Asymptotic properties of the maximum likelihood estimator in zero-inated binomial regression

Alpha Oumar Diallo, Aliou Diop, Jean-François Dupuy

Asymptotic properties of the maximum likelihood estimator in zero-inflated binomial regression

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 proposed by Hall (2000) to account for excess zeros in binomial regression. Since then, the model has been applied in various fields, such as ecology and epidemiology. In these applications, maximum likelihood estimation (MLE) is used to derive parameter estimates. However, theoretical properties of the MLE in ZIB regression have not yet been rigorously established. The current paper fills this gap and thus provides a rigorous basis for applying the model. Consistency and asymptotic normality of the MLE in ZIB regression are proved. A consistent estimator of the asymptotic variance-covariance matrix of the MLE is also provided. Finite-sample behavior of the estimator is assessed via simulations. Finally, an analysis of a data set in the field of health economics illustrates the paper.

Keywords: Asymptotic normality, consistency, count data, excess of zeros, simulations.

1. Introduction

Zero-inflated regression models have attracted a great deal of attention over the past two decades. These models account for excess zeros in count data by mixing a degenerate distribution with point mass of one at zero with a standard count regression model, such as Poisson, negative binomial or binomial. The zero-inflated Poisson (ZIP) regression model was proposed by Lambert (1992) and further developed by Dietz and Böhning (2000), Li (2011), Lim *et al.* (2014) and Monod (2014), among many others. Zero-inflated negative binomial (ZINB) regression was proposed by Ridout *et al.* (2001), see also Moghimbeigi *et al.* (2008), Mwalili *et al.* (2008), Garay *et al.* (2011). The zero-inflated binomial (ZIB) regression model was discussed by Hall (2000), Vieira *et al.* (2000) and Hall and Berenhaut (2002). Since their introduction, these models have been applied in numerous fields, such as agriculture, econometrics, epidemiology, insurance, species abundance, terrorism study, traffic safety research... In particular, ZIB regression model was recently used in dental caries epidemiology (Gilthorpe *et al.*, 2009; Matranga *et al.*, 2013). This increasing interest for

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)

August 1, 2016

zero-inflated models renders necessary to establish theoretical properties for their parameter estimates. So far, however, mathematical considerations in zero-inflated models (such as asymptotic properties of maximum likelihood estimates) have attracted much less attention than applications. Moreover, the existing literature essentially focuses on the ZIP regression model. See, for example, Min and Czado (2010) who establish asymptotic properties of maximum likelihood estimates (MLE) in a zero-modified generalized Poisson regression model. But to the best of our knowledge, no asymptotic results have been provided for the zero-inflated binomial regression model. In this paper, we investigate this issue.

In the ZIB model proposed by Hall (2000), the individual observation is a bounded count which can be thought of as the number of successes occurring out of a finite number of trials. The mixing probabilities and success probabilities are assumed to follow logistic regression models with parameters γ and β respectively. We provide rigorous proofs of consistency and asymptotic normality of the maximum likelihood estimators of γ and β. We also conduct a simulation study to evaluate finite-sample performance of these estimators. All these results provide a firm basis for making statistical inference in the zero-inflated binomial regression model.

The remainder of this paper is organized as follows. In Section 2, we recall the definition of the ZIB model, we describe maximum likelihood estimation and we introduce some useful notations. In Section 3, we state some regularity conditions and establish consistency and asymptotic normality of the maximum likelihood estimator in ZIB regression. Section 4 reports results of the simulation study. An application of ZIB model to the analysis of health-care utilization by elderlies in United States is described in Section 5. A discussion and some perspectives are provided in Section 6.

2. Zero-inflated binomial regression model

In this section, we briefly recall the definition of the ZIB model, we describe maximum likelihood estimation in ZIB regression and we introduce some useful notations.

2.1. Model and estimation

Let $(Z_i, \mathbf{X}_i, \mathbf{W}_i)$, $i = 1, \ldots, n$ be independent random vectors defined on the probability space $(\Omega, \mathcal{C}, \mathbb{P})$. For every $i = 1, \ldots, n$, the response variable Z_i is generated from the following two-state process:

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

where $\mathcal{B}(m, \pi)$ denotes the binomial distribution with size m and success (or event) probability π. Thus, Z_i follows a standard binomial distribution with probability $1 - p_i$. The first state (also called zero state) occurs with probability p_i. No success can occur in the zero state. The ZIB model reduces to a standard binomial distribution if $p_i = 0$, while $p_i > 0$ leads to zero-inflation. In Hall (2000), the mixing probabilities p_i and event probabilities π_i ($i = 1, \ldots, n$) are modeled by the logistic regression models

$$\text{logit}(p_i) = \gamma^\top \mathbf{W}_i \tag{2.2}$$

and

$$\text{logit}(\pi_i) = \beta^\top \mathbf{X}_i \tag{2.3}$$

respectively, where $\mathbf{X}_i = (1, X_{i2}, \ldots, X_{ip})^\top$ and $\mathbf{W}_i = (1, W_{i2}, \ldots, W_{iq})^\top$ are random vectors of predictors or covariates (both categorical and continuous covariates are allowed) and $\top$ denotes the transpose operator. Let $\psi = (\beta^\top, \gamma^\top)^\top$ be the unknown k-dimensional ($k := p + q$) parameter in models (2.1)-(2.3). The log-likelihood of ψ, based on observations $(Z_i, \mathbf{X}_i, \mathbf{W}_i)$, $i = 1, \ldots, n$, is

$$\begin{aligned} l_n(\psi) &= \sum_{i=1}^{n} \left\{ J_i \log\left(e^{\gamma^\top \mathbf{W}_i} + (1 + e^{\beta^\top \mathbf{X}_i})^{-m_i}\right) - \log\left(1 + e^{\gamma^\top \mathbf{W}_i}\right) \right. \\ &\qquad \left. + (1 - J_i)\left[Z_i \beta^\top \mathbf{X}_i - m_i \log\left(1 + e^{\beta^\top \mathbf{X}_i}\right)\right]\right\}, \\ &:= \sum_{i=1}^{n} l_{[i]}(\psi), \end{aligned} \tag{2.4}$$

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

$$\dot{l}_n(\psi) := \frac{\partial l_n(\psi)}{\partial \psi} = 0. \tag{2.5}$$

In what follows, we establish consistency and asymptotic normality of $\hat{\psi}_n$. First, we need to introduce some further notations.

2.2. Some further notations

Define first the $(p \times n)$ and $(q \times n)$ matrices

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

and let $\mathbb{V}$ be the $(k \times 2n)$ block-matrix defined as

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

where $0_{a,b}$ denotes the $(a \times b)$ matrix whose components are all equal to zero. Let also $C(\psi) = (C_j(\psi))_{1 \le j \le 2n}$ be the $2n$-dimensional column vector defined by

$$C(\psi) = (A_1(\psi), \ldots, A_n(\psi), B_1(\psi), \ldots, B_n(\psi))^\top,$$

where for every $i = 1, \ldots, n$,

$$\begin{aligned} A_i(\psi) &= -J_i \frac{m_i e^{\beta^\top \mathbf{X}_i}}{e^{\gamma^\top \mathbf{W}_i}(h_i(\beta))^{m_i+1} + h_i(\beta)} + (1 - J_i)\left(Z_i - \frac{m_i e^{\beta^\top \mathbf{X}_i}}{h_i(\beta)}\right), \\ B_i(\psi) &= \frac{J_i e^{\gamma^\top \mathbf{W}_i}(h_i(\beta))^{m_i}}{e^{\gamma^\top \mathbf{W}_i}(h_i(\beta))^{m_i} + 1} - \frac{e^{\gamma^\top \mathbf{W}_i}}{1 + e^{\gamma^\top \mathbf{W}_i}}, \end{aligned}$$

and $h_i(\beta) := 1 + e^{\beta^\top \mathbf{X}_i}$. Finally, let $k_i(\psi) := e^{\gamma^\top \mathbf{W}_i}(h_i(\beta))^{m_i+1} + h_i(\beta)$, $i = 1, \ldots, n$. Then, some simple algebra shows that the score equation (2.5) can be rewritten as

$$\dot{l}_n(\psi) = \mathbb{V}C(\psi) = 0.$$

If $A = (A_{ij})_{1 \le i \le a, 1 \le j \le b}$ denotes some $(a \times b)$ matrix, we will denote by $A_{\bullet j}$ its j-th column ($j = 1, \ldots, b$) that is, $A_{\bullet j} = (A_{1j}, \ldots, A_{aj})^\top$. Then, it will be useful to rewrite the score vector as

$$\dot{l}_n(\psi) = \sum_{j=1}^{2n} \mathbb{V}_{\bullet j} C_j(\psi).$$

We shall further denote by $\ddot{l}_n(\psi)$ the $(k \times k)$ matrix of second derivatives of $l_n(\psi)$ that is, $\ddot{l}_n(\psi) = \partial^2 l_n(\psi)/\partial\psi\partial\psi^\top$. Let $\mathbb{D}(\psi) = (\mathbb{D}_{ij}(\psi))_{1 \le i,j \le 2n}$ be the $(2n \times 2n)$ block matrix defined as

$$\mathbb{D}(\psi) = \begin{bmatrix} \mathbb{D}_1(\psi) & \mathbb{D}_3(\psi) \\ \mathbb{D}_3(\psi) & \mathbb{D}_2(\psi) \end{bmatrix},$$

where $\mathbb{D}_1(\psi), \mathbb{D}_2(\psi)$ and $\mathbb{D}_3(\psi)$ are $(n \times n)$ diagonal matrices, with i-th diagonal elements ($i = 1, \ldots, n$) respectively given by

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

Then, some tedious albeit not difficult algebra shows that $\ddot{l}_n(\psi)$ can be expressed as

$$\ddot{l}_n(\psi) = -\mathbb{V}\mathbb{D}(\psi)\mathbb{V}^\top.$$

Note that $C(\psi), \mathbb{V}$ and $\mathbb{D}(\psi)$ depend on n. However, in order to simplify notations, n will not be used as a lower index for these quantities.

In the next section, we establish rigorously the existence, consistency and asymptotic normality of the maximum likelihood estimator $\widehat{\psi}_n$ in the ZIB models (2.1)-(2.3).

3. Regularity conditions and asymptotic properties of the MLE

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

C1 The covariates are bounded that is, there exist compact sets $\mathcal{X} \subset \mathbb{R}^p$ and $\mathcal{W} \subset \mathbb{R}^q$ such that $\mathbf{X}_i \in \mathcal{X}$ and $\mathbf{W}_i \in \mathcal{W}$ for every $i = 1, 2, \ldots$ For every $i = 1, 2, \ldots$, $j = 2, \ldots, p$ and $\ell = 2, \ldots, q$, $\text{var}[X_{ij}] > 0$ and $\text{var}[W_{i\ell}] > 0$. For every $i = 1, 2, \ldots$, the X_{ij} ($j = 1, \ldots, p$) are linearly independent and the $W_{i\ell}$ ($\ell = 1, \ldots, q$) 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 $\mathcal{B} \times \mathcal{G} \subset \mathbb{R}^p \times \mathbb{R}^q$.

C3 The Hessian matrix $\ddot{l}_n(\psi)$ is negative definite and of full rank, for every $n = 1, 2, \ldots$, and $\frac{1}{n}\ddot{l}_n(\psi)$ converges to a negative definite matrix. Let λ_n and Λ_n be respectively the smallest and largest eigenvalues of $\mathbb{V}\mathbb{D}(\psi_0)\mathbb{V}^\top$. There exists a finite positive constant c_1 such that $\Lambda_n/\lambda_n < c_1$ for every $n = 1, 2, \ldots$ The matrix $\mathbb{V}\mathbb{V}^\top$ is positive definite for every $n = 1, 2, \ldots$ and its smallest eigenvalue $\tilde{\lambda}_n$ tends to $+\infty$ as $n \to \infty$.

C4 For every $i = 1, \ldots, n$, $m_i \in \{2, \ldots, M\}$ 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). Recall that for a symmetric real $(k \times k)$-matrix A with eigenvalues $\lambda_1, \ldots, \lambda_k$, $\|A\| = \max_i |\lambda_i|$.

We first prove existence and consistency of $\widehat{\psi}_n$:

Theorem 3.1 (Existence and consistency). *Under conditions C1-C4, the maximum likelihood estimator $\widehat{\psi}_n$ exists almost surely as $n \to \infty$ and converges almost surely to ψ_0.*

Proof of Theorem 3.1. The proof is inspired by the proof of consistency of the MLE in usual logistic regression (Gouriéroux and Monfort, 1981) but technical details are different. We first prove an intermediate technical lemma.

Lemma 3.2. *Let $\phi_n : \mathbb{R}^k \longrightarrow \mathbb{R}^k$ be defined as: $\phi_n(\psi) = \psi + (\mathbb{V}\mathbb{D}(\psi_0)\mathbb{V}^\top)^{-1}\dot{l}_n(\psi)$. Then there exists an open ball $B(\psi_0, r)$ (with $r > 0$) and a constant c $(0 < c < 1)$ such that:*

$$\left\|\phi_n(\psi) - \phi_n(\tilde{\psi})\right\| \leq c\|\psi - \tilde{\psi}\| \text{ for all } \psi, \tilde{\psi} \in B(\psi_0, r). \tag{3.6}$$

Proof of Lemma 3.2. The property (3.6) holds if we can prove that $\left\|\frac{\partial \phi_n(\psi)}{\partial \psi^\top}\right\| \leq c$ for all $\psi \in B(\psi_0, r)$.

Letting I_k be the identity matrix of order k, we have:

$$\begin{aligned}\left\|\frac{\partial\phi_n(\psi)}{\partial\psi^\top}\right\| &= \left\|I_k + (\mathbb{V}\mathbb{D}(\psi_0)\mathbb{V}^\top)^{-1}\ddot{l}_n(\psi)\right\| \\ &= \left\|I_k - (\mathbb{V}\mathbb{D}(\psi_0)\mathbb{V}^\top)^{-1}\mathbb{V}\mathbb{D}(\psi)\mathbb{V}^\top\right\| \\ &= \left\|(\mathbb{V}\mathbb{D}(\psi_0)\mathbb{V}^\top)^{-1}\mathbb{V}(\mathbb{D}(\psi_0)-\mathbb{D}(\psi))\mathbb{V}^\top\right\| \\ &\leq \left\|(\mathbb{V}\mathbb{D}(\psi_0)\mathbb{V}^\top)^{-1}\right\|\left\|\mathbb{V}(\mathbb{D}(\psi_0)-\mathbb{D}(\psi))\mathbb{V}^\top\right\| \\ &= \lambda_n^{-1}\left\|\mathbb{V}(\mathbb{D}(\psi_0)-\mathbb{D}(\psi))\mathbb{V}^\top\right\|.\end{aligned}$$

Now, let $\mathcal{I}$ denote the set of indices $\{(i,j)\in\{1,2,\ldots,2n\}^2$ such that $\mathbb{D}_{ij}(\psi_0)\neq 0\}$. Then the following holds:

$$\begin{aligned}\left\|\mathbb{V}(\mathbb{D}(\psi_0)-\mathbb{D}(\psi))\mathbb{V}^\top\right\| &= \left\|\sum_{i=1}^{2n}\sum_{j=1}^{2n}\mathbb{V}_{\bullet i}\mathbb{V}_{\bullet j}^\top(\mathbb{D}_{ij}(\psi)-\mathbb{D}_{ij}(\psi_0))\right\| \\ &\leq \sum_{(i,j)\in\mathcal{I}}\left\|\mathbb{V}_{\bullet i}\mathbb{V}_{\bullet j}^\top\mathbb{D}_{ij}(\psi_0)\right\|\left|\frac{\mathbb{D}_{ij}(\psi)-\mathbb{D}_{ij}(\psi_0)}{\mathbb{D}_{ij}(\psi_0)}\right|.\end{aligned}$$

From C1 and C2, there exists a constant c_2 ($c_2>0$) such that $|\mathbb{D}_{ij}(\psi_0)|>c_2$ for every $(i,j)\in\mathcal{I}$. For example, consider the case where $\mathbb{D}_{ij}(\psi_0)$ coincides with some $\mathbb{D}_{3,\ell\ell}(\psi_0)$, for $\ell\in\{1,\ldots,n\}$. For every $\psi\in\mathcal{B}\times\mathcal{G}$, we have:

$$|\mathbb{D}_{3,\ell\ell}(\psi)| = \frac{m_\ell e^{\beta^\top\mathbf{X}_\ell+\gamma^\top\mathbf{W}_\ell}(1+e^{\beta^\top\mathbf{X}_\ell})^{m_\ell-1}}{\left(1+e^{\gamma^\top\mathbf{W}_\ell}(1+e^{\beta^\top\mathbf{X}_\ell})^{m_\ell}\right)^2} > \frac{m_{\mathbf{X}}^{m_\ell}m_{\mathbf{W}}}{(1+M_{\mathbf{W}}(1+M_{\mathbf{X}})^{m_\ell})^2},$$

where $m_{\mathbf{X}}:=\min_{\beta,\mathbf{X}}e^{\beta^\top\mathbf{X}}$, $m_{\mathbf{W}}:=\min_{\gamma,\mathbf{W}}e^{\gamma^\top\mathbf{W}}$, $M_{\mathbf{X}}:=\max_{\beta,\mathbf{X}}e^{\beta^\top\mathbf{X}}$, $M_{\mathbf{W}}:=\max_{\gamma,\mathbf{W}}e^{\gamma^\top\mathbf{W}}$. By C1, C2 and C4, there exists a positive constant d_3 such that $\frac{m_{\mathbf{X}}^{m_\ell}m_{\mathbf{W}}}{(1+M_{\mathbf{W}}(1+M_{\mathbf{X}})^{m_\ell})^2}>d_3$. Using similar arguments, we obtain that for every $\psi\in\mathcal{B}\times\mathcal{G}$, $|\mathbb{D}_{1,\ell\ell}(\psi)|>d_1$ and $|\mathbb{D}_{2,\ell\ell}(\psi)|>d_2$ for some $d_1,d_2>0$. Letting $c_2=\min_{1\leq i\leq 3}d_i$, we conclude that $|\mathbb{D}_{ij}(\psi_0)|>c_2$ for every $(i,j)\in\mathcal{I}$. Moreover, $\mathbb{D}_{ij}(\cdot)$ is uniformly continuous on $\mathcal{B}\times\mathcal{G}$ (by Heine theorem) thus for every $\epsilon>0$, there exists a positive number r such that for all $\psi\in B(\psi_0,r)$, $|\mathbb{D}_{ij}(\psi)-\mathbb{D}_{ij}(\psi_0)|<\epsilon$. It follows that

$$\begin{aligned}\left\|\mathbb{V}(\mathbb{D}(\psi_0)-\mathbb{D}(\psi))\mathbb{V}^\top\right\| &\leq \frac{\epsilon}{c_2}\sum_{(i,j)\in\mathcal{I}}\left\|\mathbb{V}_{\bullet i}\mathbb{V}_{\bullet j}^\top\mathbb{D}_{ij}(\psi_0)\right\| \\ &\leq \frac{\epsilon}{c_2}\text{trace}\left(\sum_{(i,j)\in\mathcal{I}}\mathbb{V}_{\bullet i}\mathbb{V}_{\bullet j}^\top\mathbb{D}_{ij}(\psi_0)\right) \\ &= \frac{\epsilon}{c_2}\text{trace}\left(\sum_{i=1}^{2n}\sum_{j=1}^{2n}\mathbb{V}_{\bullet i}\mathbb{V}_{\bullet j}^\top\mathbb{D}_{ij}(\psi_0)\right) \\ &= \frac{\epsilon}{c_2}\text{trace}\left(\mathbb{V}\mathbb{D}(\psi_0)\mathbb{V}^\top\right) \\ &\leq \frac{\epsilon}{c_2}k\Lambda_n.\end{aligned}$$

This in turn implies that $\left\|\frac{\partial\phi_n(\psi)}{\partial\psi^\top}\right\| \leq \frac{\epsilon k\Lambda_n}{c_2\lambda_n} < \frac{\epsilon k c_1}{c_2}$. Now, choosing $\epsilon = c\frac{c_2}{kc_1}$ with $0 < c < 1$, we get that $\left\|\frac{\partial\phi_n(\psi)}{\partial\psi^\top}\right\| \leq c$ for all $\psi \in B(\psi_0, r)$, which concludes the proof. □

We now turn to proof of Theorem 3.1. Define the function $\psi \mapsto \eta_n(\psi)$ by $\eta_n(\psi) := \psi - \phi_n(\psi) = -(\mathbb{V}\mathbb{D}(\psi_0)\mathbb{V}^\top)^{-1}\dot{l}_n(\psi)$. Then $\eta_n(\psi_0)$ converges almost surely to 0 as $n \to \infty$. To see this, note that

$$\eta_n(\psi_0) = (\ddot{l}_n(\psi_0))^{-1} \cdot \dot{l}_n(\psi_0) = \left(\frac{1}{n}\ddot{l}_n(\psi_0)\right)^{-1} \cdot \left(\frac{1}{n}\dot{l}_n(\psi_0)\right).$$

By C3, $\left(\frac{1}{n}\ddot{l}_n(\psi_0)\right)^{-1}$ converges to some matrix Σ. Moreover,

$$\frac{1}{n}\dot{l}_n(\psi_0) = \frac{1}{n}\mathbb{V}C(\psi_0) = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_{i1}A_i(\psi_0) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n X_{ip}A_i(\psi_0) \\ \frac{1}{n}\sum_{i=1}^n W_{i1}B_i(\psi_0) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n W_{iq}B_i(\psi_0) \end{pmatrix}$$

converges to 0 almost surely as $n \to \infty$. To see this, note that for every $i = 1, \ldots, n$ and $j = 1, \ldots, p$:

$$\mathbb{E}[X_{ij}A_i(\psi_0)] = \mathbb{E}[\mathbb{E}[X_{ij}A_i(\psi_0)|\mathbf{X}_i, \mathbf{W}_i]] = \mathbb{E}[X_{ij}\mathbb{E}[A_i(\psi_0)|\mathbf{X}_i, \mathbf{W}_i]],$$

and

$$\begin{aligned}
\mathbb{E}[A_i(\psi_0)|\mathbf{X}_i, \mathbf{W}_i] &= \mathbb{E}\left[-J_i\frac{m_i e^{\beta_0^\top\mathbf{X}_i}}{e^{\gamma_0^\top\mathbf{W}_i}(h_i(\beta_0))^{m_i+1} + h_i(\beta_0)} + (1-J_i)\left(Z_i - \frac{m_i e^{\beta_0^\top\mathbf{X}_i}}{h_i(\beta_0)}\right)\middle|\mathbf{X}_i, \mathbf{W}_i\right] \\
&= -\frac{m_i e^{\beta_0^\top\mathbf{X}_i}}{e^{\gamma_0^\top\mathbf{W}_i}(h_i(\beta_0))^{m_i+1} + h_i(\beta_0)}\mathbb{E}\left[J_i|\mathbf{X}_i, \mathbf{W}_i\right] + \mathbb{E}\left[(1-J_i)Z_i|\mathbf{X}_i, \mathbf{W}_i\right] \\
&\quad -\frac{m_i e^{\beta_0^\top\mathbf{X}_i}}{h_i(\beta_0)}\mathbb{E}\left[1-J_i|\mathbf{X}_i, \mathbf{W}_i\right].
\end{aligned}$$

Now,

$$\begin{aligned}
\mathbb{E}\left[J_i|\mathbf{X}_i, \mathbf{W}_i\right] &= \mathbb{P}(Z_i = 0|\mathbf{X}_i, \mathbf{W}_i) \\
&= p_i + (1-\pi_i)^{m_i}(1-p_i) \\
&= \frac{e^{\gamma_0^\top\mathbf{W}_i}}{1+e^{\gamma_0^\top\mathbf{W}_i}} + \frac{1}{(h_i(\beta_0))^{m_i}\left(1+e^{\gamma_0^\top\mathbf{W}_i}\right)} \\
&= \frac{e^{\gamma_0^\top\mathbf{W}_i}(h_i(\beta_0))^{m_i+1} + h_i(\beta_0)}{(h_i(\beta_0))^{m_i+1}\left(1+e^{\gamma_0^\top\mathbf{W}_i}\right)}
\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\left[(1-J_i)Z_i|\mathbf{X}_i,\mathbf{W}_i\right] &= m_i(1-p_i)\pi_i \\ &= \frac{m_i e^{\beta_0^\top \mathbf{X}_i}}{h_i(\beta_0)\left(1+e^{\gamma_0^\top \mathbf{W}_i}\right)}.\end{aligned}$$

Thus

$$\begin{aligned}\mathbb{E}[A_i(\psi_0)|\mathbf{X}_i,\mathbf{W}_i] &= -\frac{m_i e^{\beta_0^\top \mathbf{X}_i}}{(h_i(\beta_0))^{m_i+1}\left(1+e^{\gamma_0^\top \mathbf{W}_i}\right)} + \frac{m_i e^{\beta_0^\top \mathbf{X}_i}}{h_i(\beta_0)\left(1+e^{\gamma_0^\top \mathbf{W}_i}\right)} - \frac{m_i e^{\beta_0^\top \mathbf{X}_i}}{h_i(\beta_0)} \\ &\quad + \frac{m_i e^{\beta_0^\top \mathbf{X}_i}}{h_i(\beta_0)} \times \frac{e^{\gamma_0^\top \mathbf{W}_i}(h_i(\beta_0))^{m_i+1} + h_i(\beta_0)}{(h_i(\beta_0))^{m_i+1}\left(1+e^{\gamma_0^\top \mathbf{W}_i}\right)} \\ &= \frac{m_i e^{\beta_0^\top \mathbf{X}_i}}{h_i(\beta_0)\left(1+e^{\gamma_0^\top \mathbf{W}_i}\right)} - \frac{m_i e^{\beta_0^\top \mathbf{X}_i}}{h_i(\beta_0)} + \frac{m_i e^{\beta_0^\top \mathbf{X}_i} e^{\gamma_0^\top \mathbf{W}_i}}{h_i(\beta_0)\left(1+e^{\gamma_0^\top \mathbf{W}_i}\right)} \\ &= \frac{m_i e^{\beta_0^\top \mathbf{X}_i}}{h_i(\beta_0)\left(1+e^{\gamma_0^\top \mathbf{W}_i}\right)}\left[1-\left(1+e^{\gamma_0^\top \mathbf{W}_i}\right)\right] + \frac{m_i e^{\beta_0^\top \mathbf{X}_i} e^{\gamma_0^\top \mathbf{W}_i}}{h_i(\beta_0)\left(1+e^{\gamma_0^\top \mathbf{W}_i}\right)} \\ &= -\frac{m_i e^{\beta_0^\top \mathbf{X}_i} e^{\gamma_0^\top \mathbf{W}_i}}{h_i(\beta_0)\left(1+e^{\gamma_0^\top \mathbf{W}_i}\right)} + \frac{m_i e^{\beta_0^\top \mathbf{X}_i} e^{\gamma_0^\top \mathbf{W}_i}}{h_i(\beta_0)\left(1+e^{\gamma_0^\top \mathbf{W}_i}\right)} \\ &= 0.\end{aligned}$$

It follows that $\mathbb{E}[X_{ij}A_i(\psi_0)] = 0$. Similarly, for every $i = 1, \ldots, n$ and $\ell = 1, \ldots, q$, we have:

$$\mathbb{E}[W_{i\ell}B_i(\psi_0)] = \mathbb{E}[\mathbb{E}[W_{i\ell}B_i(\psi_0)|\mathbf{X}_i,\mathbf{W}_i]] = \mathbb{E}[W_{i\ell}\mathbb{E}[B_i(\psi_0)|\mathbf{X}_i,\mathbf{W}_i]]$$

and

$$\begin{aligned}\mathbb{E}[B_i(\psi_0)|\mathbf{X}_i,\mathbf{W}_i] &= \frac{e^{\gamma_0^\top \mathbf{W}_i}(h_i(\beta_0))^{m_i}}{e^{\gamma_0^\top \mathbf{W}_i}(h_i(\beta_0))^{m_i}+1}\mathbb{E}\left[J_i|\mathbf{X}_i,\mathbf{W}_i\right] - \frac{e^{\gamma_0^\top \mathbf{W}_i}}{1+e^{\gamma_0^\top \mathbf{W}_i}} \\ &= 0,\end{aligned}$$

thus $\mathbb{E}[W_{i\ell}B_i(\psi_0)] = 0$. Moreover, for every $i = 1, \ldots, n$ and $\ell = 1, \ldots, q$,

$$\begin{aligned}\mathrm{var}(W_{i\ell}B_i(\psi_0)) &= \mathbb{E}[\mathrm{var}(W_{i\ell}B_i(\psi_0)|\mathbf{X}_i,\mathbf{W}_i)] + \mathrm{var}(\mathbb{E}[W_{i\ell}B_i(\psi_0)|\mathbf{X}_i,\mathbf{W}_i]) \\ &= \mathbb{E}[W_{i\ell}^2 \mathrm{var}(B_i(\psi_0)|\mathbf{X}_i,\mathbf{W}_i)] \\ &= \mathbb{E}\left[W_{i\ell}^2\left(\frac{e^{\gamma_0^\top \mathbf{W}_i}(h_i(\beta_0))^{m_i}}{e^{\gamma_0^\top \mathbf{W}_i}(h_i(\beta_0))^{m_i}+1}\right)^2 \mathrm{var}(J_i|\mathbf{X}_i,\mathbf{W}_i)\right] \\ &\leq \mathbb{E}\left[W_{i\ell}^2\left(\frac{e^{\gamma_0^\top \mathbf{W}_i}(h_i(\beta_0))^{m_i}}{e^{\gamma_0^\top \mathbf{W}_i}(h_i(\beta_0))^{m_i}+1}\right)^2\right].\end{aligned}$$

Therefore, by C1, C2 and C4, there exists a finite constant c_3 such that $\text{var}(W_{i\ell}B_i(\psi_0)) \leq c_3$. Similarly, there exists a finite constant c_4 such that $\text{var}(X_{ij}A_i(\psi_0)) \leq c_4$ for every $i = 1, \ldots, n$ and $j = 1, \ldots, p$. It follows that

$$\sum_{i=1}^{\infty} \frac{\text{var}(W_{i\ell}B_i(\psi_0))}{i^2} \leq c_3 \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty$$

and

$$\sum_{i=1}^{\infty} \frac{\text{var}(X_{ij}A_i(\psi_0))}{i^2} \leq c_4 \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty.$$

Kolmogorov's strong law of large numbers (see for example Jiang (2010), Theorem 6.7) implies that for every $j = 1, \ldots, p$,

$$\frac{1}{n} \sum_{i=1}^{n} \{X_{ij}A_i(\psi_0) - \mathbb{E}\left[X_{ij}A_i(\psi_0)\right]\} = \frac{1}{n} \sum_{i=1}^{n} X_{ij}A_i(\psi_0)$$

converges almost surely to 0. Similarly, for every $\ell = 1, \ldots, q$, $\frac{1}{n}\sum_{i=1}^{n} W_{i\ell}B_i(\psi_0)$ converges almost surely to 0. Finally, $\frac{1}{n}\dot{l}_n(\psi_0)$ and $\eta_n(\psi_0)$ converge almost surely to 0 as $n \to \infty$.

Now, let ϵ be an arbitrary positive value. Almost sure convergence of $\eta_n(\psi_0)$ implies that for almost every $\omega \in \Omega$, there exists an integer $n(\epsilon, \omega)$ such that for any $n \geq n(\epsilon, \omega)$, $\|\eta_n(\psi_0)\| \leq \epsilon$ or equivalently, $0 \in B(\eta_n(\psi_0), \epsilon)$. In particular, let $\epsilon = (1-c)s$ with $0 < c < 1$ such as in Lemma 3.2. Since ϕ_n satisfies the Lipschitz condition (3.6), Lemma 2 of Gouriéroux and Monfort (1981) ensures that there exists an element of $B(\psi_0, s)$ (let $\widehat{\psi}_n$ denote this element) such that $\eta_n(\widehat{\psi}_n) = 0$ that is,

$$(\mathbb{V}\mathbb{D}(\psi_0)\mathbb{V}^\top)^{-1}\dot{l}_n(\widehat{\psi}_n) = 0.$$

Condition C3 implies that $\dot{l}_n(\widehat{\psi}_n) = 0$ and that $\widehat{\psi}_n$ is the unique maximizer of l_n.

To summarize, we have shown that for almost every $\omega \in \Omega$ and for every $s > 0$, there exists an integer value $n(s, \omega)$ such that if $n \geq n(s, \omega)$, then the maximum likelihood estimator $\widehat{\psi}_n$ exists, and $\|\widehat{\psi}_n - \psi_0\| \leq s$ (that is, $\widehat{\psi}_n$ converges almost surely to ψ_0). □

We now turn to asymptotic normality of the maximum likelihood estimator in the ZIB regression model.

Theorem 3.3 (Asymptotic normality). *Let* $\widehat{\Sigma}_n := \mathbb{V}\mathbb{D}(\widehat{\psi}_n)\mathbb{V}^\top$. *Then, under conditions C1-C4,* $\widehat{\Sigma}_n^{\frac{1}{2}}(\widehat{\psi}_n - \psi_0)$ *converges in distribution, as* $n \to \infty$, *to the Gaussian vector* $\mathcal{N}(0, I_k)$.

Proof of Theorem 3.3. A Taylor expansion of the score function yields

$$0 = \dot{l}_n(\widehat{\psi}_n) = \dot{l}_n(\psi_0) + \ddot{l}_n(\widetilde{\psi}_n)(\widehat{\psi}_n - \psi_0),$$

where $\widetilde{\psi}_n$ lies between $\widehat{\psi}_n$ and ψ_0. Thus, $\dot{l}_n(\psi_0) = -\ddot{l}_n(\widetilde{\psi}_n)(\widehat{\psi}_n - \psi_0)$. Letting $\widetilde{\Sigma}_n := -\ddot{l}_n(\widetilde{\psi}_n) = \mathbb{V}\mathbb{D}(\widetilde{\psi}_n)\mathbb{V}^\top$ and $\Sigma_{n,0} := \mathbb{V}\mathbb{D}(\psi_0)\mathbb{V}^\top$, we have:

$$\widehat{\Sigma}_n^{\frac{1}{2}}(\widehat{\psi}_n - \psi_0) = \left[\widehat{\Sigma}_n^{\frac{1}{2}}\widetilde{\Sigma}_n^{-\frac{1}{2}}\right]\left[\widetilde{\Sigma}_n^{-\frac{1}{2}}\Sigma_{n,0}^{\frac{1}{2}}\right]\Sigma_{n,0}^{-\frac{1}{2}}\left(\widetilde{\Sigma}_n(\widehat{\psi}_n - \psi_0)\right). \tag{3.7}$$

The terms $[\widehat{\Sigma}_n^{\frac{1}{2}}\widetilde{\Sigma}_n^{-\frac{1}{2}}]$ and $[\widetilde{\Sigma}_n^{-\frac{1}{2}}\Sigma_{n,0}^{\frac{1}{2}}]$ in (3.7) converge almost surely to I_k. To see this, we show for example that $\|\widetilde{\Sigma}_n^{-\frac{1}{2}}\Sigma_{n,0}^{\frac{1}{2}} - I_k\| \xrightarrow{a.s.} 0$ as $n \to \infty$. First, note that

$$\left\|\widetilde{\Sigma}_n^{-\frac{1}{2}}\Sigma_{n,0}^{\frac{1}{2}} - I_k\right\| \leq \Lambda_n^{\frac{1}{2}}\left\|\widetilde{\Sigma}_n^{-\frac{1}{2}}\right\|\left\|\Lambda_n^{-\frac{1}{2}}\left(\Sigma_{n,0}^{\frac{1}{2}} - \widetilde{\Sigma}_n^{\frac{1}{2}}\right)\right\|, \tag{3.8}$$

and

$$\Lambda_n^{-1}\left\|\Sigma_{n,0} - \widetilde{\Sigma}_n\right\| = \Lambda_n^{-1}\left\|\mathbb{V}(\mathbb{D}(\psi_0) - \mathbb{D}(\widetilde{\psi}_n))\mathbb{V}^\top\right\|.$$

By Theorem 3.1, $\widetilde{\psi}_n$ converges almost surely to ψ_0. Let $\omega \in \Omega$ be outside the negligible set where this convergence does not hold. By the same arguments as in proof of Lemma 3.2, for every $\epsilon > 0$, there exists $n(\epsilon, \omega) \in \mathbb{N}$ such that if $n \geq n(\epsilon, \omega)$, then $\Lambda_n^{-1}\|\mathbb{V}(\mathbb{D}(\psi_0) - \mathbb{D}(\widetilde{\psi}_n))\mathbb{V}^\top\| \leq \epsilon$. Thus $\Lambda_n^{-1}\|\mathbb{V}(\mathbb{D}(\psi_0) - \mathbb{D}(\widetilde{\psi}_n))\mathbb{V}^\top\|$ converges almost surely to 0. By continuity of the map $A \mapsto A^{\frac{1}{2}}$, $\|\Lambda_n^{-\frac{1}{2}}(\Sigma_{n,0}^{\frac{1}{2}} - \widetilde{\Sigma}_n^{\frac{1}{2}})\|$ converges almost surely to 0. Moreover, for n sufficiently large, there exists $0 < c_5 < \infty$ such that almost surely, $\Lambda_n^{\frac{1}{2}}\|\widetilde{\Sigma}_n^{-\frac{1}{2}}\| \leq c_5\Lambda_n^{\frac{1}{2}}/\lambda_n^{\frac{1}{2}} < c_5 c_1^{\frac{1}{2}}$ (by condition C3). Thus $\|\widetilde{\Sigma}_n^{-\frac{1}{2}}\Sigma_{n,0}^{\frac{1}{2}} - I_k\|$ converges almost surely to 0. Almost sure convergence of $\|\widehat{\Sigma}_n^{\frac{1}{2}}\widetilde{\Sigma}_n^{-\frac{1}{2}} - I_k\|$ to 0 follows by similar arguments.

It remains us to show that $\Sigma_{n,0}^{-\frac{1}{2}}(\widetilde{\Sigma}_n(\widehat{\psi}_n - \psi_0))$ converges in distribution to the Gaussian vector $\mathcal{N}(0, I_k)$. Note that $\Sigma_{n,0}^{-\frac{1}{2}}(\widetilde{\Sigma}_n(\widehat{\psi}_n - \psi_0)) = \Sigma_{n,0}^{-\frac{1}{2}}\sum_{j=1}^{2n}\mathbb{V}_{\bullet j}C_j(\psi_0)$. Thus, by Eicker (1966), this convergence holds if we can check that the following conditions are fulfilled: 1) $\max_{1\leq j\leq 2n}\mathbb{V}_{\bullet j}^\top(\mathbb{V}\mathbb{V}^\top)^{-1}\mathbb{V}_{\bullet j} \to 0$ as $n \to \infty$, 2) $\sup_{1\leq j\leq 2n}\mathbb{E}[C_j(\psi_0)^2 1_{\{|C_j(\psi_0)|>c\}}] \to 0$ as $c \to \infty$, 3) $\inf_{1\leq j\leq 2n}\mathbb{E}[C_j(\psi_0)^2] > 0$. Condition 1) follows by noting that

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

and that $\|\mathbb{V}_{\bullet j}\|$ is bounded, by C1. Moreover, $1/\tilde{\lambda}_n$ tends to 0 as $n \to \infty$ by C3. Condition 2) follows by noting that the $C_j(\psi_0)$, $j = 1, \ldots, 2n$ are bounded under C1, C2, C4. Finally, we note that $\mathbb{E}[C_j(\psi_0)^2] = \mathrm{var}(C_j(\psi_0))$ since $\mathbb{E}[C_j(\psi_0)] = 0$, $j = 1, \ldots, 2n$. If $j \in \{n+1, \ldots, 2n\}$, $C_j(\psi_0) = B_{j'}(\psi_0)$, with $j' = j - n$. Then $\mathrm{var}(C_j(\psi_0)) = \mathrm{var}(B_{j'}(\psi_0)) =$

$\mathbb{E}[\mathrm{var}(B_{j'}(\psi_0)|\mathbf{X}_{j'}, \mathbf{W}_{j'})] + \mathrm{var}(\mathbb{E}[B_{j'}(\psi_0)|\mathbf{X}_{j'}, \mathbf{W}_{j'}]) = \mathbb{E}[\mathrm{var}(B_{j'}(\psi_0)|\mathbf{X}_{j'}, \mathbf{W}_{j'})]$. Now,

$$\begin{aligned}
\mathrm{var}(B_{j'}(\psi_0)|\mathbf{X}_{j'}, \mathbf{W}_{j'}) &= \left(\frac{e^{\gamma_0^\top \mathbf{W}_{j'}}(h_{j'}(\beta_0))^{m_{j'}}}{e^{\gamma_0^\top \mathbf{W}_{j'}}(h_{j'}(\beta_0))^{m_{j'}}+1}\right)^2 \mathrm{var}(J_{j'}|\mathbf{X}_{j'}, \mathbf{W}_{j'}) \\
&= \left(\frac{e^{\gamma_0^\top \mathbf{W}_{j'}}(h_{j'}(\beta_0))^{m_{j'}}}{e^{\gamma_0^\top \mathbf{W}_{j'}}(h_{j'}(\beta_0))^{m_{j'}}+1}\right)^2 \mathbb{P}(Z_{j'}=0|\mathbf{X}_{j'}, \mathbf{W}_{j'})(1-\mathbb{P}(Z_{j'}=0|\mathbf{X}_{j'}, \mathbf{W}_{j'})) \\
&= \left(\frac{e^{\gamma_0^\top \mathbf{W}_{j'}}(h_{j'}(\beta_0))^{m_{j'}}}{e^{\gamma_0^\top \mathbf{W}_{j'}}(h_{j'}(\beta_0))^{m_{j'}}+1}\right)^2 \left(p_{j'} + (1-\pi_{j'})^{m_{j'}}(1-p_{j'})\right)(1-p_{j'}) \\
&\qquad \times (1-(1-\pi_{j'})^{m_{j'}}),
\end{aligned}$$

and thus, $\mathrm{var}(B_{j'}(\psi_0)|\mathbf{X}_{j'}, \mathbf{W}_{j'}) > 0$ for every $j' = 1, \ldots, n$ by C1, C2, C4. It follows that $\mathrm{var}(C_j(\psi_0)) > 0$ for every $j = n+1, \ldots, 2n$. By similar arguments, $\mathrm{var}(C_j(\psi_0)) > 0$ for every $j = 1, \ldots, 2n$ and condition 3) is satisfied.

To summarize, we have proved that $\Sigma_{n,0}^{-\frac{1}{2}}(\widetilde{\Sigma}_n(\widehat{\psi}_n - \psi_0))$ converges in distribution to $\mathcal{N}(0, I_k)$. This result combined with Slutsky's theorem and equation (3.7) implies that $\widehat{\Sigma}_n^{\frac{1}{2}}(\widehat{\psi}_n - \psi_0)$ converges in distribution to $\mathcal{N}(0, I_k)$. □

4. Simulation study

In this section, we assess finite-sample properties of the maximum likelihood estimator $\hat{\psi}_n$.

4.1. Study design

We generate data from the following ZIB regression model:

$$\mathrm{logit}(\pi_i) = \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_4 X_{i4} + \beta_5 X_{i5} + \beta_6 X_{i6} + \beta_7 X_{i7}$$

and

$$\mathrm{logit}(p_i) = \gamma_1 W_{i1} + \gamma_2 W_{i2} + \gamma_3 W_{i3} + \gamma_4 W_{i4} + \gamma_5 W_{i5},$$

where $X_{i1} = 1$ and the $X_{i2}, \ldots, X_{i7}$ 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. We let $W_{i1} = 1$ and W_{i4} and W_{i5} be independently drawn from normal $\mathcal{N}(-1,1)$ and binomial $\mathcal{B}(1,0.5)$ distributions respectively. The linear predictors in $\mathrm{logit}(\pi_i)$ and $\mathrm{logit}(p_i)$ are allowed to share some common terms by letting $W_{i2} = X_{i2}$ and $W_{i3} = X_{i6}$. The regression parameter β is chosen as $\beta = (-0.3, 1.2, 0.5, -0.75, -1, 0.8, 0)^\top$. The regression parameter γ is chosen as:

- case 1: $\gamma = (-0.55, -0.7, -1, 0.45, 0)^\top$
- case 2: $\gamma = (0.25, -0.4, 0.8, 0.45, 0)^\top$

We consider several sample sizes, namely $n = 150, 300, 500$. The numbers m_i are allowed to vary across subjects, with $m_i \in \{4, 5, 6\}$. Let $(n_4, n_5, n_6) = (\text{card}\{i : m_i = 4\}, \text{card}\{i : m_i = 5\}, \text{card}\{i : m_i = 6\})$. For $n = 150$, we let $(n_4, n_5, n_6) = (60, 50, 40)$. For $n = 300$, we let $(n_4, n_5, n_6) = (120, 100, 80)$ and for $n = 500$, we let $(n_4, n_5, n_6) = (200, 170, 130)$.
Using these values, in case 1 (respectively case 2), the average proportion of zero-inflation in the simulated data sets is 25% (respectively 50%). For each combination of the simulation design parameters (sample size and zero-inflation proportion), we simulate $N = 5000$ samples and we calculate the maximum likelihood estimate $\hat{\psi}_n$.

Computational aspects of maximum likelihood estimation in ZIB regression are discussed by Hall (2000). There, the author develops an EM algorithm for estimating ψ. Alternatively, he also suggests to use Newton-Raphson algorithm for solving (2.5). In his paper, Hall (2000) motivated his preference for the EM algorithm by programming simplicity. Since then, numerous `R` packages (R Core Team, 2013) have been developed for maximizing log-likelihoods such as (2.4) or for solving likelihood equations such as (2.5). In our simulation study, we use the `R` package `maxLik` (Henningsen and Toomet, 2011).

4.2. Results

For each configuration `sample size` $\times$ `zero-inflation proportion` of the simulation design parameters, we calculate the average bias of the estimates $\hat{\beta}_{j,n}$ and $\hat{\gamma}_{k,n}$ of the β_j and γ_k over the N estimates. Based on the N simulated samples, we also obtain the average standard error (SE) and empirical standard deviation (SD) for each estimator $\hat{\beta}_{j,n}$ $(j = 1, \ldots, 7)$ and $\hat{\gamma}_{k,n}$ $(k = 1, \ldots, 5)$. Finally, we obtain 95%-level confidence intervals for the β_j and γ_k. We provide their empirical coverage probability (CP) and average length ℓ(CI). Results are given in Table 1 (case 1) and Table 2 (case 2).

Finally, in order to assess the quality of the Gaussian approximation stated in Theorem 3.3, we provide normal Q-Q plots of the estimates (see figures 1 and 2 for $n = 300$ in case 1 and figures 3 and 4 for $n = 300$ in case 2. Plots for $n = 150$ and $n = 500$ yield similar observations and are thus omitted).

From these results, it appears as expected that the bias, SE, SD and ℓ(CI) of all estimators decrease as the sample size increases. The bias stays moderate provided that the sample size is large enough (say, $n \geq 300$). The empirical coverage probabilities are close to the nominal confidence level, even when the sample size is moderate. As may also be expected, we observe that the maximum likelihood estimator of the β_js (respectively γ_ks) performs better when the zero-inflation proportion decreases (respectively increases). Finally, it appears from normal Q-Q plots that the Gaussian approximation of the distribution of the maximum likelihood estimator in ZIB regression is reasonably satisfied, even when the sample size is moderate.

n		$\widehat{\beta}_n$							$\widehat{\gamma}_n$				
		$\widehat{\beta}_{1,n}$	$\widehat{\beta}_{2,n}$	$\widehat{\beta}_{3,n}$	$\widehat{\beta}_{4,n}$	$\widehat{\beta}_{5,n}$	$\widehat{\beta}_{6,n}$	$\widehat{\beta}_{7,n}$	$\widehat{\gamma}_{1,n}$	$\widehat{\gamma}_{2,n}$	$\widehat{\gamma}_{3,n}$	$\widehat{\gamma}_{4,n}$	$\widehat{\gamma}_{5,n}$
150													
	bias	-0.0116	0.0297	0.0132	-0.0230	-0.0317	0.0286	0.0009	-0.0753	-0.0504	-0.0736	0.0512	0.0074
	SD	0.5686	0.1673	0.1484	0.1018	0.1635	0.2755	0.1278	0.5208	0.3583	0.7332	0.3150	0.5887
	SE	0.5546	0.1635	0.1454	0.0993	0.1596	0.2706	0.1241	0.5038	0.3441	0.8796	0.3061	0.5771
	CP	0.9446	0.9419	0.9436	0.9440	0.9459	0.9486	0.9459	0.9609	0.9534	0.9659	0.9576	0.9586
	ℓ(CI)	2.1648	0.6375	0.5682	0.3877	0.6210	1.0572	0.4840	1.9387	1.3256	2.8945	1.1760	2.2280
300													
	bias	-0.0105	0.0173	0.0072	-0.0100	-0.0150	0.0097	-0.0015	-0.0359	-0.0206	-0.0732	0.0242	0.0115
	SD	0.3887	0.1140	0.1012	0.0689	0.1128	0.1887	0.0863	0.3411	0.2281	0.5015	0.2078	0.3838
	SE	0.3793	0.1120	0.0996	0.0681	0.1088	0.1849	0.0845	0.3304	0.2255	0.4953	0.1998	0.3794
	CP	0.9467	0.9499	0.9489	0.9487	0.9427	0.9457	0.9423	0.9503	0.9501	0.9595	0.9479	0.9559
	ℓ(CI)	1.4843	0.4380	0.3900	0.2663	0.4251	0.7240	0.3304	1.2894	0.8791	1.8958	0.7782	1.4824
500													
	bias	0.0009	0.0092	0.0027	-0.0066	-0.0096	0.0081	0.0010	-0.0170	-0.0094	-0.0395	0.0157	0.0010
	SD	0.2925	0.0856	0.0766	0.0518	0.0828	0.1418	0.0664	0.2545	0.1740	0.3687	0.1554	0.2922
	SE	0.2910	0.0858	0.0764	0.0521	0.0832	0.1418	0.0647	0.2497	0.1702	0.3631	0.1508	0.2874
	CP	0.9490	0.9498	0.9480	0.9508	0.9518	0.9500	0.9436	0.9514	0.9470	0.9548	0.9506	0.9510
	ℓ(CI)	1.1397	0.3359	0.2993	0.2041	0.3255	0.5555	0.2531	0.9766	0.6653	1.4107	0.5891	1.1246

Table 1: Simulation results (case 1). SE: average standard error. SD: empirical standard deviation. CP: empirical coverage probability of 95%-level confidence intervals. ℓ(CI): average length of confidence intervals. All results are based on $N = 5000$ simulated samples.

n		$\widehat{\beta}_n$							$\widehat{\gamma}_n$				
		$\widehat{\beta}_{1,n}$	$\widehat{\beta}_{2,n}$	$\widehat{\beta}_{3,n}$	$\widehat{\beta}_{4,n}$	$\widehat{\beta}_{5,n}$	$\widehat{\beta}_{6,n}$	$\widehat{\beta}_{7,n}$	$\widehat{\gamma}_{1,n}$	$\widehat{\gamma}_{2,n}$	$\widehat{\gamma}_{3,n}$	$\widehat{\gamma}_{4,n}$	$\widehat{\gamma}_{5,n}$
150													
	bias	-0.0396	0.0607	0.0278	-0.0378	-0.0589	0.0414	-0.0008	-0.0019	-0.0034	0.0507	0.0430	-0.0108
	SD	0.7568	0.2220	0.1976	0.1384	0.2257	0.4291	0.1719	0.4184	0.2593	0.4788	0.2377	0.4307
	SE	0.7228	0.2115	0.1909	0.1313	0.2154	0.4078	0.1639	0.4084	0.2490	0.4679	0.2298	0.4278
	CP	0.9395	0.9451	0.9445	0.9409	0.9443	0.9467	0.9431	0.9549	0.9517	0.9515	0.9527	0.9559
	ℓ(CI)	2.8115	0.8214	0.7435	0.5105	0.8330	1.5810	0.6358	1.5947	0.9704	1.8286	0.8948	1.6738
300													
	bias	-0.0180	0.0263	0.0121	-0.0160	-0.0278	0.0249	-0.0018	-0.0011	-0.0046	0.0255	0.0214	0.0043
	SD	0.4954	0.1447	0.1311	0.0896	0.1462	0.2784	0.1088	0.2808	0.1694	0.3184	0.1581	0.3014
	SE	0.4851	0.1415	0.1282	0.0881	0.1430	0.2707	0.1088	0.2780	0.1672	0.3174	0.1550	0.2918
	CP	0.9466	0.9482	0.9496	0.9492	0.9428	0.9448	0.9508	0.9498	0.9460	0.9522	0.9490	0.9446
	ℓ(CI)	1.8953	0.5524	0.5011	0.3441	0.5572	1.0566	0.4245	1.0883	0.6541	1.2429	0.6058	1.1431
500													
	bias	-0.0075	0.0151	0.0073	-0.0109	-0.0163	0.0142	-0.0016	-0.0027	-0.0030	0.0160	0.0108	0.0022
	SD	0.3707	0.1101	0.0982	0.0679	0.1083	0.2094	0.0837	0.2133	0.1298	0.2448	0.1175	0.2237
	SE	0.3684	0.1075	0.0974	0.0670	0.1083	0.2053	0.0824	0.2125	0.1273	0.2423	0.1178	0.2231
	CP	0.9492	0.9472	0.9502	0.9516	0.9498	0.9446	0.9458	0.9502	0.9492	0.9484	0.9516	0.9510
	ℓ(CI)	1.4413	0.4204	0.3813	0.2621	0.4229	0.8029	0.3224	0.8323	0.4984	0.9492	0.4613	0.8744

Table 2: Simulation results (case 2). SE: average standard error. SD: empirical standard deviation. CP: empirical coverage probability of 95%-level confidence intervals. ℓ(CI): average length of confidence intervals. All results are based on $N = 5000$ simulated samples.

5. An application of ZIB model to health economics

5.1. Data description and modelling

In this section, we describe an application of ZIB regression to the analysis of health-care utilization by elderlies in the United States. This application is based on data obtained from the National Medical Expenditure Survey (NMES) conducted in 1987-1988. This data set was first described by Deb and Trivedi (1997). It provides a comprehensive picture of how Americans (aged 66 years and over) use and pay for health services. Several measures of health-care utilization were reported in this study, including the number of visits to a doctor in an office setting (denoted by *ofd* in what follows), the number of visits to a non-doctor health professional (such as a nurse, optician, physiotherapist...) in an office setting (*ofnd*), the number of visits to a doctor in an outpatient setting, the number of visits to a non-doctor in an outpatient setting, the number of visits to an emergency service and the number of hospital stays. A feature of these data is the high proportion of zero counts observed for some of the health-care utilization measures. In addition to health services utilization, the data set also contains information on health status, sociodemographic characteristics and economic status. Deb and Trivedi (1997) analyse separately each measure of health-care utilization by fitting zero-inflated count data models to each type of health-care usage in turns.

Here, we consider the following issue. Consider a patient who decides to visit a health professional in an office setting. We wish to identify factors that explain patient's choice between a visit to a doctor and a visit to a non-doctor. For our study, we consider patients in the NMES data set who have a total number of office consultations comprised between 2 and 25. Among these $n = 3227$ patients, frequencies of zero in *ofnd* and *ofd* counts are 62.1% and 1.21% respectively. Let Z_i and m_i be respectively the number of non-doctor office visits and the total number of office visits for the i-th patient ($i = 1, \ldots, 3227$). Given m_i, one may model Z_i as a $\mathcal{B}(m_i, \pi_i)$ distribution. However, the high frequency of zero in *ofnd* count suggests that Z_i is affected by zero-inflation. Therefore, we suggest to use a ZIB model for Z_i. Several covariates are available in the NMES data set, including: i) socio-economic variables: gender (1 for female, 0 for male), age (in years, divided by 10), marital status (1 if married, 0 if not married), educational level (number of years of education), income (in ten-thousands of dollars), ii) various measures of health status: number of chronic conditions (cancer, arthritis, diabete...) and a variable indicating self-perceived health level (poor, average, excellent) and iii) a binary variable indicating whether individual is covered by medicaid or not (medicaid is a US health insurance for individuals with limited income and resources, we code it as 1 if the individual is covered and 0 otherwise). Self-perceived health is re-coded as two dummy variables denoted by "health1" (1 if health is perceived as poor, 0 otherwise) and "health2" (1 if health is perceived as excellent, 0 otherwise). As mentioned above, we wish to identify determinants of patients choice between a doctor and a non-doctor visit. We model zero-inflation and event probabilities p_i and π_i by (2.2) and (2.3) respectively, where $\mathbf{X}_i$ and $\mathbf{W}_i$ are the set of covariates listed above.

First, we fitted a ZIB regression model incorporating all available covariates in (2.2) and (2.3), *i.e.*, letting $\mathbf{X}_i = \mathbf{W}_i$ for every i. Then, Wald tests were used to select rele-

parameter	variable	estimate	s.e.	Wald test of $H_0 : \beta_j = 0$
β_1	intercept	-0.2095	0.2983	NS
β_2	health1	-0.3459	0.0750	VS
β_3	health2	0.2642	0.0816	VS
β_4	chronic	-0.0939	0.0167	VS
β_5	age	-0.0566	0.0360	NS
β_6	gender	0.0687	0.0487	NS
β_7	marital status	0.1372	0.0476	VS
β_8	educational	-0.0031	0.0067	NS
β_9	income	-0.0069	0.0064	NS
β_{10}	medicaid	-0.0911	0.0924	NS
γ_1	intercept	1.1095	0.1549	VS
γ_2	health1	0.3338	0.1284	S
γ_3	gender	-0.3220	0.0873	VS
γ_4	educational	-0.0746	0.0124	VS
γ_5	medicaid	0.4519	0.1621	VS

Table 3: Health-care data analysis (NS: not significant at the 5% level, S: significant at level between 1% and 5%, VS (very significant): significant at level less than 1%).

vant covariates in submodels (2.2) and (2.3). However, this procedure can be cumbersome when the number of covariates is large. Thus, we propose an alternative procedure (here, both procedures yield the same final set of significant predictors). In a first stage, we fit a standard logistic regression model with all available covariates to binary indicators $1_{\{Z_i=0\}}$, $i = 1, \dots, n$. The resulting model is not a model for zero-inflation since some of the 0 may arise from the binomial distribution $\mathcal{B}(m_i, \pi_i)$. However, we expect that this rough procedure will still select a relevant subset of covariates, that will be used in a second stage in the logistic sub-model (2.2) for p_i. Using this procedure and Wald testing, we identify four significant predictors: "health1" dummy variable, gender, educational level and medicaid status, that are included in p_i while all covariates are included in π_i. Results for the resulting ZIB model are displayed in Table 3.

5.2. Results

In Table 3, we report estimate, standard error (s.e.) and significance level (as: not significant, significant or very significant) of Wald test of nullity for each parameter.

As mentioned above, gender, educational level, medicaid status and "health1" dummy variable are identified as the most influencing factors of the decision of never resorting to non-doctor health professionals, with a probability of never resorting which increases when health level degradates (one reason is that patients whose health declines may tend to favor visits to a doctor). Medicaid recipients are more likely to renounce non-doctor office visits. One explanation is that patients with medicaid coverage may limit their consultations to those necessary, that is, to doctor visits only (recall that medicaid is a health insurance for poor people). The probability of never resorting to non-doctor office consultations decreases with the number of years of education. This is coherent with previous findings, *e.g.*, Deb and Trivedi (1997), who postulate that education may make individuals more informed consumers of medical care services. More informed patients may tend to diversify their health-care utilization.

For patients who eventually consult non-doctor health professionals in an office setting, ZIB model suggests that health status variables (number of chronic conditions and self-perceived health) are the most influencing factors of the choice between doctor and non-doctor visit. ZIB model also suggests that patients with poor health will favor visits to doctors over non-doctors, which seems a natural finding. Perhaps surprisingly, marital status has a significant effect on the choice of doctor *vs* non-doctor visit (being married increases the probability of visiting a non-doctor health professional). One explanation is that marital status may capture some income effect leading married patients to diversify their health-care utilization.

6. Discussion

Zero-inflated binomial regression is now commonly used for investigating count data with zeros excess. In this paper, we provide a rigorous basis for maximum likelihood inference in this model. Precisely, we establish consistency and asymptotic normality of the maximum likelihood estimator in ZIB regression. Moreover, our simulation study suggests that the maximum likelihood estimator performs well under a wide range of conditions pertaining to sample size and proportion of zero-inflation.

We consider here the basic ZIB regression model. Hall (2000) proposes to incorporate random effects to this model when the count data are correlated. Several other generalizations of ZIB regression may be developed to account for the increasing complexity of experimental data. For example, one may use partially linear link functions for the mixing and/or success probabilities (such as in the ZIP model, see for example Lam *et al.* (2006) and He *et al.* (2010)). Asymptotic properties of the statistical inference in these generalizations are still unknown and their rigorous derivation remains an open problem. This is a topic for our future work.

Acknowledgements

Authors are grateful to the referees and associate editor for their comments and suggestions on an earlier version of this paper. 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

Deb, P., Trivedi, P. K., 1997. Demand for medical care by the elderly: a finite mixture approach. Journal of Applied Econometrics 12(3), 313-336.

Dietz, E., Böhning, D., 2000. On estimation of the Poisson parameter in zero-modified Poisson models. Computational Statistics & Data Analysis 34(4), 441-459.

Eicker, F., 1966. A multivariate central limit theorem for random linear vector forms. The Annals of Mathematical Statistics 37(6), 1825-1828.

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.

Hall, D. B., Berenhaut, K. S., 2002. Score tests for heterogeneity and overdispersion in zero-inflated Poisson and binomial regression models. The Canadian Journal of Statistics 30(3), 415-430.

He, X., Xue, H., Shi, N.-Z., 2010. Sieve maximum likelihood estimation for doubly semiparametric zero-inflated Poisson models. Journal of Multivariate Analysis 101, 2026-2038.

Henningsen, A., Toomet, O., 2011. maxLik: A package for maximum likelihood estimation in R. Computational Statistics 26(3), 443-458.

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. Technometrics 34, 1-14.

Li, C.-S., 2011. A lack-of-fit test for parametric zero-inflated Poisson models. Journal of Statistical Computation and Simulation 81(9), 1081-1098.

Lim, H. K., Li, W. K., Yu, P. L. H., 2006. Zero-inflated Poisson regression mixture model. Computational Statistics & Data Analysis 71, 151-158.

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, A., Czado, C., 2010. Testing for zero-modification in count regression models. Statistica Sinica 20(1), 323-341.

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.

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.

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.

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.

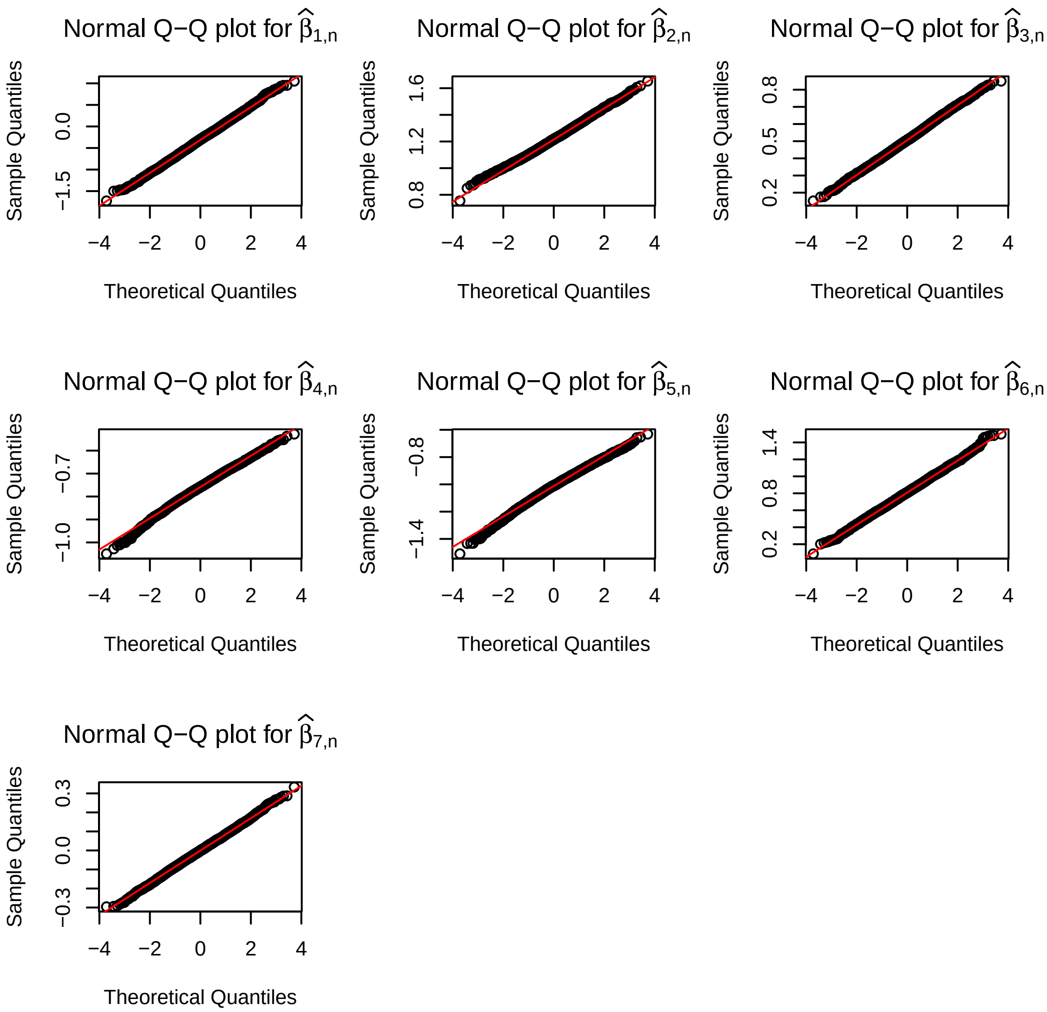

Figure 1: Normal Q-Q plots for $\hat{\beta}_{1,n}, \ldots, \hat{\beta}_{7,n}$ with $n = 300$ (case 1).

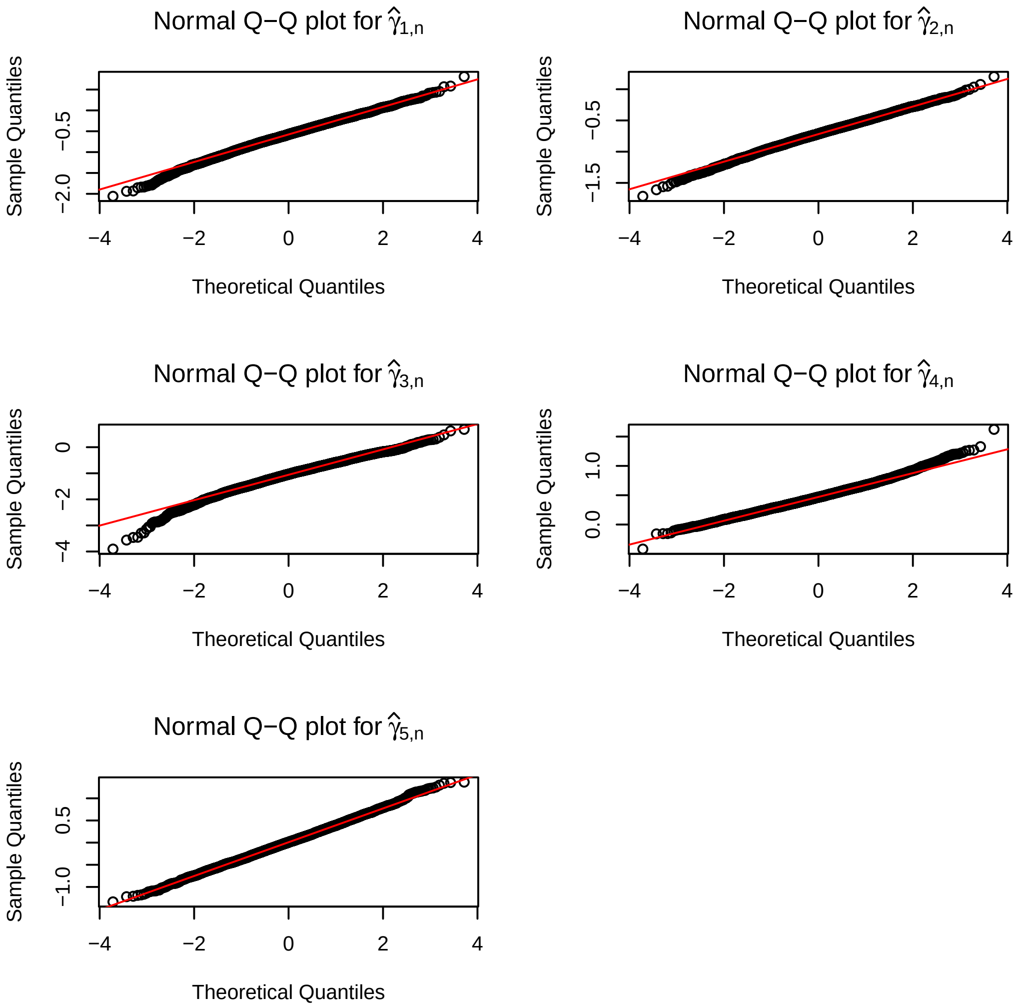

Figure 2: Normal Q-Q plots for $\hat{\gamma}_{1,n}, \ldots, \hat{\gamma}_{5,n}$ with $n = 300$ (case 1).

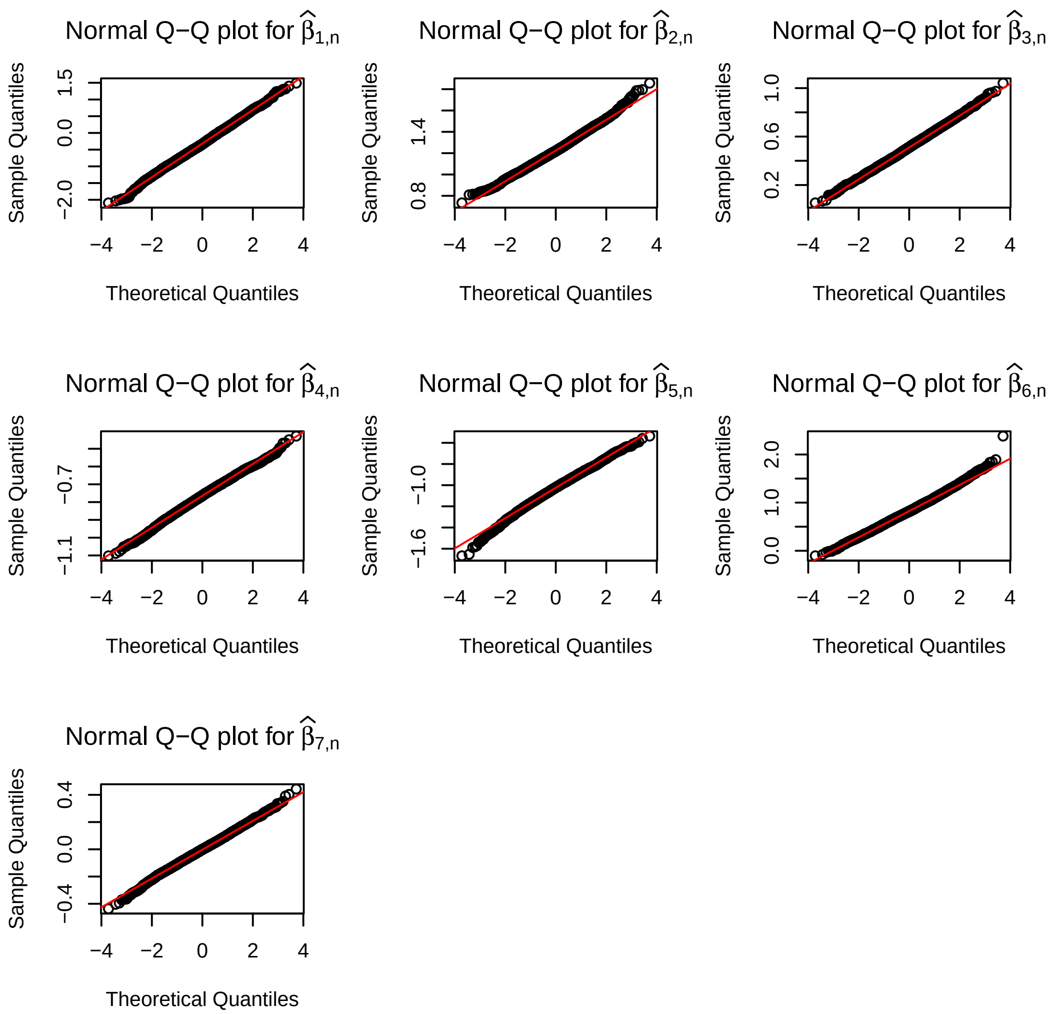

Figure 3: Normal Q-Q plots for $\hat{\beta}_{1,n}, \ldots, \hat{\beta}_{7,n}$ with $n = 300$ (case 2).

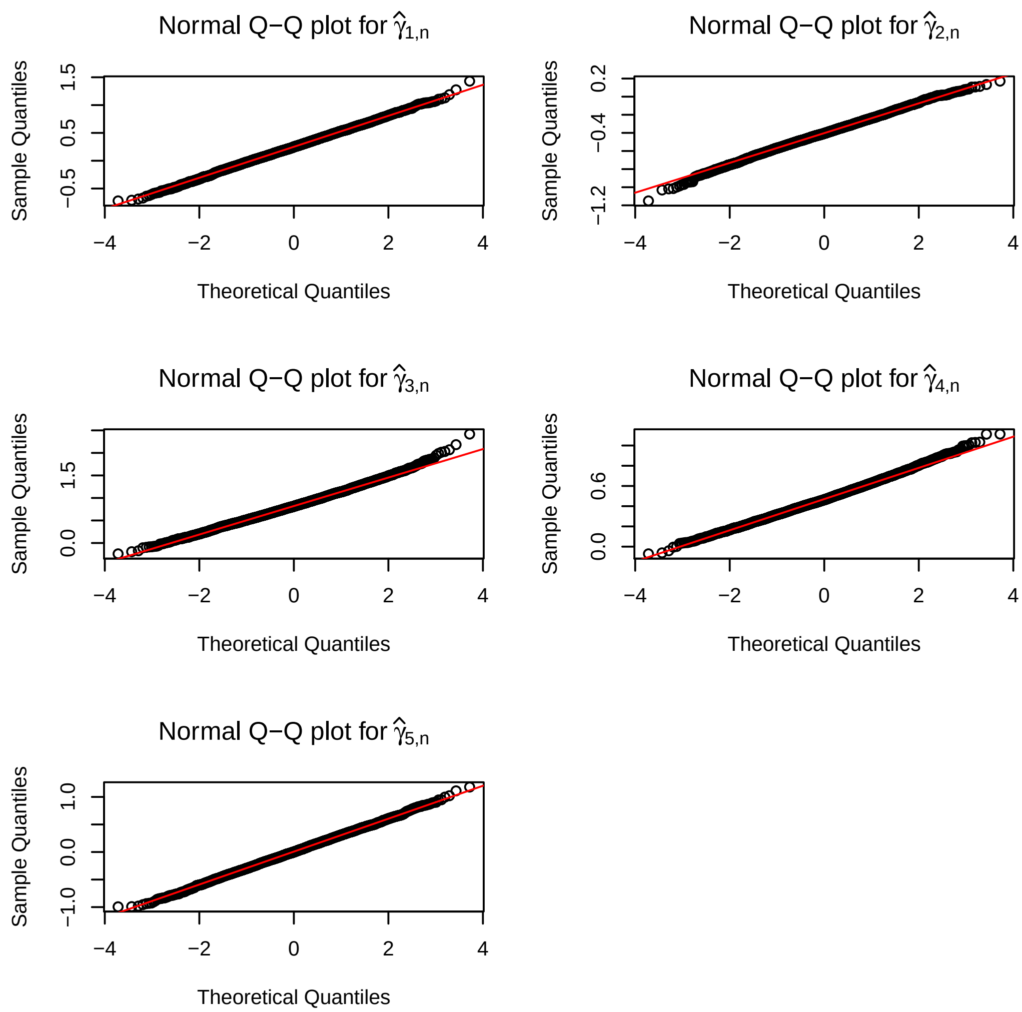

Figure 4: Normal Q-Q plots for $\hat{\gamma}_{1,n}, \ldots, \hat{\gamma}_{5,n}$ with $n = 300$ (case 2).