\partial \boldsymbol{\Theta}^\top}\right), \quad \boldsymbol{\Theta} \in \mathcal{V}_{\boldsymbol{\Theta}_0}, \boldsymbol{\omega} \in \mathcal{B}.$$

By condition C5, the individual second-order derivatives are dominated by an integrable envelope, so that $\sup_{\Theta\in\mathcal{V}_{\Theta_0}} H_{W,n}(\boldsymbol{\Theta}, \boldsymbol{\omega}_0) - \Lambda(\boldsymbol{\Theta}, \boldsymbol{\omega}_0)\| \xrightarrow{\mathbb{P}} 0$. Moreover, condition C3 (Lipschitz property of $\pi(\boldsymbol{\omega}, \boldsymbol{O})$) implies $\sup_{\Theta\in\mathcal{V}_{\Theta_0}} \left\|H_{W,n}(\boldsymbol{\Theta}, \widehat{\boldsymbol{\omega}}_n) - H_{W,n}(\boldsymbol{\Theta}, \boldsymbol{\omega}_0)\right\| \leq C\|\widehat{\boldsymbol{\omega}}_n - \boldsymbol{\omega}_0\| = o_{\mathbb{P}}(1)$, for some deterministic $C > 0$. Hence, we get $H_{W,n}(\boldsymbol{\Theta}, \widehat{\boldsymbol{\omega}}_n) \xrightarrow{\mathbb{P}} \Lambda(\boldsymbol{\Theta}, \boldsymbol{\omega}_0)$ uniformly on $\mathcal{V}_{\Theta_0}$ and $\Lambda(\boldsymbol{\Theta}_0, \boldsymbol{\omega}_0)$ is positive definite by condition C5. The conditions F1-F3 of the Foutz theorem are therefore satisfied, establishing the consistency $\widehat{\boldsymbol{\Theta}}_{W,n} \xrightarrow{\mathbb{P}} \Theta_0$.

Perform a first-order Taylor expansion of the score around $(\boldsymbol{\Theta}_0, \boldsymbol{\omega}_0)$ such as

$$\mathbf{0} = U_{W,n}(\widehat{\boldsymbol{\Theta}}_{W,n}, \widehat{\boldsymbol{\omega}}_n) = U_{W,n}(\boldsymbol{\Theta}_0, \boldsymbol{\omega}_0) + nH_{W,n}(\widetilde{\boldsymbol{\Theta}}_n, \widehat{\boldsymbol{\omega}}_n)(\widehat{\boldsymbol{\Theta}}_{W,n} - \boldsymbol{\Theta}_0) + \frac{\partial U_{W,n}(\boldsymbol{\Theta}_0, \boldsymbol{\omega}_0)}{\partial \omega^\top}(\widehat{\boldsymbol{\omega}}_n - \boldsymbol{\omega}_0),$$

where $\widetilde{\boldsymbol{\Theta}}_n$ lies on the line segment joining $\boldsymbol{\Theta}_0$ and $\widehat{\boldsymbol{\Theta}}_{W,n}$. Divide by $\sqrt{n}$ and rearrange the expression stated as $\sqrt{n}(\widehat{\boldsymbol{\Theta}}_{W,n} - \boldsymbol{\Theta}_0) = H_{W,n}^{-1}(\widetilde{\boldsymbol{\Theta}}_n, \widehat{\boldsymbol{\omega}}_n)(-n^{-1/2}U_{W,n}(\boldsymbol{\Theta}_0, \boldsymbol{\omega}_0) - \widehat{\boldsymbol{\Gamma}}_n\sqrt{n}(\widehat{\boldsymbol{\omega}}_n - \boldsymbol{\omega}_0))$, with $\widehat{\boldsymbol{\Gamma}}_n = n^{-1/2}\partial U_{W,n}(\boldsymbol{\Theta}_0, \boldsymbol{\omega}_0)/\partial \boldsymbol{\omega}^\top$.

By condition C4 and the multivariate central limit theorem, $n^{-1/2}U_{W,n}(\boldsymbol{\Theta}_0, \boldsymbol{\omega}_0) \xrightarrow{\mathbb{D}} \text{Normal}(\mathbf{0}, \boldsymbol{\Phi})$. Standard ML theory for the selection model gives to us the expression formulated as $\sqrt{n}(\widehat{\boldsymbol{\omega}}_n - \boldsymbol{\omega}_0) \xrightarrow{\mathbb{D}} \text{Normal}(\mathbf{0}, \boldsymbol{\Omega}^{-1})$ and $\widehat{\boldsymbol{\Gamma}}_n \xrightarrow{\mathbb{P}} \boldsymbol{\Gamma}$. Together with the uniform convergence of the Hessian, the Slutsky lemma yields $\sqrt{n}(\widehat{\boldsymbol{\Theta}}_{W,n} - \boldsymbol{\Theta}_0) \xrightarrow{\mathbb{D}} \text{Normal}(\mathbf{0}, \boldsymbol{\Lambda}^{-1}(\boldsymbol{\Phi} - \boldsymbol{\Gamma}\boldsymbol{\Omega}^{-1}\boldsymbol{\Gamma}^\top)(\boldsymbol{\Lambda}^{-1})^\top) \equiv \text{Normal}(\mathbf{0}, \boldsymbol{\Delta}_W)$.

■

4. Simulation study

4.1 *Context*

Next, we assesss the finite-sample behavior of the proposed IPW–MZIBell estimator. The Monte Carlo design is as follows:

- Sample sizes $n \in \{500, 1000\}$;
- Mean zero-inflation rates of 31% and 60%;
- Number of replications $N = 1000$;
- Covariate-missingness rates of 15%, 22%, and 35%.

The data-generating mechanism is given by

$$\begin{aligned} \operatorname{logit}(\psi_i) &= \gamma_1 z_{i1} + \gamma_2 z_{i2} + \gamma_3 z_{i3} + \gamma_4 z_{i4} + \gamma_5 z_{i5}, \\ \log(\nu_i) &= \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \alpha_4 x_{i4}, \end{aligned}$$

with $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{i4})^\top$ and $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{i5})^\top$ being the observed values of $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{i4})^\top$ and $\boldsymbol{Z}_i = (Z_{i1}, \ldots, Z_{i5})^\top$, respectively, where $X_{i1} = Z_{i1} = 1$ and, $X_{i2} \sim \text{Normal}(-1, 0.8^2)$, $X_{i3} \sim \text{Uniform}(0.2, 1)$, $X_{i4} \sim \text{Bernoulli}(0.7)$, $Z_{i2} \sim \text{Normal}(0, 1)$, $Z_{i3} = X_{i3}$, $Z_{i4} \sim \text{Exponential}(1)$, and $Z_{i5} \sim \text{Uniform}(0.6, 1.3)$, which are independent random variables. The regression coefficients are fixed at $\boldsymbol{\alpha} = (-0.5, 1.2, 0.4, 0.9)^\top$, and

$$\boldsymbol{\gamma} = \begin{cases} (0.4, 0.5, -0.8, -1.1, 0.6)^\top & \text{(Scenario 1)}; \\ (0.6, 0.2, -0.7, -0.3, 0.85)^\top & \text{(Scenario 2)}; \end{cases}$$

with Scenario 1 generating approximately 31% zeros and Scenario 2 approximately 60%.

Missing-data indicators are generated under a missing-at-random mechanism stated as

$$\delta_i \sim \text{Bernoulli}\left(\pi_i(\boldsymbol{\omega}, \boldsymbol{O}_i)\right), \quad \pi_i(\boldsymbol{\omega}, \boldsymbol{O}_i) = \operatorname{logit}^{-1}(\boldsymbol{\omega}^\top \boldsymbol{O}_i),$$

with $\boldsymbol{O}_i = (1, y_i, x_{i1}, x_{i3}, z_{i4})^\top$. The vector $\boldsymbol{\omega}$ is calibrated to yield 15%, 22%, or 35% missing covariate values, matching the three simulation settings.

For each combination of (sample size) × (missingness rate) × (zero-inflation level), we generate $N = 1000$ data sets and compute the following:

- $\widehat{\boldsymbol{\Theta}}_{\text{ML}}$ —Full ML (no missing data);
- $\widehat{\boldsymbol{\Theta}}_{W,n}$ —IPW-MZIBell (proposed);
- $\widehat{\boldsymbol{\Theta}}_{\text{CC}}$ —Complete-case (CC) estimator.

All simulations were conducted in the `R` software (version 3.5.2) using the `maxLik` package [16]. For each parameter γ_j, for $j = 1, \ldots, 5$, and α_k, for $k = 1, \ldots, 4$, we report the bias, empirical standard deviation (SD), average model-based standard error (SE), root mean squared error (RMSE), empirical 95% coverage probability (CP), and mean confidence interval length (CIL). The numerical results are shown in Tables 1–4, which compare the following three estimators:

- The "oracle" ML fit using the complete data set;
- The proposed IPW;
- The naive CC estimator.

The ML values serve solely as a benchmark: after fitting the full data, observations are removed at random to generate incomplete data sets on which CC and IPW estimators are computed.

For both CC and IPW, the bias, empirical SD, mean model-based SE, RMSE, and average CIL decrease as n increases and the missing-data rate decreases. The CC estimator exhibits noticeable bias at moderate sample sizes, and its SD, SE, and RMSE are typically larger than those of IPW. Coverage for CC often falls well below the nominal 95% level when 35% of covariate values are missing (Tables 1 and 3).

Table 1: Simulation results for sample size $n = 500$, zero-inflation rate = 31%, and missing-data rates of 15% (top), 22% (middle), and 35% (bottom).

		$\widehat{\boldsymbol{\gamma}}_n$					$\widehat{\boldsymbol{\alpha}}_n$			
		$\widehat{\gamma}_{1,n}$	$\widehat{\gamma}_{2,n}$	$\widehat{\gamma}_{3,n}$	$\widehat{\gamma}_{4,n}$	$\widehat{\gamma}_{5,n}$	$\widehat{\alpha}_{1,n}$	$\widehat{\alpha}_{2,n}$	$\widehat{\alpha}_{3,n}$	$\widehat{\alpha}_{4,n}$
ML	Bias	0.111	0.040	−0.705	−0.647	0.125	0.221	−0.463	0.103	0.125
	SD	0.510	0.210	0.222	0.265	0.500	0.134	0.046	0.090	0.134
	SE	0.525	0.222	0.272	0.290	0.531	0.142	0.045	0.075	0.112
	RMSE	0.526	0.298	0.273	0.242	0.496	0.125	0.049	0.075	0.118
	CP	0.957	0.946	0.945	0.962	0.959	0.947	0.945	0.942	0.946
	CIL	2.024	0.438	1.021	0.925	1.976	0.460	0.175	0.291	0.437
					15%					
CC	Bias	−1.348	0.748	−3.184	−2.718	0.483	0.615	−1.064	−0.102	0.543
	SD	0.645	0.260	0.404	0.465	0.645	0.128	0.064	0.090	0.133
	SE	0.693	0.326	0.384	0.459	0.673	0.119	0.045	0.073	0.114
	RMSE	0.812	0.157	0.440	0.460	0.659	0.266	0.090	0.092	0.116
	CP	0.938	0.940	0.952	0.938	0.947	0.788	0.679	0.892	0.927
	CIL	2.695	0.570	1.461	1.369	2.623	0.466	0.176	0.284	0.446
IPW	Bias	−0.126	0.026	−0.960	−0.459	−0.152	−0.102	0.603	−0.205	0.015
	SD	0.597	0.142	0.394	0.328	0.594	0.131	0.050	0.084	0.122
	SE	0.524	0.114	0.282	0.445	0.513	0.118	0.055	0.077	0.113
	RMSE	0.595	0.142	0.407	0.533	0.195	0.131	0.050	0.084	0.128
	CP	0.920	0.894	0.842	0.862	0.913	0.908	0.932	0.916	0.934
	CIL	2.046	0.445	1.079	0.934	1.999	0.461	0.175	0.291	0.438
					22%					
CC	Bias	−0.760	−0.087	−1.274	−1.348	0.079	0.134	−0.617	−0.122	0.300
	SD	0.657	0.149	0.376	0.315	0.654	0.133	0.055	0.103	0.141
	SE	0.669	0.146	0.345	0.296	0.648	0.136	0.057	0.094	0.131
	RMSE	0.658	0.150	0.377	0.321	0.656	0.133	0.059	0.104	0.140
	CP	0.931	0.914	0.924	0.902	0.932	0.917	0.894	0.875	0.894
	CIL	2.610	0.532	1.332	1.139	2.534	0.528	0.211	0.363	0.513
IPW	Bias	0.446	−0.183	−1.031	−0.362	−0.030	−0.105	0.498	−0.214	−0.002
	SD	0.621	0.140	0.357	0.300	0.614	0.137	0.053	0.096	0.130
	SE	0.606	0.144	0.315	0.270	0.586	0.135	0.053	0.078	0.113
	RMSE	0.621	0.142	0.357	0.303	0.616	0.148	0.054	0.098	0.130
	CP	0.962	0.921	0.942	0.931	0.973	0.937	0.953	0.947	0.964
	CIL	2.327	0.472	1.194	1.033	2.267	0.457	0.168	0.304	0.440
					35%					
CC	Bias	1.230	0.029	−1.727	−0.674	−0.189	−0.123	0.599	−0.224	0.005
	SD	0.801	0.172	0.398	0.374	0.798	0.195	0.074	0.135	0.190
	SE	0.696	0.155	0.361	0.324	0.683	0.165	0.065	0.124	0.155
	RMSE	0.798	0.176	0.398	0.382	0.797	0.196	0.074	0.137	0.189
	CP	0.826	0.834	0.738	0.838	0.512	0.794	0.384	0.760	0.782
	CIL	2.714	0.603	1.396	1.237	2.660	0.644	0.254	0.479	0.607
IPW	Bias	−0.177	0.206	−1.313	−0.438	−0.027	0.119	−1.029	−0.019	0.265
	SD	0.719	0.161	0.359	0.335	0.708	0.172	0.065	0.126	0.157
	SE	0.562	0.123	0.288	0.263	0.543	0.180	0.071	0.085	0.116
	RMSE	0.750	0.163	0.368	0.335	0.708	0.201	0.072	0.118	0.159
	CP	0.921	0.938	0.944	0.919	0.926	0.882	0.912	0.968	0.941
	CIL	2.150	0.475	1.099	0.997	2.110	0.477	0.185	0.327	0.419

Table 2: Simulation results for sample size $n = 1000$, zero-inflation rate = 31%, and missing-data rates of 15% (top), 22% (middle), and 35% (bottom).

		$\widehat{\boldsymbol{\gamma}}_n$					$\widehat{\boldsymbol{\alpha}}_n$			
		$\widehat{\gamma}_{1,n}$	$\widehat{\gamma}_{2,n}$	$\widehat{\gamma}_{3,n}$	$\widehat{\gamma}_{4,n}$	$\widehat{\gamma}_{5,n}$	$\widehat{\alpha}_{1,n}$	$\widehat{\alpha}_{2,n}$	$\widehat{\alpha}_{3,n}$	$\widehat{\alpha}_{4,n}$
ML	Bias	−0.024	0.012	−0.497	−0.785	0.713	0.190	0.010	0.036	0.108
	SD	0.290	0.053	0.153	0.138	0.292	0.058	0.027	0.038	0.057
	SE	0.309	0.065	0.156	0.137	0.305	0.062	0.023	0.039	0.060
	RMSE	0.283	0.057	0.153	0.138	0.292	0.057	0.022	0.038	0.056
	CP	0.965	0.972	0.956	0.941	0.958	0.956	0.949	0.952	0.964
	CIL	1.210	0.240	0.607	0.522	1.109	0.240	0.090	0.159	0.239
					15%					
CC	Bias	−1.026	−0.210	−1.606	−1.755	0.086	0.284	−0.789	−0.126	0.306
	SD	0.325	0.067	0.179	0.163	0.316	0.058	0.035	0.037	0.059
	SE	0.356	0.070	0.176	0.159	0.344	0.069	0.034	0.039	0.062
	RMSE	0.392	0.071	0.175	0.171	0.327	0.103	0.025	0.048	0.062
	CP	0.943	0.956	0.953	0.939	0.958	0.912	0.926	0.918	0.972
	CIL	1.403	0.274	0.688	0.620	1.347	0.251	0.094	0.157	0.251
IPW	Bias	0.502	−0.070	−1.129	−0.287	−0.097	−0.099	0.525	−0.186	0.036
	SD	0.319	0.068	0.168	0.158	0.309	0.058	0.032	0.039	0.059
	SE	0.311	0.063	0.154	0.137	0.301	0.062	0.057	0.034	0.061
	RMSE	0.312	0.063	0.171	0.158	0.309	0.059	0.024	0.039	0.059
	CP	0.955	0.946	0.938	0.910	0.950	0.941	0.939	0.959	0.963
	CIL	1.215	0.239	0.606	0.544	1.181	0.244	0.094	0.152	0.240
					22%					
CC	Bias	−0.685	−0.084	−1.223	−1.584	−0.014	0.094	−0.823	−0.131	0.281
	SD	0.356	0.073	0.172	0.158	0.349	0.078	0.030	0.058	0.082
	SE	0.353	0.072	0.178	0.153	0.342	0.074	0.028	0.052	0.071
	RMSE	0.356	0.073	0.172	0.158	0.349	0.078	0.030	0.058	0.082
	CP	0.910	0.900	0.930	0.914	0.914	0.872	0.870	0.826	0.848
	CIL	1.383	0.282	0.694	0.595	1.340	0.290	0.109	0.203	0.279
IPW	Bias	0.409	−0.195	−0.595	−0.311	−0.111	−0.098	0.501	−0.210	−0.001
	SD	0.338	0.070	0.165	0.148	0.328	0.069	0.026	0.049	0.073
	SE	0.311	0.066	0.153	0.138	0.308	0.066	0.023	0.040	0.076
	RMSE	0.356	0.072	0.182	0.149	0.328	0.071	0.235	0.059	0.078
	CP	0.952	0.939	0.948	0.953	0.939	0.947	0.889	0.937	0.939
	CIL	1.213	0.246	0.604	0.533	1.185	0.247	0.089	0.158	0.256
					35%					
CC	Bias	−0.541	−0.118	−1.139	−1.582	−0.033	−0.067	−0.813	−0.166	0.297
	SD	0.477	0.083	0.220	0.186	0.463	0.111	0.045	0.093	0.120
	SE	0.381	0.079	0.198	0.167	0.373	0.091	0.034	0.079	0.089
	RMSE	0.455	0.079	0.261	0.163	0.361	0.280	0.033	0.096	0.088
	CP	0.812	0.880	0.842	0.878	0.816	0.748	0.718	0.594	0.698
	CIL	1.493	0.309	0.775	0.652	1.458	0.357	0.134	0.274	0.349
IPW	Bias	−0.406	−0.102	−1.087	−0.317	−0.111	−0.093	0.601	−0.239	−0.013
	SD	0.382	0.080	0.198	0.168	0.374	0.096	0.031	0.073	0.090
	SE	0.314	0.067	0.160	0.141	0.304	0.093	0.034	0.076	0.061
	RMSE	0.455	0.079	0.261	0.163	0.361	0.098	0.033	0.076	0.084
	CP	0.896	0.948	0.859	0.946	0.964	0.899	0.930	0.899	0.926
	CIL	1.330	0.249	0.625	0.542	1.193	0.248	0.095	0.172	0.238

Table 3: Simulation results for sample size $n = 500$ and zero-inflation rate of 60% under missing-data rates of 15% (top), 22% (middle), and 35% (bottom).

		$\widehat{\boldsymbol{\gamma}}_n$					$\widehat{\boldsymbol{\alpha}}_n$			
		$\widehat{\gamma}_{1,n}$	$\widehat{\gamma}_{2,n}$	$\widehat{\gamma}_{3,n}$	$\widehat{\gamma}_{4,n}$	$\widehat{\gamma}_{5,n}$	$\widehat{\alpha}_{1,n}$	$\widehat{\alpha}_{2,n}$	$\widehat{\alpha}_{3,n}$	$\widehat{\alpha}_{4,n}$
ML	Bias	−0.012	0.878	−0.614	−0.129	0.055	0.085	−0.197	0.219	0.147
	SD	0.385	0.088	0.207	0.121	0.387	0.130	0.050	0.083	0.103
	SE	0.312	0.092	0.204	0.116	0.362	0.135	0.057	0.095	0.117
	RMSE	0.398	0.092	0.215	0.116	0.387	0.139	0.057	0.094	0.118
	CP	0.945	0.967	0.972	0.959	0.937	0.941	0.955	0.961	0.954
	CIL	1.393	0.383	0.806	0.458	1.419	0.548	0.220	0.331	0.456
					15%					
CC	Bias	−0.141	−0.579	−1.422	−0.940	0.096	0.379	−0.518	−0.403	0.299
	SD	0.430	0.105	0.250	0.137	0.413	0.159	0.067	0.104	0.119
	SE	0.425	0.110	0.239	0.132	0.404	0.150	0.058	0.099	0.118
	RMSE	0.440	0.106	0.243	0.138	0.428	0.196	0.059	0.103	0.120
	CP	0.916	0.964	0.944	0.965	0.947	0.894	0.959	0.947	0.953
	CIL	1.655	0.394	0.921	0.508	1.560	0.535	0.250	0.362	0.474
IPW	Bias	0.101	−0.395	−1.019	0.285	0.102	−0.126	0.373	−0.207	0.058
	SD	0.420	0.098	0.237	0.121	0.383	0.149	0.061	0.100	0.101
	SE	0.385	0.095	0.216	0.119	0.383	0.145	0.057	0.096	0.115
	RMSE	0.405	0.094	0.239	0.123	0.397	0.141	0.059	0.100	0.121
	CP	0.929	0.957	0.921	0.945	0.929	0.935	0.952	0.949	0.949
	CIL	1.502	0.410	0.842	0.453	1.520	0.529	0.227	0.349	0.460
					22%					
CC	Bias	0.400	−0.601	−1.375	−0.285	−0.095	0.020	−0.310	−0.413	0.287
	SD	0.487	0.118	0.277	0.139	0.466	0.175	0.079	0.137	0.167
	SE	0.447	0.120	0.250	0.129	0.425	0.166	0.075	0.125	0.142
	RMSE	0.499	0.119	0.268	0.139	0.433	0.248	0.075	0.130	0.142
	CP	0.870	0.897	0.857	0.911	0.863	0.871	0.856	0.867	0.855
	CIL	1.759	0.438	0.974	0.505	1.669	0.658	0.282	0.486	0.558
IPW	Bias	0.328	−0.507	−1.074	0.200	0.100	−0.070	0.211	−0.218	−0.011
	SD	0.459	0.114	0.256	0.128	0.433	0.170	0.071	0.128	0.146
	SE	0.390	0.097	0.218	0.114	0.411	0.158	0.068	0.119	0.130
	RMSE	0.483	0.115	0.260	0.118	0.423	0.174	0.072	0.119	0.142
	CP	0.927	0.951	0.942	0.930	0.934	0.902	0.949	0.939	0.934
	CIL	1.604	0.381	0.880	0.452	1.459	0.535	0.224	0.385	0.463
					35%					
CC	Bias	0.701	−0.671	−1.262	−0.869	0.023	−0.399	−0.479	−0.302	0.280
	SD	0.587	0.141	0.322	0.166	0.581	0.236	0.104	0.203	0.203
	SE	0.568	0.139	0.318	0.161	0.555	0.236	0.106	0.210	0.212
	RMSE	0.772	0.158	0.350	0.169	0.580	0.240	0.106	0.203	0.204
	CP	0.841	0.829	0.828	0.838	0.840	0.729	0.686	0.649	0.689
	CIL	2.219	0.545	1.248	0.626	2.170	0.931	0.417	0.506	0.822
IPW	Bias	1.042	−0.510	−1.151	0.214	0.010	−0.161	0.217	−0.265	−0.020
	SD	0.564	0.132	0.310	0.153	0.563	0.218	0.098	0.196	0.198
	SE	0.530	0.101	0.239	0.127	0.419	0.172	0.094	0.190	0.194
	RMSE	0.606	0.143	0.334	0.153	0.533	0.267	0.101	0.201	0.196
	CP	0.859	0.910	0.934	0.904	0.952	0.895	0.940	0.962	0.916
	CIL	1.679	0.393	0.929	0.499	1.625	0.572	0.248	0.462	0.510

Table 4: Simulation results for sample size $n = 1000$, zero-inflation rate = 60%, and missing-data rates of 15% (top), 22% (middle), and 35% (bottom).

		$\widehat{\boldsymbol{\gamma}}_n$					$\widehat{\boldsymbol{\alpha}}_n$			
		$\widehat{\gamma}_{1,n}$	$\widehat{\gamma}_{2,n}$	$\widehat{\gamma}_{3,n}$	$\widehat{\gamma}_{4,n}$	$\widehat{\gamma}_{5,n}$	$\widehat{\alpha}_{1,n}$	$\widehat{\alpha}_{2,n}$	$\widehat{\alpha}_{3,n}$	$\widehat{\alpha}_{4,n}$
ML	Bias	−0.005	0.100	−0.401	−0.092	0.411	0.922	−0.514	0.238	1.321
	SD	0.215	0.052	0.112	0.065	0.194	0.080	0.034	0.050	0.062
	SE	0.210	0.052	0.109	0.062	0.200	0.072	0.031	0.051	0.062
	RMSE	0.214	0.052	0.115	0.064	0.201	0.073	0.030	0.045	0.064
	CP	0.943	0.952	0.959	0.948	0.938	0.956	0.945	0.952	0.956
	CIL	0.810	0.201	0.437	0.243	0.757	0.284	0.122	0.191	0.246
					15%					
CC	Bias	−0.063	−0.642	−1.523	−0.918	0.090	0.330	−0.500	−0.345	0.284
	SD	0.236	0.054	0.120	0.071	0.218	0.073	0.031	0.048	0.067
	SE	0.228	0.054	0.125	0.069	0.215	0.073	0.030	0.049	0.064
	RMSE	0.331	0.058	0.120	0.077	0.233	0.142	0.031	0.061	0.067
	CP	0.814	0.934	0.966	0.930	0.928	0.620	0.944	0.904	0.942
	CIL	0.893	0.213	0.488	0.271	0.843	0.288	0.119	0.191	0.251
IPW	Bias	−0.292	−0.500	−0.904	0.233	0.099	−0.222	0.312	−0.318	0.019
	SD	0.221	0.053	0.118	0.068	0.207	0.073	0.032	0.050	0.067
	SE	0.206	0.051	0.114	0.062	0.196	0.072	0.031	0.049	0.063
	RMSE	0.221	0.053	0.118	0.068	0.207	0.073	0.032	0.050	0.067
	CP	0.938	0.942	0.946	0.916	0.936	0.948	0.938	0.946	0.930
	CIL	0.807	0.200	0.447	0.242	0.766	0.284	0.120	0.192	0.246
					22%					
CC	Bias	0.368	−0.588	−1.366	−0.866	−0.022	0.031	−0.520	−0.297	0.279
	SD	0.253	0.062	0.142	0.072	0.239	0.094	0.041	0.072	0.080
	SE	0.238	0.060	0.133	0.068	0.225	0.088	0.038	0.065	0.075
	RMSE	0.286	0.063	0.139	0.076	0.214	0.191	0.043	0.066	0.075
	CP	0.889	0.890	0.900	0.910	0.904	0.876	0.870	0.818	0.870
	CIL	1.296	0.258	0.600	0.298	0.904	0.394	0.261	0.273	0.304
IPW	Bias	0.298	−0.610	−0.698	0.210	0.118	−0.095	0.604	−0.213	−0.010
	SD	0.235	0.067	0.139	0.068	0.215	0.089	0.038	0.065	0.070
	SE	0.209	0.054	0.116	0.063	0.198	0.075	0.032	0.055	0.065
	RMSE	0.252	0.067	0.138	0.079	0.210	0.097	0.048	0.070	0.076
	CP	0.901	0.929	0.970	0.916	0.940	0.901	0.926	0.938	0.922
	CIL	0.815	0.211	0.460	0.242	0.768	0.282	0.121	0.194	0.243
					35%					
CC	Bias	0.724	−0.679	−1.274	−0.839	−0.037	−0.382	−0.467	−0.294	0.284
	SD	0.378	0.094	0.222	0.106	0.373	0.168	0.083	0.155	0.154
	SE	0.294	0.074	0.166	0.083	0.286	0.124	0.057	0.106	0.109
	RMSE	0.589	0.108	0.215	0.103	0.262	0.596	0.065	0.107	0.108
	CP	0.590	0.826	0.870	0.848	0.970	0.600	0.916	0.925	0.954
	CIL	1.152	0.293	0.690	0.323	1.129	0.485	0.222	0.452	0.426
IPW	Bias	0.314	−0.514	−0.670	0.225	0.049	−0.110	0.622	−0.247	−0.010
	SD	0.268	0.073	0.174	0.083	0.260	0.126	0.056	0.107	0.106
	SE	0.214	0.053	0.121	0.064	0.206	0.075	0.033	0.058	0.066
	RMSE	0.378	0.095	0.224	0.109	0.376	0.168	0.086	0.162	0.154
	CP	0.858	0.868	0.944	0.932	0.912	0.846	0.902	0.914	0.880
	CIL	0.839	0.207	0.474	0.249	0.805	0.295	0.129	0.224	0.257

Across all scenarios, the bias remains small for IPW, and the empirical coverage probabilities stay close to 95%, even with $n = 500$. For a fixed missingness rate, estimation of the α_k coefficients improves as the proportion of structural zeros decreases, whereas estimation of the γ_j coefficients benefits from higher zero-inflation, in line with model intuition.

Overall, the results confirm that IPW provides reliable inference for MZIBell regression with covariates which are missing at random, markedly outperforming the naive CC approach while approaching the oracle ML benchmark as information increases.

4.2 *Diagnostic check of asymptotic normality.*

To assess the quality of the Gaussian approximation stated in Theorem 1, we plot histograms of the standardized estimators $(\widehat{\alpha}_{k,n} - \alpha_k)/\mathrm{SE}(\widehat{\alpha}_{k,n})$, for $k = 1, \ldots, 4$, and $(\widehat{\gamma}_{j,n} - \gamma_j)/\mathrm{SE}(\widehat{\gamma}_{j,n})$, for $j = 1, \ldots, 5$, under the setting $n = 1000$, zero-inflation 31%, and missingness 15%; see Figures 1 and 2. The empirical distributions closely follow the standard normal distribution and similar behavior is observed in the remaining scenarios and is therefore omitted.

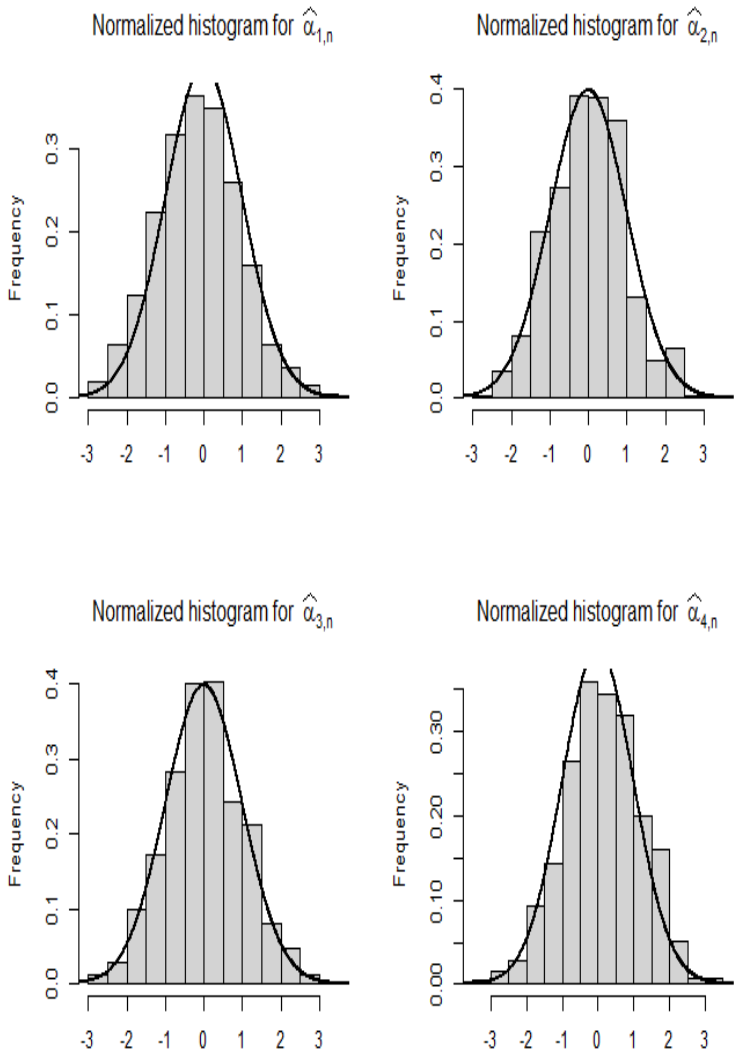

Figure 1: Histograms of the standardized estimates $(\widehat{\alpha}_{k,n} - \alpha_k)/\mathrm{SE}(\widehat{\alpha}_{k,n})$, for $k = 1, \ldots, 4$, with $n = 1000$, zero-inflation rate of 60%, and 15% missing data.

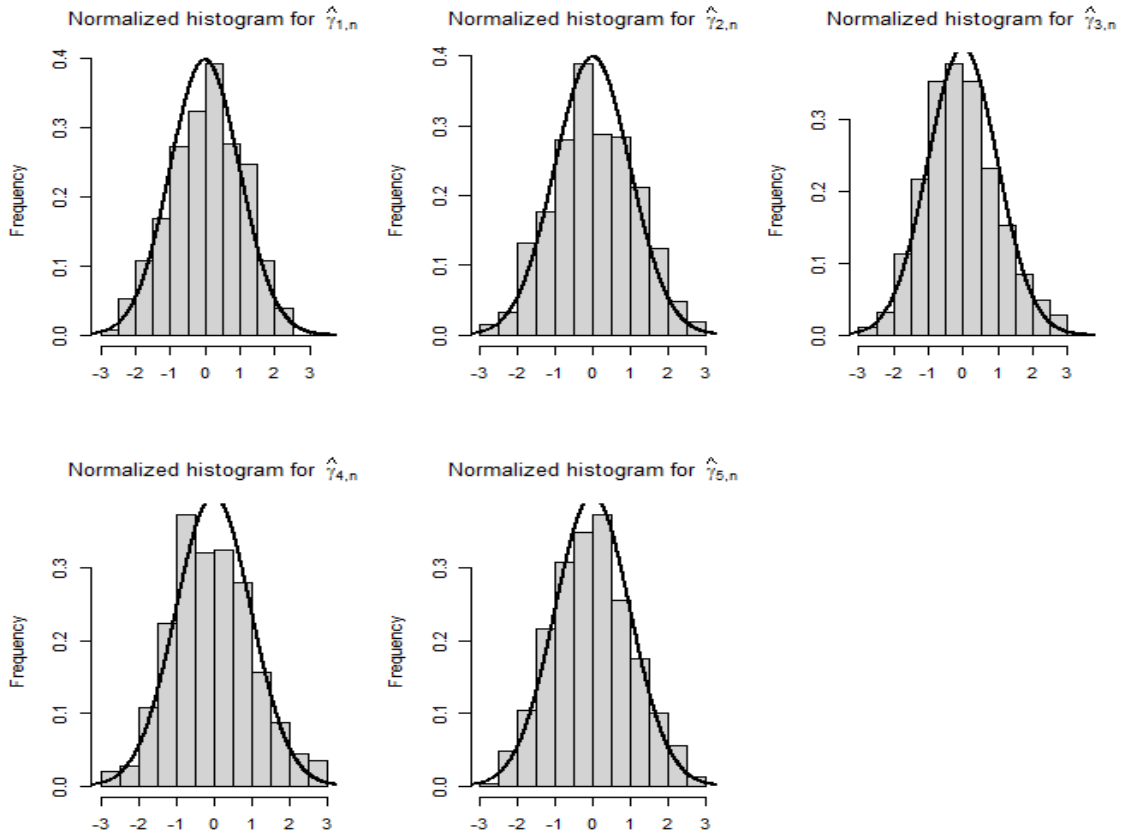

Figure 2: Histograms of the standardized estimates $(\widehat{\gamma}_{j,n} - \gamma_j)/\mathrm{SE}(\widehat{\gamma}_{j,n})$ for $j = 1, \ldots, 5$, with $n = 1000$, zero-inflation rate of 60%, and 15% missing data.

5. Application with real data

5.1 *Context*

The RAND Health Insurance Experiment is one of the most influential policy studies ever conducted in the United States (1974–1982). Roughly 8,000 individuals in 2,823 families, recruited at six sites, were randomly assigned to one of 14 health insurance plans and followed for up to five years. The public-use microfile contains 20,190 person-year records with information on medical utilization, expenditures, demographics, and health status.

The response variable analyzed below is the number of doctor visits (`mdvis`) during the previous two weeks. Structural zeros are common: 6,308/20,190 $\approx$ 0.31, confirming substantial zero-inflation. Explanatory variables include age (in years), gender (1 = female), education (`educdec`, in years of schooling), household income (in thousands of US dollars), and a general health index (`ghindx`; higher values indicate poorer health). Descriptive statistics for selected variables are presented in Table 5 from which we can verify the statistical distribution of each variable.

Table 5: Descriptive statistics for selected variables with RAND Health Insurance Experiment data and $n = 20{,}190$.

Variable	Minimum	First quartile	Median	Third quartile	Maximum	Mean	SD
`mdvis`	0	0	1	4	77	2.860	4.500
`age`	0	11.460	24.200	37.420	64.280	25.720	16.770
`income`	0	5,334.000	7,976.000	10,475.000	29,238.000	8,037.000	4,058.000
`educdec`	0	11	12	13	25	11.960	2.810
`ghindx`	0	0	67	81.500	100	54.180	34.840

5.2 *Missingness and model fit*

Both `ghindx` and `educdec` contain missing values. The nonresponse rate is 25.87% for `ghindx` and less than 0.1% for `educdec`, so that 25.89% of subjects exhibit at least one missing covariate. We estimate the MZIBell regression by (i) CC analysis, and (ii) the proposed parametric IPW approach. Results are summarized in Table 6, which reports the IPW estimates, their SEs and two-sided p-values.

Table 6: Estimates, SE, and p-values from the MZIBell regression analysis for the RAND Health Insurance Experiment data.

		IPW estimator			CC estimator		
Parameter	Variable	Estimate	SE	p-value	Estimate	SE	p-value
Marginal mean submodel (α_i)	`intercept`	0.973	0.053	< 0.001	1.064	0.044	< 0.001
	`age`	0.007	0.001	< 0.001	0.008	0.001	< 0.001
	`gender`	0.290	0.017	< 0.001	0.264	0.014	< 0.001
	`income`	2.590×10^{-5}	2.170×10^{-6}	< 0.001	3.320×10^{-5}	1.690×10^{-6}	< 0.001
	`educdec`	0.020	0.003	< 0.001	0.020	0.003	< 0.001
	`ghindx`	-0.011	0.001	< 0.001	-0.012	0.001	< 0.001
Zero−inflation submodel (γ_i)	`intercept`	0.688	0.121	< 0.001	0.573	0.097	< 0.001
	`age`	-0.008	0.002	< 0.001	-0.009	0.002	< 0.001
	`gender`	-0.487	0.056	< 0.001	-0.473	0.045	< 0.001
	`income`	-9.420×10^{-5}	7.940×10^{-6}	< 0.001	-8.250×10^{-5}	6.320×10^{-6}	< 0.001
	`educdec`	-0.071	0.010	< 0.001	-0.061	0.008	< 0.001

For comparison, we also fitted marginalized zero-inflated negative binomial (MZINB) and Poisson (MZIP) regression models to the RAND data. Because `ghindx` and `educdec` contain missing values, the same IPW strategy used for MZIBell was applied to MZIP and MZINB. Tables 7, 8, and 9 report the IPW estimates, their SE, and two-sided p-values.

Table 7: Estimates, SE, and p-values from the MZIBell model on RAND data.

Parameter	Variable	Estimate	SE	p-value
Marginal mean submodel (α_i)	`intercept`	1.064	0.044	< 0.001
	`age`	0.008	0.001	< 0.001
	`gender`	0.264	0.014	< 0.001
	`income`	0.000	0.000	< 0.001
	`educdec`	0.020	0.003	< 0.001
	`ghindx`	-0.012	0.001	< 0.001
Zero-inflation submodel (γ_i)	`intercept`	0.573	0.097	< 0.001
	`age`	-0.009	0.002	< 0.001
	`gender`	-0.473	0.045	< 0.001
	`income`	-0.000	0.000	< 0.001
	`educdec`	-0.061	0.008	< 0.001
	AIC	-61883.160		
	BIC	-61796.110		

Table 8: Estimates, SE, and p-values from the MZIP model on RAND data.

Parameter	Variable	Estimate	SE	p-value
Marginal mean submodel (α_i)	`intercept`	1.013	0.033	< 0.001
	`age`	0.010	0.001	< 0.001
	`gender`	0.026	0.012	< 0.001
	`income`	0.000	0.000	< 0.001
	`educdec`	0.014	0.002	< 0.001
	`ghindx`	-0.011	0.000	< 0.001
Zero-inflation submodel (γ_i)	`intercept`	0.681	0.070	< 0.001
	`age`	-0.015	0.001	< 0.001
	`gender`	-0.437	0.032	< 0.001
	`income`	-0.000	0.000	< 0.001
	`educdec`	-0.035	0.006	< 0.001
	AIC	-35709.850		
	BIC	-35622.810		

Table 9: Estimates, SE, and p-values from the MZINB model on RAND data.

Parameter	Variable	Estimate	SE	p-value
Marginal mean submodel (α_i)	`intercept`	1.572	0.248	< 0.001
	`age`	-0.023	0.005	< 0.001
	`gender`	-0.820	0.142	< 0.001
	`income`	-0.000	0.000	< 0.001
	`educdec`	-0.145	0.021	< 0.001
	`ghindx`	0.987	0.070	< 0.001
Zero-inflation submodel (γ_i)	`intercept`	0.006	0.001	< 0.001
	`age`	0.261	0.021	< 0.001
	`gender`	0.000	0.000	< 0.001
	`income`	0.023	0.004	< 0.001
	`educdec`	-0.011	0.001	< 0.001
	Dispersion parameter	0.442	0.007	< 0.001
	AIC	-60048.590		
	BIC	-59953.640		

From Tables 7, 8, and 9, the three models yield broadly similar point estimates, yet the IPW–MZIBell achieves the smallest Akaike (AIC) and Bayesian (BIC) information criteria. Recall that $\text{AIC} = -2\ell_{\max} + 2k$ and $\text{BIC} = -2\ell_{\max} + k\log(n)$, where the sign of the maximized log-likelihood $\ell_{\max}$ is immaterial, so lower AIC and BIC —whether positive or negative— indicate a better trade-off between goodness of fit and model complexity. Hence, the MZIBell model provides the best overall compromise for these data.

Point estimates are almost identical across methods —as expected with a large sample and moderate missingness— but IPW yields noticeably smaller standard errors, indicating efficiency gains over CC.

Age, gender, income, and education enter the logit part of the model with negative and statistically significant coefficients: older individuals, women, and respondents with higher income or more years of schooling are less likely to belong to the always-zero group, that is, they are more likely to have at least one medical contact. In the marginal mean equation, age, gender, income, and education all present positive coefficients, so the expected number of visits increases for older, female, wealthier, or better-educated subjects. Conversely, the coefficient for `ghindx` is negative (-0.012), indicating that respondents who report worse health tend to have fewer visits during these two weeks —a pattern that may reflect barriers to access or postponement of care among those in poorer condition.

Overall, the IPW–MZIBell procedure provides a coherent and efficient analysis that accommodates both excess zeros and covariate missingness, offering a practical alternative to ad hoc CC analyses in large observational studies.

6. Conclusions

We have extended the MZIBell regression model to the realistic setting in which some covariates are missing at random. Estimation is performed via parametric IPW; under mild regularity conditions, the IPW estimator is consistent and asymptotically normal. Simulation studies confirmed the finite-sample validity of the asymptotic approximations, even with moderate sample sizes and up to 35% covariate nonresponse. When applied to the RAND Health Insurance Experiment, the IPW–MZIBell produced tighter standard errors than CC analysis and outperformed marginalized Poisson and negative binomial competitors in both AIC and BIC.

Some limitations and future work are as follows. The present IPW formulation assumes a monotone missing-data pattern—that is, if a subject is missing one covariate, all covariates to its right in a fixed ordering are also missing. In many surveys, the pattern is nonmonotone, making the missingness mechanism harder to model and standard IPW less effective. Developing doubly robust or semiparametric alternatives that remain valid under arbitrary MAR patterns is an important topic for future research.

Another promising extension is a multivariate MZIBell framework capable of analyzing several zero-inflated, overdispersed counts jointly—for instance, simultaneous modeling of doctor visits and prescription drug use. Such a model would capture both marginal effects and cross-outcome correlations, offering a more comprehensive understanding of complex healthcare behavior. These directions, together with computational refinements for very large datasets, will further enhance the practical utility of the MZIBell family.

Appendix A: Theoretical aspects

The Lambert function $W(x)$ is defined as the inverse of $f(w) = w\exp(w)$, that is, $W(x)\exp(W(x)) = x$. We work with the principal branch W: $(-\exp(-1), \infty) \to (-1, \infty)$.

Implicit differentiation of $W(x)\exp(W(x)) = x$ gives $W'(x)\exp(W(x))(1+W(x)) = 1$. Since $\exp(W(x)) = x/W(x)$ for $x \neq 0$, we obtain

$$W'(x) = \frac{W(x)}{x(1+W(x))}, \quad x \neq -\exp(-1).$$

From $W'(x) = W(x)/(x(1+W(x)))$, set $u(x) = x(1+W(x))$. Then, we have that $W'(x) = W(x)/u(x)$ and, by the quotient rule,

$$W''(x) = \frac{u(x)W'(x) - W(x)u'(x)}{u(x)^2}.$$

Compute $u'(x) = 1 + W(x) + xW'(x)$. Using the expression for $W'(x)$,

$$u'(x) = 1 + W(x) + \frac{W(x)}{1+W(x)} = \frac{(1+W(x))^2 + W(x)}{1+W(x)} = \frac{W(x)(2+W(x))}{1+W(x)}.$$

Substituting into the expression for $W''(x)$ and simplifying,

$$W''(x) = -\frac{W(x)(2+W(x))}{x^2(1+W(x))^3}, \quad x \neq -\exp(-1).$$

Hence, W is of class $\mathcal{C}^2$ —indeed, $\mathcal{C}^\infty$— on every open interval contained in $(-\exp(-1), \infty)$. Higher-order derivatives can be obtained recursively from the implicit definition.

Appendix B: Computational codes in R

B1. R code for the simulation study

```
##################################################
# SIMULATION OF THE ZIBell MODEL #
##################################################

# Required libraries
library(maxLik)
library(bellreg)
library(LambertW)
library(lamW)

# Simulation parameters
n <- 1000
b <- c(-0.5, 1.2, 0.4, 0.9)
g <- c(0.6, 0.2, -0.7, -0.3, 0.85)
phi <- c(g, b)
nbrep <- 1000
```

```r

# Storage for results
estimatesFD <- matrix(0, length(phi), nbrep)
est.covFD <- estimatesFD
estimatesCC <- estimatesFD
est.covCC <- estimatesFD
estimatesIPW <- estimatesFD
est.covIPW <- estimatesFD
missing <- rep(0, nbrep)
zeroinflp <- rep(0, nbrep)

# Monte Carlo Simulation
for (ind in 1:nbrep) {
 cat("Simulation:", ind, "\n")

 # Covariates generation
 inter <- rep(1, n)
 X1 <- rnorm(n, -1, 0.8)
 X2 <- runif(n, 0.2, 1)
 X3 <- rbinom(n, 1, 0.7)
 Z2 <- X2
 Z3 <- rexp(n)
 Z4 <- runif(n, 0.6, 1.3)

 X <- rbind(inter, X1, X2, X3)
 Z <- rbind(inter, X1, Z2, Z3, Z4)

 pi <- exp(t(g) %*% Z) / (1 + exp(t(g) %*% Z))
 mu <- (1 + exp(t(g) %*% Z)) * exp(t(b) %*% X)
 S <- rbinom(n, 1, pi)
 zeroinflp[ind] <- mean(S)

 Y <- ifelse(S == 1, 0, rbell(n, W(mu)))
 J <- as.integer(Y == 0)

 # Full Data Log-likelihood
 loglikfun <- function(param) {
 g <- param[1:length(g)]
 b <- param[(length(g) + 1):length(phi)]
 lpz <- t(g) %*% Z
 lpx <- t(b) %*% X
 wmu <- (1 + exp(lpz)) * exp(lpx)
 wval <- W(wmu)
 sum(J * log(exp(lpz) + exp(1 - exp(wval))) - log(1 + exp(lpz)) +
 (1 - J) * (1 - exp(wval) + Y * log(wval)))
 }

 mle <- maxLik(logLik = loglikfun, start = rep(0, length(phi)))
 estimatesFD[, ind] <- mle$est
 est.covFD[, ind] <- diag(vcov(mle))

 # Missing data mechanism
 So <- rbind(rep(1, n), Y, X1, X3, Z4)
 alpha <- c(-0.4, 0.1, -0.5, 0.7, 1.2)
 pr <- exp(t(alpha) %*% So) / (1 + exp(t(alpha) %*% So))
 delta <- rbinom(n, 1, pr)
 missing[ind] <- (1 - mean(delta)) * 100
 modele <- glm(delta ~ Y + X1 + X3 + Z4, family = binomial)
 pr.est <- predict(modele, type = "response")

 # Complete Case Estimation
 loglikfCC <- function(param) {
 g <- param[1:length(g)]
 b <- param[(length(g) + 1):length(phi)]
 lpz <- t(g) %*% Z
 lpx <- t(b) %*% X
 wmu <- (1 + exp(lpz)) * exp(lpx)
 wval <- W(wmu)
 sum(delta * (J * log(exp(lpz) + exp(1 - exp(wval))) - log(1 + exp(lpz)) +
 (1 - J) * (1 - exp(wval) + Y * log(wval))))
 }

 mleCC <- maxLik(logLik = loglikfCC, start = rep(0, length(phi)))
 estimatesCC[, ind] <- mleCC$est
```

```r
 est.covCC[, ind] <- diag(vcov(mleCC))

 # IPW Estimation
 loglikfunIPW <- function(param) {
 g <- param[1:length(g)]
 b <- param[(length(g) + 1):length(phi)]
 lpz <- t(g) %*% Z
 lpx <- t(b) %*% X
 wmu <- (1 + exp(lpz)) * exp(lpx)
 wval <- W(wmu)
 sum((delta / pr.est) * (J * log(exp(lpz) + exp(1 - exp(wval))) - log(1 +
     exp(lpz)) +
 (1 - J) * (1 - exp(wval) + Y * log(wval))))
 }

 mleIPW <- maxLik(logLik = loglikfunIPW, start = rep(0, length(phi)))
 estimatesIPW[, ind] <- mleIPW$est
 est.covIPW[, ind] <- diag(vcov(mleIPW))
}

# Summary statistics
biais <- rbind(
 rowMeans(estimatesFD),
 rowMeans(estimatesCC),
 rowMeans(estimatesIPW)
) - phi

rel.biais <- abs(biais / phi) * 100

SD <- rbind(
 apply(estimatesFD, 1, sd),
 apply(estimatesCC, 1, sd),
 apply(estimatesIPW, 1, sd)
)

SE <- sqrt(rbind(
 rowMeans(est.covFD),
 rowMeans(est.covCC),
 rowMeans(est.covIPW)
))

MSE <- rbind(
 rowMeans((estimatesFD - phi)^2),
 rowMeans((estimatesCC - phi)^2),
 rowMeans((estimatesIPW - phi)^2)
)

# Confidence intervals and coverage
CI <- function(est, cov) list(
 lower = est - 1.96 * sqrt(cov),
 upper = est + 1.96 * sqrt(cov)
)

ICFD <- CI(estimatesFD, est.covFD)
ICCC <- CI(estimatesCC, est.covCC)
ICIPW <- CI(estimatesIPW, est.covIPW)

lengthIC <- function(lower, upper) rowMeans(upper - lower)

coverage <- function(lower, upper, trueval) {
 apply(lower <= phi & phi <= upper, 1, mean)
}

lengthICFD <- lengthIC(ICFD$lower, ICFD$upper)
lengthICCC <- lengthIC(ICCC$lower, ICCC$upper)
lengthICIPW <- lengthIC(ICIPW$lower, ICIPW$upper)

covProbaFD <- coverage(ICFD$lower, ICFD$upper, phi)
covProbaCC <- coverage(ICCC$lower, ICCC$upper, phi)
covProbaIPW <- coverage(ICIPW$lower, ICIPW$upper, phi)

# Final results (for display)
round(biais, 4)
round(rel.biais, 4)
```

```r
round(SD, 4)
round(SE, 4)
round(sqrt(MSE), 4)
t(round(cbind(covProbaFD, covProbaCC, covProbaIPW), 4))
t(round(cbind(lengthICFD, lengthICCC, lengthICIPW), 4))
```

Listing 1 Simulation and estimation for the ZIBell model in R

B2. R code for the data analysis

```r
# Load required libraries
library(maxLik)
library(bellreg)
library(LambertW)
library(lamW)
library(foreign)
library(tibble)
library(MASS)
# Clear environment
rm(list=ls())
# Load data
data <- read.dta("randdata.dta")
attach(data)
table(mdvis) # Check distribution of visits
# Create barplot
mdvis.fac <- factor(mdvis, levels=0:max(mdvis))
barplot(table(mdvis.fac),
 xlab="Number of outpatient visits to an MD",
 ylab="Frequency", col="lightblue")

# Preprocessing variables
Y <- data$mdvis
school <- data$educdec
ghindx <- data$ghindx
age <- data$xage
income <- data$income
female <- as.numeric(as.character(data$female))
lnmeddol <- data$lnmeddol
lnmeddol[is.na(lnmeddol)] <- 0
school[is.na(school)] <- 0
ghindx[is.na(ghindx)] <- 0

# Design matrices
inter <- rep(1, length(Y))
W <- rbind(inter, age, female, income, school)
X <- rbind(inter, age, female, income, school, ghindx)

# Handle missingness
Z <- cbind(school, ghindx)
ind <- function(data) {
 apply(data, 1, function(x) as.integer(all(!is.na(x))))
}
delta <- ind(Z)
(1 - mean(delta)) * 100 # Missing data percentage

# Initial values
b <- rep(1, nrow(X))
g <- rep(1, nrow(W))
phi <- c(b, g)

# Logistic regression for missingness
J <- as.integer(Y == 0)
modele <- glm(delta ~ Y + age + female + income, family=binomial)
pr.est <- predict(modele, type="response")

# Complete case estimation (ZIBell)
loglikfunCC <- function(param) {
 b <- param[1:length(b)]
 g <- param[(length(b)+1):length(phi)]
 etaZ <- t(g) %*% W
 etaX <- t(b) %*% X
 mu <- (1 + exp(etaZ)) * exp(etaX)
 Wmu <- W(mu)
 sum(delta * (J * log(exp(etaZ) + exp(1 - exp(Wmu))) -
 log(1 + exp(etaZ)) +
```

```r
 (1 - J) * (1 - exp(Wmu) + Y * log(Wmu))))
}
mleCC <- maxLik(logLik=loglikfunCC, start=rep(0, length(phi)))
summary(mleCC)

# MZIP estimation
e <- rep(1, nrow(X))
a <- rep(1, nrow(W))
phi <- c(e, a)

loglikfunMZIP <- function(param) {
 e <- param[1:length(e)]
 a <- param[(length(e)+1):length(phi)]
 etaZ <- t(W) %*% a
 etaX <- t(X) %*% e
 sum((delta/pr.est) * (
 -log(1 + exp(etaZ)) +
 J * log(exp(etaZ) + exp(-exp(etaX) * (1 + exp(etaZ)))) +
 (1 - J) * ( - (1 + exp(etaZ)) * exp(etaX) +
 Y * log(1 + exp(etaZ)) + etaX * Y )))
}
mleMZIP <- maxLik(logLik=loglikfunMZIP, start=rep(0, length(phi)))
summary(mleMZIP)
cat("AIC MZIP =", 2 * length(mleMZIP$est) - 2 * mleMZIP$maximum)
cat("BIC MZIP =", log(length(Y)) * length(mleMZIP$est) - 2 * mleMZIP$maximum)

# MZIBell estimation (IPW)
b <- rep(1, nrow(X))
g <- rep(1, nrow(W))
phi <- c(b, g)

loglikfunIPW <- function(param) {
 b <- param[1:length(b)]
 g <- param[(length(b)+1):length(phi)]
 etaZ <- t(g) %*% W
 etaX <- t(b) %*% X
 mu <- (1 + exp(etaZ)) * exp(etaX)
 Wmu <- W(mu)
 sum((delta/pr.est) * (
 J * log(exp(etaZ) + exp(1 - exp(Wmu))) -
 log(1 + exp(etaZ)) +
 (1 - J) * (1 - exp(Wmu) + Y * log(Wmu))))
}
mleIPW <- maxLik(logLik=loglikfunIPW, start=rep(0, length(phi)))
summary(mleIPW)
cat("AIC MZIBell =", 2 * length(mleIPW$est) - 2 * abs(mleIPW$maximum))
cat("BIC MZIBell =", log(length(Y)) * length(mleIPW$est) - 2 * abs(mleIPW$maximum))
# (Optional) MZINB estimation can be added as needed
```

Listing 2 Data preparation and model estimation using the RAND dataset

Statements

Acknowledgements

The authors would like to thank reviewers and editors for their valuable comments and suggestions, which have considerably improved this article.

Author contributions

Conceptualization: K.J.G. Kouakou, K.M. Amani, L. Bamba, B. Soro, O. Hili; data curation: K.J.G. Kouakou, K.M. Amani, L. Bamba; formal analysis: K.J.G. Kouakou, K.M. Amani, L. Bamba, B. Soro; investigation: K.J.G. Kouakou, K.M. Amani, L. Bamba, B. Soro; methodology: K.J.G. Kouakou, K.M. Amani, O. Hili; software: K.J.G. Kouakou, K.M. Amani; supervision: K.J.G. Kouakou, K.M. Amani, O. Hili; validation: K.J.G. Kouakou, K.M. Amani, O. Hili; visualization: K.J.G. Kouakou; writing original draft: K.J.G. Kouakou, K.M. Amani, L. Bamba; writing review and editing: K.J.G. Kouakou. All authors have read and agreed to the published version of the article.

Conflict of interest

The authors declare that they have no conflicts of interest.

Data and code availability

The computational code is available in Appendix B of this article.

Declaration on the use of artificial intelligence (AI) technologies

The authors declare that no generative AI was used in the preparation of this article.

Funding

The authors are grateful to the Centre d'Excellence Africain en Mathématiques, Informatique et TIC (CEA-MITIC) for the first author's outgoing mobility grant.

References

[1] Kouakou, K.J.G., Hili, O., and Dupuy, J.F., 2024. Estimation of zero-inflated bivariate Poisson regression with missing covariates. Communications in Statistics: Theory and Methods, 53, 7216–7243. DOI: 10.1080/03610926.2023.2262637

[2] De Andrade, B.B., Matsuhita, R.Y., Rathie, P.N., Ozelim, L., and De Oliveira, S.B., 2021. On a weighted Poisson distribution and its associated regression model. Chilean Journal of Statistics, 12, 229–252. URL: https://soche.cl/chjs/volumes/12/ChJS-12-02-06.pdf

[3] Castellares, F., Ferrari, S.L., and Lemonte, A.J., 2018. On the Bell distribution and its associated regression model for count data. Applied Mathematical Modelling, 56, 172–185. DOI: 10.1016/j.apm.2017.12.014

[4] Bell, E.T., 1934. Exponential polynomials. Annals of Mathematics, 35, 258–277. DOI: 10.2307/1968431

[5] Long, D., Preisser, J.S., Herring, A.H., and Golin, C.E., 2014. A marginalized zero-inflated Poisson regression model with overall exposure effects. Statistics in Medicine, 33, 5151–5165. DOI: 10.1002/sim.6293

[6] Amani, K.M., Kouakou, K.J.G., and Hili, O., 2025. Marginalized zero-inflated Bell regression models for overdispersed count data. Journal of Statistical Theory and Practice, 19, 17. DOI: 10.1007/s42519-025-00433-7

[7] Schafer, J.L. and Graham, J.W., 2002. Missing data: our view of the state of the art. Psychological Methods, 7, 147–177. DOI: 10.1037/1082-989X.7.2.147

[8] Diallo, A., Diop, A., and Dupuy, J.F., 2019. Estimation in zero-inflated binomial regression with missing covariates. Statistics, 53, 839–865. DOI: 10.1080/02331888.2019.1619741

[9] Lee, S.M., Lukusa, M.T., and Li, C.S., 2020. Estimation of a zero-inflated Poisson regression model with missing covariates via nonparametric multiple imputation methods. Computational Statistics, 35, 725–754. DOI: 10.1007/s00180-019-00930-x

[10] Rubin, D.B., 1976. Inference and missing data. Biometrika, 63, 581–592. DOI: 10.1093/biomet/63.3.581

[11] Lukusa, T.M., Lee, S.M., and Li, C.S., 2016. Semiparametric estimation of a zero-inflated Poisson regression model with missing covariates. Metrika, 79, 457–483. DOI: 10.1007/s00184-015-0563-7

[12] Horvitz, D.G. and Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association, 47, 663–685. DOI: 10.2307/2280784

[13] van der Vaart, A.W., 1998. Asymptotic Statistics. Cambridge University Press, Cambridge, UK.

[14] Corless, R.M., Gonnet, G.H., Hare, D.E.G., Jeffrey, D.J., and Knuth, D.E., 1996. On the Lambert function. Advances in Computational Mathematics, 5, 329–359. DOI: 10.1007/BF02124750

[15] Foutz, R.V., 1977. On the unique consistent solution to the likelihood equations. Journal of the American Statistical Association, 72, 147–148. DOI: 10.1080/01621459.1977.10479926

[16] Henningsen, A. and Toomet, O., 2011. maxLik: A package for maximum likelihood estimation in R. Computational Statistics, 26, 443–458. DOI: 10.1007/s00180-010-0217-1