



**HAL**  
open science

# A constrained marginal zero-inflated binomial regression model

Essoham Ali, Aliou Diop, Jean-François Dupuy

► **To cite this version:**

Essoham Ali, Aliou Diop, Jean-François Dupuy. A constrained marginal zero-inflated binomial regression model. *Communications in Statistics - Theory and Methods*, 2022, 51 (18), pp.6396-6422. 10.1080/03610926.2020.1861296 . hal-02399725

**HAL Id: hal-02399725**

**<https://hal.science/hal-02399725>**

Submitted on 9 Dec 2019

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

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

# A constrained marginal zero-inflated binomial regression model

Essoham ALI<sup>a</sup>, Aliou DIOP<sup>a</sup>, Jean-François DUPUY<sup>b</sup>

<sup>a</sup>*University Gaston Berger, LERSTAD, Saint-Louis, Senegal*

<sup>b</sup>*Univ Rennes, INSA Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France*

---

## Abstract

Zero-inflated models have become a popular tool for assessing the relationships between explanatory variables and a zero-inflated count outcome. In these models, regression coefficients have latent class interpretations, where the latent classes correspond to a susceptible subpopulation with observations generated from a count distribution and a non-susceptible subpopulation that provides only zero counts. However, it is often of interest to evaluate covariates effects in the overall mixture population, that is, on the marginal mean of the zero-inflated count response. Marginal zero-inflated models, such as the marginal zero-inflated Poisson and negative binomial models, have been developed for that purpose. They specify independent submodels for the susceptibility probability and the marginal mean of the count response. When the count outcome is bounded, it is tempting to formulate a marginal zero-inflated binomial model in the same fashion. This, however, is not possible, due to the inherent constraints that relate, in the zero-inflated binomial model, the susceptibility probability and the latent and marginal means of the count outcome. In this paper, we propose a marginal zero-inflated binomial regression model that accommodates these constraints. We construct maximum likelihood estimates of the regression parameters. Their asymptotic properties are established and their finite-sample behaviour is examined by simulations. An application of the proposed model to the analysis of health-care demand is provided for illustration.

*Keywords:* Asymptotic properties; Count data; Excess of zeros; Health-care demand; Simulations

---

## 1. Introduction

Zero-inflated models have become a popular tool to analyze count data with excess zeros. The classical formulation of zero-inflated models assumes that the population under study is divided into two unobserved latent classes, one corresponding to a susceptible (or at-risk) subpopulation with observations generated from a non-degenerate count distribution, the other corresponding to a non-susceptible subpopulation that provides only zero counts. A

---

*Email addresses:* [aliessoham@ugb.edu.sn](mailto:aliessoham@ugb.edu.sn) (Essoham ALI), [aliou.diop@ugb.edu.sn](mailto:aliou.diop@ugb.edu.sn) (Aliou DIOP), [Jean-Francois.Dupuy@insa-rennes.fr](mailto:Jean-Francois.Dupuy@insa-rennes.fr) (Jean-François DUPUY)  
corresponding author: Jean-François Dupuy

zero-inflated model can thus be viewed as a mixture of a degenerate distribution with point mass at zero and a non-degenerate count distribution. Covariate effects can be incorporated in these distributions, which yields zero-inflated regression models. Well-known examples include the zero-inflated Poisson (ZIP) regression model (Lambert, 1992) and its extensions (e.g., Lam et al., 2006; He et al., 2010; Feng and Zhu, 2011; Monod, 2014) and the zero-inflated negative binomial (ZINB) regression model (Ridout et al., 2001; Moghimbeigi et al., 2008). When counts have an upper bound, ZIP and ZINB models are no longer appropriate and Hall (2000) introduced the zero-inflated binomial (ZIB) model, see also Diallo et al. (2019). Diallo et al. (2018) recently proposed a zero-inflated regression model for multinomial counts with joint zero-inflation, see also Dupuy (2018).

In these models, regression coefficients have latent class interpretations. However, it is often of interest to evaluate covariate effects directly on the marginal mean of the count response (that is, on the mean of the overall mixture population) instead of the mean of the susceptible population. Moreover, Preisser et al. (2012) note that several articles reporting results of latent class zero-inflated models tend to misinterpret covariate effects on the mean of the susceptible subpopulation as overall effects. Marginal effects can well be estimated from latent class zero-inflated models. This, however, necessitates additional calculations after fitting the model. Moreover, the delta method or bootstrap are needed to obtain appropriate standard errors (see Albert et al., 2014; Todem et al., 2016). As noted by Long et al. (2014), the computational tools needed for these calculations are typically not readily available in standard softwares, making this approach rather cumbersome for applied analysts. Marginal zero-inflated regression models have thus been proposed to facilitate marginal inference. These models directly relate covariates to the marginal mean of the zero-inflated count. Under this formulation, the relationship between the latent mean of the susceptible subpopulation and the covariates is implied by the assumed models for the marginal mean and latent class membership. For example, a marginal zero-inflated Poisson (MZIP) regression model was recently introduced by Long et al. (2014) and extended to include random effects (Long et al., 2015). Preisser et al. (2016) propose a marginal zero-inflated negative binomial (MZINB) regression model. Instead of modeling the Poisson and negative binomial means in the susceptible latent class, these models directly relate the marginal mean to the covariates (by using a log link). As in the latent class formulation, a logistic regression model is assumed for the susceptibility (or at-risk) probability.

Estimation and inference in MZIP and MZINB models are essentially based on same methods as those developed for latent class ZIP and ZINB models. Thus, at first sight, the extension of latent class zero-inflated models to marginal zero-inflated models may appear conceptually modest. However, in the case of bounded counts, this extension surprisingly raises some unexpected difficulties, as we explain now. The zero-inflated binomial (ZIB) regression model is defined as  $Z \sim \pi\delta_0 + (1 - \pi)\mathcal{B}(m, p)$ , where  $Z$  is the observed count,  $\pi$  is the susceptibility probability,  $\delta_0$  is the degenerate distribution at 0 and  $\mathcal{B}(m, p)$  denotes the binomial distribution with size  $m$  and latent success probability  $p$  (that is,  $p$  is the success probability for a susceptible subject). The latent class ZIB regression model proposed by Hall (2000) specifies  $\pi$  and  $p$  (using logistic regressions). On the other hand, a marginal ZIB

regression model will specify  $\pi$  and the marginal success probability  $q = p(1 - \pi)$ . In this case, the model relating  $p$  to the covariates is implied by the relation  $p = q(1 - \pi)^{-1}$ .

The relation  $q = p(1 - \pi)$  imposes the following constraints on the probabilities involved in a ZIB model (whether it is latent class or marginal) :

$$q \leq p \quad \text{and} \quad q \leq 1 - \pi \quad (\text{or equivalently } \pi \leq 1 - q). \quad (1.1)$$

In the latent class ZIB model, one specifies  $\pi$  and  $p$  and obtains  $q$  as  $q = p(1 - \pi)$ . The constraints (1.1) are thus automatically satisfied. In contrast, in a marginal ZIB model, once  $q$  has been specified,  $\pi$  has to be chosen such that  $\pi \leq 1 - q$  (then  $q \leq p$  automatically holds since  $p = q(1 - \pi)^{-1}$ ). In this paper, we propose a marginal ZIB regression model which satisfies the constraints (1.1). A logistic regression model is used for the marginal success probability  $q$ . Then, a constrained regression model is chosen for the susceptibility probability  $\pi$ . Of course, it is possible to proceed the other way around: one may first model the at-risk probability and then set a constrained model for the marginal success probability. Choosing between these two approaches depends on the aim of the analysis. If the objective is to interpret covariates effects on the marginal probability of success  $q$ , one should rather restrict the at-risk probability  $\pi$  and use, for example, a classical logistic regression for  $q$ , which will enable the analyst to use odds-ratios to interpret covariates effects. If the objective is to interpret covariates effects on the susceptibility probability, one should rather use a restricted model for  $q$ .

The paper is organized as follows. In Section 2, we briefly review the latent class ZIB regression model. Then we discuss the constraints that relate latent success, marginal success and susceptibility probabilities in a ZIB regression model, and we explain how these constraints affect the formulation of a marginal ZIB model. A novel constrained marginal ZIB regression model is finally proposed. In section 3, we investigate model identifiability and asymptotic properties of maximum likelihood estimates in this model. Section 4 report the results of a simulation study. An application of the proposed marginal ZIB regression model to the analysis of health-care demand data is described in Section 5. Some concluding remarks are given in Section 6. Technical proofs are postponed to an appendix.

## 2. Zero-inflated binomial regression

### 2.1. A brief review of the latent class ZIB regression model

Let  $Z_i$  denote the random count of interest for individual  $i$ ,  $i = 1, \dots, n$ . The ZIB model is defined as the mixture model

$$Z_i \sim \begin{cases} 0 & \text{with probability } \pi_i, \\ \mathcal{B}(m_i, p_i) & \text{with probability } 1 - \pi_i, \end{cases} \quad (2.2)$$

where  $\pi_i$  and  $p_i$  are the susceptibility and event probability respectively. To be more specific, let  $S_i$  denote the unobserved latent class indicator, that is, the random variable which is 1 if subject  $i$  is susceptible to the event of interest and 0 otherwise. Then  $p_i$  is the probability  $\mathbb{P}(Z_i = 1 | S_i = 1)$ .

**Remark 1.** The marginal success probability  $q_i := \mathbb{P}(Z_i = 1)$  is equal to :

$$\begin{aligned} q_i &= \mathbb{P}(Z_i = 1|S_i = 0)\pi_i + \mathbb{P}(Z_i = 1|S_i = 1)(1 - \pi_i) \\ &= p_i(1 - \pi_i). \end{aligned}$$

Suppose that covariates are present. The latent class ZIB regression model directly relates  $\pi_i$  and  $p_i$  to the covariates. Logit link functions are usually assumed (e.g., Hall, 2000; Diallo et al., 2019) but other links, such as probit and complementary log-log, are also applicable. Using logit links, the model writes as :

$$\text{logit}(p_i) = \beta^\top \mathbf{X}_i, \quad \text{logit}(\pi_i) = \gamma^\top \mathbf{W}_i, \quad (2.3)$$

where  $\beta$  and  $\gamma$  are vectors of unknown regression coefficients and  $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip})^\top$  and  $\mathbf{W}_i = (W_{i1}, W_{i2}, \dots, W_{iq})^\top$  are covariates vectors (they may share common components or be distinct), with  $X_{i1} = W_{i1} = 1$ .

**Remark 2.** The parameter  $\beta$  describes covariates effects on the mean count response of susceptible individuals, since  $\mathbb{E}[Z_i|\mathbf{X}_i, S_i = 1] = m_i p_i$ . Assessing covariates effects on the marginal mean of  $Z_i$  requires calculation of  $\mathbb{E}[Z_i|\mathbf{X}_i, \mathbf{W}_i] = m_i p_i (1 - \pi_i)$ , which is :

$$\mathbb{E}[Z_i|\mathbf{X}_i, \mathbf{W}_i] = \frac{m_i e^{\beta^\top \mathbf{X}_i}}{(1 + e^{\beta^\top \mathbf{X}_i})(1 + e^{\gamma^\top \mathbf{W}_i})} \quad (2.4)$$

under (2.3). We note that (2.4) involves all covariates and parameters from both submodels for zero-inflation and count response. Interpreting covariates effects on the marginal mean of the latent class ZIB regression model is thus not trivial, see Long et al. (2014) for a detailed discussion of this issue.

Suppose that we observe a sample of  $n$  independent copies  $(Z_i, \mathbf{X}_i, \mathbf{W}_i), i = 1, \dots, n$  of  $(Z, \mathbf{X}, \mathbf{W})$ . The log-likelihood of  $\theta = (\beta^\top, \gamma^\top)^\top$  in the latent class ZIB model (2.2)-(2.3) is :

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

where  $J_i := 1_{\{Z_i=0\}}$ . The maximum likelihood estimator (MLE)  $\hat{\theta}_n := (\hat{\beta}_n^\top, \hat{\gamma}_n^\top)^\top$  of  $\theta$  is obtained by solving the score equation  $\partial \ell \ell_n^{LC}(\theta) / \partial \theta = 0$ , which can be achieved with the EM algorithm (Hall, 2000) or by direct maximization of  $\ell \ell_n^{LC}(\theta)$ .

## 2.2. About inherent constraints of ZIB regression

As mentioned above, the interpretation of covariates effects on the marginal mean of a latent class ZIB model is difficult. Marginal zero-inflated regression models have therefore been proposed to facilitate marginal inference. Examples of such models include the marginal

zero-inflated Poisson (MZIP) and marginal zero-inflated negative binomial (MZINB) models. By mimicking the formulation of MZIP and MZINB models, a marginal ZIB regression model was also recently proposed (Martin and Hall, 2017). This model directly relates the marginal success probability  $q_i = p_i(1 - \pi_i)$  to the covariates, by assuming:

$$\text{logit}(q_i) = \beta^\top \mathbf{X}_i. \quad (2.5)$$

As in the latent class ZIB model, a logistic regression

$$\text{logit}(\pi_i) = \gamma^\top \mathbf{W}_i \quad (2.6)$$

is assumed for the susceptibility probability. Here,  $\beta$  directly describes covariates effects on the marginal mean  $\mathbb{E}[Z_i|\mathbf{X}_i] = m_i q_i$ . The model relating the latent mean (i.e., the count mean response for a susceptible subject) to covariates is implied by the relation  $p_i = q_i(1 - \pi_i)^{-1}$  and is used for purpose of likelihood construction, rather than for interpretation. The log-likelihood of  $(\beta, \gamma)$  in the marginal ZIB model defined by (2.2), (2.5) and (2.6) is :

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

Martin and Hall (2017) use a EM algorithm for estimation. Direct maximization of the log-likelihood is also possible, using one of the various optimisation routines available in standard softwares (such as `optim` or `maxLik` in R, R Core Team, 2018).

The formulation of a marginal ZIB regression model thus seems straightforward. However, it raises a specific - and somewhat hidden - difficulty, which was not addressed so far. The relation  $q_i = p_i(1 - \pi_i)$  imposes some constraints on the latent class success probability  $p_i$ , marginal success probability  $q_i$  and susceptibility probability  $\pi_i$ . These constraints are :

$$q_i \leq p_i \quad \text{and} \quad q_i \leq 1 - \pi_i, \quad i = 1, \dots, n, \quad (2.7)$$

and come from the fact that  $p_i$  and  $\pi_i$  are both less than 1. In the latent class ZIB model, one assumes models for  $p_i$  and  $\pi_i$  (taking care of  $0 \leq p_i, \pi_i \leq 1$ ) and (2.7) is automatically satisfied. In the marginal ZIB model, one assumes models for  $q_i$  and  $\pi_i$ . But these quantities cannot be modeled independently, since  $p_i = q_i(1 - \pi_i)^{-1}$  must be less than 1. More precisely, once  $q_i$  has been specified,  $\pi_i$  must be chosen such that  $q_i \leq 1 - \pi_i$ .

Consider the marginal ZIB model defined by (2.2), (2.5) and (2.6). It is not difficult to see that the condition  $q_i \leq 1 - \pi_i$  is equivalent to

$$\beta^\top \mathbf{X}_i + \gamma^\top \mathbf{W}_i \leq 0,$$

and this should hold for every  $i = 1, \dots, n$ , which seems highly implausible in practice. Simulation results reported in Table 1 of Martin and Hall (2017) are obtained for  $\mathbf{X}_i = \mathbf{W}_i =$

$(1, X_{i2}, X_{i3})^\top$ , where  $X_{i2}$  takes its value in  $\{0, 1\}$  and  $\beta = (\beta_1, \beta_2, \beta_3)^\top$  and  $\gamma = (\gamma_1, \gamma_2, \gamma_3)^\top$  are chosen such that  $\beta_3 = -\gamma_3$ . In this case, the condition  $\beta^\top \mathbf{X}_i + \gamma^\top \mathbf{W}_i \leq 0$  reduces to:

$$\beta_1 + \gamma_1 \leq \min(0, -(\beta_2 + \gamma_2)), \quad (2.8)$$

which is satisfied in Martin and Hall (2017) since  $\beta = (-.405, .811, -.25)^\top$  and  $\gamma = (-1.386, .539, .25)^\top$  (let  $\mathcal{C}$  denote the set of values of  $(\beta, \gamma)$  such that (2.8) is satisfied). This special case, however, masks the issue raised by constraints (2.7). For example, consider  $\gamma = (-1.386, .989, .25)^\top$  ( $\gamma_2$  above has been changed from .539 to .989,  $\gamma_1$  and  $\gamma_3$  are unchanged). This value may appear, a priori, as admissible as  $\gamma = (-1.386, .539, .25)^\top$ . However, the corresponding marginal ZIB model cannot be defined since under this value,  $q_i > 1 - \pi_i$  for some  $i = 1, \dots, n$  and the corresponding  $p_i$  are greater than 1.

Moreover, from our experience, optimisation algorithms may fail to converge in model (2.2)-(2.5)-(2.6) even when  $\beta$  and  $\gamma$  satisfy (2.8) but their values lie close to the boundary of the admissible set  $\mathcal{C}$ .

**Remark 3.** *Marginalized ZIP and ZINB models are not affected by the kind of constraint described above. For example, a ZIP model for  $Z$  is defined as  $Z \sim \pi\delta_0 + (1 - \pi)\mathcal{P}(\lambda)$ , where  $\mathcal{P}(\lambda)$  denotes Poisson distribution with mean  $\lambda > 0$ . A marginal ZIP model specifies the susceptibility probability  $\pi \in [0, 1]$  and the marginal mean  $\nu := \lambda(1 - \pi)$  of  $Z$  (see Long et al., 2014). The latent class mean  $\lambda$  is implied by the relation  $\lambda = \nu(1 - \pi)^{-1}$ . Since  $\lambda$  is unbounded, no restriction applies on  $\pi$  once  $\nu$  is specified. Therefore,  $\pi$  and  $\nu$  can be modeled independently of one another.*

### 2.3. A constrained marginal ZIB regression model

In this section, we propose a marginal ZIB regression model which accommodates the constraints described above. First, we formulate the model.

#### 2.3.1. Model formulation and estimation

Consider the ZIB model (2.2), defined as  $Z_i \sim \pi_i\delta_0 + (1 - \pi_i)\mathcal{B}(m_i, p_i)$ ,  $i = 1, \dots, n$ . In order to construct a marginal ZIB model, we first relate the marginal success probability  $q_i := \mathbb{P}(Z_i = 1 | \mathbf{X}_i)$  to the covariate  $\mathbf{X}_i$  using a logit link :

$$\text{logit}(q_i) = \beta^\top \mathbf{X}_i. \quad (2.9)$$

Then we specify the susceptibility probability  $\pi_i$ , which must satisfy  $\pi_i \leq 1 - q_i = (1 + e^{\beta^\top \mathbf{X}_i})^{-1}$ . Our proposal is to set :

$$\pi_i = \frac{1}{1 + e^{\beta^\top \mathbf{X}_i} + e^{\gamma^\top \mathbf{X}_i}}, \quad (2.10)$$

where  $\beta$  and  $\gamma$  are  $p$ -dimensional unknown vectors of regression coefficients. Note that here, we do not distinguish between covariates that act on the marginal success probability and covariates that act on the susceptibility probability. All covariates are included in both submodels (under the common notation  $\mathbf{X}_i$ ) and Wald tests will be used to identify the ones which significantly influence each process.

**Remark 4.** *The intuition behind (2.10) is as follows. We need that  $\pi_i \leq 1 - q_i = (1 + e^{\beta^\top \mathbf{X}_i})^{-1}$  or equivalently :  $1 + e^{\beta^\top \mathbf{X}_i} \leq \pi_i^{-1}$ . If  $\mathbf{X}_i$  is bounded and we assume that  $\beta$  belongs to some compact set of  $\mathbb{R}^p$ , there exists a finite constant  $M$  such that  $1 + e^{\beta^\top \mathbf{X}_i} \leq M$ , and  $\pi_i = M^{-1}$  provides a mathematically admissible model for  $\pi_i$ . This, however, is not satisfactory since the susceptibility probability is likely to depend on  $i$ , through covariates. Thus, a genuine upper bound for  $\pi_i$  should depend in some way on  $1 + e^{\beta^\top \mathbf{X}_i}$ .*

*A more refined upper bound for  $1 + e^{\beta^\top \mathbf{X}_i}$  is given by  $1 + a \cdot e^{\beta^\top \mathbf{X}_i}$  with  $a > 1$ , or equivalently by  $1 + e^{\beta^\top \mathbf{X}_i}(1 + e^\alpha)$ , where  $\alpha$  is some finite real number (the parameterization  $1 + e^\alpha$  is convenient since it ensures that  $1 + e^\alpha > 1$  for every  $\alpha$ ). Now,*

$$\begin{aligned} 1 + e^{\beta^\top \mathbf{X}_i}(1 + e^\alpha) &= 1 + e^{\beta^\top \mathbf{X}_i} + e^{\alpha + \beta^\top \mathbf{X}_i}, \\ &:= 1 + e^{\beta^\top \mathbf{X}_i} + e^{\tilde{\beta}^\top \mathbf{X}_i}, \end{aligned}$$

*where  $\tilde{\beta}$  coincides with  $\beta = (\beta_1, \dots, \beta_p)$ , except its first component which is given by  $\alpha + \beta_1$ . We have  $(1 + e^{\beta^\top \mathbf{X}_i} + e^{\tilde{\beta}^\top \mathbf{X}_i})^{-1} \leq 1 - q_i$ . Therefore, an alternative and rather general model for  $\pi_i$ , satisfying the constraint  $\pi_i \leq 1 - q_i$ , is given by  $\pi_i = (1 + e^{\beta^\top \mathbf{X}_i} + e^{\tilde{\beta}^\top \mathbf{X}_i})^{-1}$ . Finally, in order to further increase model flexibility, we set  $\pi_i = (1 + e^{\beta^\top \mathbf{X}_i} + e^{\gamma^\top \mathbf{X}_i})^{-1}$ , where  $\gamma$  is a  $p$ -dimensional vector possibly unrelated to  $\beta$ .*

The proposed marginal ZIB regression model is thus defined by equations (2.2), (2.9) and (2.10). The latent class success probability implied by this model is :

$$p_i = \frac{e^{\beta^\top \mathbf{X}_i}(1 + e^{\beta^\top \mathbf{X}_i} + e^{\gamma^\top \mathbf{X}_i})}{(1 + e^{\beta^\top \mathbf{X}_i})(e^{\beta^\top \mathbf{X}_i} + e^{\gamma^\top \mathbf{X}_i})}.$$

**Remark 5.** *The interpretation of covariates effects on  $p_i$  is not trivial. However, in a marginal zero-inflated model, the latent class success probability is of secondary interest since one is primarily interested in the marginal success probability  $q_i$ .*

Let  $\theta = (\beta^\top, \gamma^\top)^\top$  be the whole  $2p$ -dimensional parameter in the marginal ZIB regression model (2.2)-(2.9)-(2.10). For  $i = 1, \dots, n$ , let  $J_i = 1_{\{Z_i=0\}}$  and

$$f_i(\theta) = 1 + e^{\beta^\top \mathbf{X}_i} + e^{\gamma^\top \mathbf{X}_i}, \quad g_i(\theta) = e^{\beta^\top \mathbf{X}_i} + e^{\gamma^\top \mathbf{X}_i}, \quad h_i(\theta) = \frac{e^{\gamma^\top \mathbf{X}_i}}{(1 + e^{\beta^\top \mathbf{X}_i})g_i(\theta)}.$$

Then, the log-likelihood of  $\theta$ , based on  $n$  independent observations  $(Z_i, \mathbf{X}_i), i = 1, \dots, n$ , is :

$$\begin{aligned} \ell \ell_n(\theta) &= \sum_{i=1}^n J_i \log(1 + g_i(\theta)h_i^{m_i}(\theta)) - \sum_{i=1}^n \log(f_i(\theta)) \\ &\quad + \sum_{i=1}^n (1 - J_i) [Z_i(\beta - \gamma)^\top \mathbf{X}_i + \log(g_i(\theta)) + Z_i \log(f_i(\theta)) + m_i \log(h_i(\theta))], \\ &= \sum_{i=1}^n \ell_i(\theta). \end{aligned}$$

The MLE  $\hat{\theta}_n = (\hat{\beta}_n^\top, \hat{\gamma}_n^\top)^\top$  of  $\theta$  is obtained by solving the score equation  $\dot{\ell}_n(\theta) = 0$ , where

$$\dot{\ell}_n(\theta) = \frac{1}{\sqrt{n}} \frac{\partial \ell_n(\theta)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial \ell_i(\theta)}{\partial \theta}. \quad (2.11)$$

In Section 3, we establish consistency and asymptotic normality of  $\hat{\theta}_n$ . Before this, we introduce some further notations.

### 2.3.2. Notations

Let

$$\mathbb{X} = \begin{pmatrix} 1 & \cdots & 1 \\ X_{12} & \cdots & X_{n2} \\ \vdots & \ddots & \vdots \\ X_{1p} & \cdots & X_{np} \end{pmatrix}$$

be the  $(p \times n)$  design matrix and  $\mathbb{V}$  be the  $(2p \times 2n)$  block-matrix defined as :

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

where  $0_{p,n}$  denotes the  $(p \times n)$  matrix whose components are all equal to zero. Let also  $S(\theta) = (S_j(\theta))_{1 \leq j \leq 2n}$  be the  $2n$ -dimensional column vector defined by :

$$S(\theta) = (\dot{\ell}_{\beta,1}(\theta), \dots, \dot{\ell}_{\beta,n}(\theta), \dot{\ell}_{\gamma,1}(\theta), \dots, \dot{\ell}_{\gamma,n}(\theta))^\top,$$

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

$$\dot{\ell}_{\beta,i}(\theta) = -J_i \frac{e^{\beta^\top \mathbf{X}_i} Q_i(\theta)}{(1 + e^{\beta^\top \mathbf{X}_i}) g_i(\theta) k_i(\theta)} + Z_i \left( \frac{e^{\beta^\top \mathbf{X}_i}}{f_i(\theta)} + 1 \right) - \frac{e^{\beta^\top \mathbf{X}_i}}{f_i(\theta)} + \frac{e^{\beta^\top \mathbf{X}_i} Q_i(\theta)}{(1 + e^{\beta^\top \mathbf{X}_i}) g_i(\theta)},$$

and

$$\dot{\ell}_{\gamma,i}(\theta) = -J_i \frac{m_i e^{\beta^\top \mathbf{X}_i} + e^{\gamma^\top \mathbf{X}_i}}{g_i(\theta) k_i(\theta)} + Z_i \left( \frac{e^{\gamma^\top \mathbf{X}_i}}{f_i(\theta)} - 1 \right) - \frac{e^{\gamma^\top \mathbf{X}_i}}{f_i(\theta)} + \frac{m_i e^{\beta^\top \mathbf{X}_i} + e^{\gamma^\top \mathbf{X}_i}}{g_i(\theta)},$$

with  $k_i(\theta) = 1 + g_i(\theta) h_i^{m_i}(\theta)$  and  $Q_i(\theta) = 1 + e^{\beta^\top \mathbf{X}_i} - m_i(f_i(\theta) + e^{\beta^\top \mathbf{X}_i})$ . Finally, if  $A = (A_{ij})_{1 \leq i \leq a, 1 \leq j \leq b}$  is a  $(a \times b)$  matrix, we let  $A_{\bullet j}$  denote its  $j$ -th column ( $j = 1, \dots, b$ ) that is,  $A_{\bullet j} = (A_{1j}, \dots, A_{aj})^\top$ . Then the score vector  $\dot{\ell}_n(\theta)$  can be written as  $\dot{\ell}_n(\theta) = \frac{1}{\sqrt{n}} \sum_{j=1}^{2n} \mathbb{V}_{\bullet j} S_j(\theta)$ .

### 3. Model identifiability and asymptotic results

We start by giving some regularity conditions that are needed to ensure identifiability of the proposed model and to establish large-sample properties of the MLE.

1. There exists a finite positive constant  $c_1$  such that  $|X_{ij}| \leq c_1$  for every  $i = 1, 2, \dots$  and  $j = 1, \dots, p$ . For every  $i = 1, 2, \dots$  and  $j = 2, \dots, p$ ,  $\text{var}[X_{ij}] > 0$ . For every  $i = 1, 2, \dots$  the  $X_{ij}$  ( $j = 1, \dots, p$ ) are linearly independent.
2. The true parameter values  $\beta_0$  and  $\gamma_0$  belong to the interior of a known compact set  $\mathcal{C}$  of  $\mathbb{R}^p$ .
3. As  $n \rightarrow \infty$ ,  $n^{-1} \sum_{i=1}^n \mathbb{E} \left[ \frac{\partial^2 \ell_i(\theta)}{\partial \theta \partial \theta^\top} \right]$  converges to some invertible matrix  $L(\theta)$  and the smallest eigenvalue  $\lambda_n$  of  $\mathbb{V}\mathbb{V}^\top$  tends to  $+\infty$ .
4. For every  $i = 1, \dots, n$ , we have :  $m_i \in \{2, \dots, M\}$  for some finite integer  $M$ .

In what follows, all random variables  $Z_i$  and  $\mathbf{X}_i$ ,  $i = 1, 2, \dots$  are defined on a same probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . The space  $\mathbb{R}^{2p}$  is equipped with the Euclidean norm  $\|\cdot\|_2$  and the space of  $(2p \times 2p)$  real matrices is equipped with the norm  $\max_{\|x\|_2=1} \|Ax\|_2$  (for notations simplicity, we use  $\|\cdot\|$  for both norms). We are now in position to state our first result:

**Theorem 3.1.** *Assume that conditions 1-4 hold. Then the marginal ZIB regression model defined by (2.2), (2.9) and (2.10) is identifiable, that is,  $\ell_i(\theta) = \ell_i(\theta^*)$  almost surely implies  $\theta = \theta^*$ .*

The proof is given in Appendix A.

Asymptotic properties of  $\hat{\theta}_n$  are now presented in the following theorems. Here, we present a sketch of the proofs ; details are provided in Appendix B.

**Theorem 3.2.** *Assume that conditions 1-4 hold. Then, as  $n \rightarrow \infty$ ,  $\hat{\theta}_n$  converges almost surely to  $\theta_0 = (\beta_0^\top, \gamma_0^\top)^\top$ .*

The proof hinges on the fact that  $n^{-1/2} \dot{\ell}_n(\theta)$  converges to a limit which is null at  $\theta_0$ . Then we appeal to the inverse function theorem of Foutz (1977).

**Theorem 3.3.** *Assume that conditions 1-4 hold. Then  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  is asymptotically distributed as a multivariate normal with mean-zero and covariance matrix  $L(\theta_0)^{-1}$ . A consistent estimator of the asymptotic variance is given by  $(n^{-1/2} \partial \dot{\ell}_n(\hat{\theta}_n) / \partial \theta^\top)^{-1}$ .*

To prove asymptotic normality, we show that  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  is asymptotically equivalent to a linear combination of random (non identically distributed) vectors and we apply a multivariate central limit theorem for random linear vector forms of Eicker (1966). More precisely, a Taylor series expansion of  $\dot{\ell}_n(\hat{\theta}_n)$  at  $\theta_0$  yields

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = - \left( \frac{1}{\sqrt{n}} \frac{\partial \dot{\ell}_n(\theta_0)}{\partial \theta^\top} \right)^{-1} \frac{1}{\sqrt{n}} \sum_{j=1}^{2n} \mathbb{V}_{\bullet, j} S_j(\theta_0) + o_{\mathbb{P}}(1).$$

By Lemma 6.3 in Appendix B,  $n^{-1/2}\partial\dot{\ell}_n(\theta_0)/\partial\theta^\top$  converges to  $L(\theta_0)$ , and we show that  $\frac{1}{\sqrt{n}}\sum_{j=1}^{2n}\mathbb{V}_{\bullet_j}S_j(\theta_0)$  is asymptotically distributed as a multivariate normal with mean zero and covariance matrix  $L(\theta_0)$ . An application of Slutsky's theorem concludes the proof. All details are given in Appendix B, along with expressions of the terms of the matrix of second derivatives  $\partial\dot{\ell}_n(\theta_0)/\partial\theta^\top$ .

#### 4. A simulation study

In this section, we evaluate the finite-sample performance of the MLE via Monte Carlo experiments.

##### 4.1. Simulation experiments

The count data  $Z_i$  are simulated from the marginal ZIB regression model (2.2)-(2.9)-(2.10) with  $\mathbf{X}_i = (X_{i1}, \dots, X_{i8})$  and  $X_{i1} = 1$ . Covariates  $X_{i2}$  to  $X_{i8}$  are generated from the standard normal distribution, uniform distribution on  $[-2, 2]$ , normal distribution with mean 1 and standard deviation 1.5, normal distribution with mean -2 and standard deviation 1, Bernoulli distribution with probability 0.3, uniform distribution on  $[1, 3]$  and Bernoulli distribution with probability 0.8 respectively. Here, we take  $\beta_0 = (-0.25, -0.9, 0.3, -0.8, 1, 0, 0, 0)^\top$ , that is,  $X_{i6}$ ,  $X_{i7}$  and  $X_{i8}$  have no influence on the marginal success probability  $q_i$ .

We consider successively four values for  $\gamma$ , namely :  $\gamma = (-0.7, 0, -0.4, 0.6, -0.5, 0.7, 0, 0)^\top$ ,  $\gamma = (-0.4, 0, 0.8, -0.6, -0.5, -0.7, 0, 0)^\top$ ,  $\gamma = (-0.4, 0, 0.8, -0.6, -0.5, -0.7, -0.7, -0.8)$  and  $\gamma = (-0.9, 0, 0.8, -0.6, 0.5, -0.7, -0.7, -0.8)$ . With these values, the average proportion  $c$  of zero-inflated data in the simulated data sets is 0.25, 0.50, 0.75 and 0.90 respectively. We consider the following sample sizes:  $n = 500, 2000$ . For each individual  $i$ , the binomial size  $m_i$  is taken as the maximum between 2 and a random draw from the binomial distribution  $\mathcal{B}(15, 0.3)$ .

We simulate  $N = 1000$  replications for each combination [**sample size**  $\times$  **proportion of zero-inflation**] of the design parameters. Simulations are conducted using the statistical software R (R Core Team, 2018). We use the package `maxLik` (Henningsen and Toomet, 2011) to solve the score equation (2.11) via a Newton-Raphson algorithm.

##### 4.2. Results

For each simulation scenario and each estimator  $\hat{\beta}_{j,n}$  ( $j = 1, \dots, 8$ ) and  $\hat{\gamma}_{k,n}$  ( $k = 1, \dots, 8$ ), we calculate the average bias, standard deviation, average standard error and root mean square error of the estimate over the  $N$  simulated samples. We also obtain the empirical coverage probability and average length of 95%-level Wald confidence intervals for the  $\beta_j$  and  $\gamma_k$ . Tables 1 and 2 report results for  $n = 500$  and  $n = 2000$  respectively.

We also assess the normal approximation stated by Theorem 3.3, by plotting estimated densities obtained from the  $N$  normalized estimates  $(\hat{\beta}_{j,n} - \beta_j)/\text{standard error}(\hat{\beta}_{j,n})$  and  $(\hat{\gamma}_{k,n} - \gamma_k)/\text{standard error}(\hat{\gamma}_{k,n})$ ,  $j, k = 1, \dots, 8$ , and by comparing with the density of the standard normal distribution. Standard errors are obtained as the square roots of the diagonal elements of the estimated variance matrix  $(n^{-1/2}\partial\dot{\ell}_n(\hat{\theta}_n)/\partial\theta^\top)^{-1}$ . Figures 1 and 2

provide results for  $n = 500$  and 50% of zero-inflation (plots for the other scenarios are similar and thus are not given).

From these results, it appears, as expected, that the bias, variability and length of confidence intervals of all estimates decrease as the sample size increases. For fixed  $n$ , we observe that : *i*) performances of the  $\hat{\beta}_{j,n}$ s remain stable when the proportion of zero-inflation varies from small to moderate values (here, from 0.25 to 0.50) and deteriorate when zero-inflation achieves higher values, and : *ii*) performances of the  $\hat{\gamma}_{k,n}$ s improve and then deteriorate as zero-inflation increases.

These observations illustrate the general fact that accurate estimation in a zero-inflated regression model requires a balance between susceptible and non-susceptible subpopulations (that is, a sufficient amount of zero and non-zero observations should be available to accurately estimate the zero-inflation probabilities and count submodel). Also, empirical coverage probabilities are close to the nominal level, which indicates that the normal approximation of the distribution of the MLE is appropriate, even when the sample size is moderate. This is confirmed by Figures 1 and 2.

## 5. Application

### 5.1. Data description and modelling

In this section, we describe an application of the marginal ZIB regression model to the analysis of health-care utilization by elderlies in the United States. We use data from the National Medical Expenditure Survey (NMES) conducted in 1987-1988 in the United States. This survey provides a comprehensive picture of how Americans (aged 66 years and over) use health services. Several measures of health-care utilization were reported in this study, such as the number of visits to a non-physician health professional in an office setting and the number of visits to a physician in an office setting. Informations on patients health are also reported, along with sociodemographic and economic variables. A detailed description of the data can be found in Deb and Trivedi (1997).

Here, we address the issue of identifying factors that determine patients decision to consult a non-physician health professional when visiting in an office setting. Let  $Z_i$  and  $m_i$  be respectively the number of non-physician office visits and the total number of office visits for patient  $i$ . Given  $m_i$ , one may model  $Z_i$  by a  $\mathcal{B}(m_i, \pi_i)$  distribution (in view of condition 4, we consider patients with at least two office visits over the study period, that is,  $m_i \geq 2$ ). However, the high frequency of zero in the  $Z_i$ s (60.2% among the  $n = 3391$  patients included in our analysis) suggests that the response is zero-inflated and that a ZIB regression model should be used instead. A latent class ZIB model will only provide insights into the determinants of patients decision in the unobservable class of at-risk patients (that is, patients who have not abdicated from consulting non-physicians). More meaningful interpretations can be obtained at the population level, and for this, we suggest to use a marginal ZIB regression model.

As mentioned above, available covariates include: *i*) socio-economic variables: gender (1 for female, 0 for male), age (in years, divided by 10), marital status, educational level (number of years of education), income, *ii*) various measures of health condition: number

of chronic conditions (cancer, arthritis, gallbladder problems. . . ) 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).

We fit a marginal ZIB regression model to the data, where the marginal success probability  $q_i$  and susceptibility probability  $\pi_i$  are given by (2.9) and (2.10) respectively, and  $\mathbf{X}_i$  includes all available covariates listed above. We use Wald tests to select significant covariates (the least significant covariate - at the level 5% - is removed and the model is fitted again, until all remaining covariates are significant ; note that the BIC criterion decreases at each step of this procedure). Table 3 presents the final marginal ZIB model.

## 5.2. Results

Most influencing factors of the decision of never resorting to a non-physician are health condition, age, gender, educational level and medicaid status. The probability of never resorting increases when health condition degrades. One hypothesis is that patients whose health declines will tend to favor visits to a physician over visits to a non-physician. The probability of never resorting to a non-physician decreases when the number of years of education increases. This is coherent with previous findings in the literature. Given the insignificance of income, this observation could be due to an income effect signalled through educational level. An alternative explanation, hypothesized by Deb and Trivedi (1997) is that education may make individuals more informed consumers of medical care services. More informed patients may in turn tend to diversify their health-care utilization. Men are more likely than women to renounce non-physician office visits. Similar observation has been reported several times in the literature, and may be explained by social codes and stereotype influence (such as masculinity being associated to resistance to harm). Finally, medicaid beneficiaries are more likely to renounce non-physician office visits. One explanation is that medicaid recipients, who have low income, may limit their consultations to those necessary, that is, to physician visits.

In the latent class ZIB model, covariates effects on the probability of consulting a non-physician when visiting in an office setting must be interpreted with respect to the susceptible subpopulation. This subpopulation being unobserved, these effects are difficult to interpret. On the contrary, the marginal ZIB regression model allows population-wide interpretations. From Table 3, we observe that in the overall population, significant determinants of the decision to consult a non-physician when visiting in an office setting include health status, age, gender, educational level and medicaid status. Patients with poor health will favor office visits to a physician over office visits to a non-physician, which seems a natural observation. Women and people with higher education have higher probability to consult a non-physician, while medicaid recipients are more likely to visit physicians than non-physicians. The probability of visiting a non-physician when consulting in an office setting decreases with age. This may be due to several factors, such as decreasing mobility associated with ageing (aged

patients will tend to limit their consultations to those considered as the most necessary, that is, to physician visits) and worsening of the health condition with ageing (patients whose health declines are likely to favor visits to a physician).

## 6. Concluding Remarks

Zero-inflated regression models provide a useful framework for analyzing count data with excess zeros. Most classical zero-inflated models are based on a latent class formulation. In this setting, regression parameters in the count submodel must be interpreted with respect to the class of susceptible individuals, which often yields misleading, or even erroneous, statements. For example, Preisser et al. (2012) observe that in dental epidemiology, many researchers misinterpret parameters in zero-inflated Poisson regression models in terms of overall caries incidence, instead of incidence within the susceptible subpopulation.

Marginal zero-inflated models were recently proposed to allow population-wide inference from count data with excess zeros. This approach directly models the marginal mean of the count response, rather than the conditional mean given that the individual is susceptible. For example, marginal zero-inflated Poisson and negative binomial models were introduced by Long et al. (2014) and Preisser et al. (2016) respectively. The formulation of these models is relatively straightforward. On the contrary, formulating a marginal zero-inflated binomial model requires some care, owing to the inherent constraints which apply, in this case, to the susceptibility probability and marginal mean of the count response. In this paper, we propose a formulation that takes these constraints into account. Asymptotic theory and simulations studies both suggest the appropriateness of maximum likelihood inference in this new model.

Several issues now deserve attention and the proposed marginal ZIB model should be extended to accommodate additional data complexity. For example, random effects could be incorporated to the model, in order to take account of correlation among the individuals. Non-linear effects may also be introduced in the linear predictors, through unknown functions of the covariates. These extensions require specific theoretical and numerical developments that are the topics for our future work.

## Appendix A : proof of Theorem 3.1

Assume that  $\ell_i(\theta) = \ell_i(\theta^*)$  almost surely and note that under  $\theta$ ,

$$\begin{aligned}
 \mathbb{P}(J_i = 0 | \mathbf{X}_i) &= 1 - \mathbb{P}(Z_i = 0 | \mathbf{X}_i) \\
 &= 1 - (\pi_i + (1 - \pi_i)(1 - p_i)^{m_i}) \\
 &= 1 - \frac{1}{f_i(\theta)} - \frac{g_i(\theta)}{f_i(\theta)} \left( \frac{e^{\gamma^\top \mathbf{X}_i}}{(1 + e^{\beta^\top \mathbf{X}_i})g_i(\theta)} \right)^{m_i} \\
 &= 1 - \frac{k_i(\theta)}{f_i(\theta)}.
 \end{aligned}$$

Under conditions 1 and 2, it is not difficult to see that  $0 < \mathbb{P}(J_i = 0 | \mathbf{X}_i) < 1$  for every  $\mathbf{x}_i$  and  $\theta$ . Thus, there exists a  $\omega \in \Omega$ , with  $\omega$  outside the negligible set where  $\ell_i(\theta) \neq \ell_i(\theta^*)$ , such that  $J_i(\omega) = 0$ . For this  $\omega$ , the equality  $\ell_i(\theta) = \ell_i(\theta^*)$  reduces to :

$$\begin{aligned} z_i(\beta - \gamma)^\top \mathbf{x}_i + \log(g_i(\theta)) + (z_i - 1) \log(f_i(\theta)) + m_i \log(h_i(\theta)) \\ = z_i(\beta^* - \gamma^*)^\top \mathbf{x}_i + \log(g_i(\theta^*)) + (z_i - 1) \log(f_i(\theta^*)) + m_i \log(h_i(\theta^*)), \end{aligned}$$

or equivalently, to :

$$z_i(\beta - \gamma - \beta^* + \gamma^*)^\top \mathbf{x}_i + (z_i - 1) \log \frac{f_i(\theta)}{f_i(\theta^*)} = \log \frac{g_i(\theta^*)}{g_i(\theta)} + m_i \log \frac{h_i(\theta^*)}{h_i(\theta)}. \quad (6.12)$$

The right-hand side of (6.12) does not depend on  $z_i$  and thus is constant for distinct positive values of  $z_i$ . Consider, for example,  $z_i = z$  and  $z_i = z + 1$  (by condition 4,  $z_i$  can take at least two distinct positive values). We obtain :

$$(\beta - \gamma - \beta^* + \gamma^*)^\top \mathbf{x}_i = -\log \frac{f_i(\theta)}{f_i(\theta^*)}. \quad (6.13)$$

Assume that  $\log(f_i(\theta)/f_i(\theta^*)) \neq 0$  and differentiate (6.13) with respect to  $\beta$ . We obtain  $\mathbf{x}_i(1 + e^{\beta^\top \mathbf{x}_i}/f_i(\theta)) = 0$ . Condition 1 implies that  $1 + e^{\beta^\top \mathbf{x}_i}/f_i(\theta) = 0$ , which is a contradiction since  $f_i(\theta) > 0$ . Thus  $f_i(\theta) = f_i(\theta^*)$ , which implies that  $(\beta - \gamma - \beta^* + \gamma^*)^\top \mathbf{x}_i = 0$ . Linear independence of the  $x_{i1}, \dots, x_{ip}$  (condition 1) implies that  $\beta - \gamma = \beta^* - \gamma^*$ . Now,  $f_i(\theta) = f_i(\theta^*)$  is equivalent to  $e^{\beta^\top \mathbf{x}_i} + e^{\gamma^\top \mathbf{x}_i} = e^{\beta^* \top \mathbf{x}_i} + e^{\gamma^* \top \mathbf{x}_i}$ , which implies :

$$e^{\gamma^\top \mathbf{x}_i} \left( e^{(\beta - \gamma)^\top \mathbf{x}_i} + 1 \right) = e^{\gamma^* \top \mathbf{x}_i} \left( e^{(\beta^* - \gamma^*)^\top \mathbf{x}_i} + 1 \right),$$

and finally,  $(\gamma - \gamma^*)^\top \mathbf{x}_i = 0$ . Condition 1 implies that  $\gamma = \gamma^*$ , which in turn implies that  $\beta = \beta^*$ . We have proved that  $\theta = \theta^*$ , which concludes the proof.  $\square$

## Appendix B : proofs of asymptotic results

**Proof of Theorem 3.2.** To prove the consistency of  $\hat{\theta}_n$ , we verify the conditions of the inverse function theorem of Foutz (1977). These conditions are proved in a series of technical lemmas.

**Lemma 6.1.**  $\partial \dot{\ell}_n(\theta) / \partial \theta^\top$  exists and is continuous in an open neighborhood of  $\theta_0$ .

**Proof of Lemma 6.1.** The  $\ell_i(\theta), i = 1, \dots, n$  are twice differentiable with respect to  $\theta$ . Continuity of  $\partial^2 \ell_i(\theta) / \partial \theta \partial \theta^\top$  is straightforward and is omitted (the expression of  $\partial^2 \ell_i(\theta) / \partial \theta_\ell \partial \theta_j$  is given by (6.18), see below).  $\square$

**Lemma 6.2.** As  $n \rightarrow \infty$ ,  $n^{-1/2} \dot{\ell}_n(\theta_0)$  converges almost surely to 0.

**Proof of Lemma 6.2.** To prove this lemma, we verify the conditions of the strong law of large numbers - SLLN (see Jiang, 2010, Theorem 6.7). First, some simple algebra yields :

$$n^{-1/2} \ell \dot{\ell}_n(\theta_0) = \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n X_{i1} \dot{\ell}_{\beta,i}(\theta_0) \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n X_{ip} \dot{\ell}_{\beta,i}(\theta_0) \\ \frac{1}{n} \sum_{i=1}^n X_{i1} \dot{\ell}_{\gamma,i}(\theta_0) \\ \vdots \\ \frac{1}{n} \sum_{i=1}^n X_{ip} \dot{\ell}_{\gamma,i}(\theta_0) \end{pmatrix}.$$

We start by proving that  $\mathbb{E}[X_{ij} \dot{\ell}_{\gamma,i}(\theta_0)] = 0$  for every  $i = 1, \dots, n$  and  $j = 1, \dots, p$ . To see this, note that :

$$\mathbb{E} \left[ X_{ij} \dot{\ell}_{\gamma,i}(\theta_0) \right] = \mathbb{E} \left[ X_{ij} \mathbb{E} \left[ \dot{\ell}_{\gamma,i}(\theta_0) | \mathbf{X}_i \right] \right],$$

and

$$\begin{aligned} \mathbb{E} \left[ \dot{\ell}_{\gamma,i}(\theta_0) | \mathbf{X}_i \right] &= -\mathbb{E} [J_i | \mathbf{X}_i] \frac{m_i e^{\beta_0^\top \mathbf{X}_i} + e^{\gamma_0^\top \mathbf{X}_i}}{g_i(\theta_0) k_i(\theta_0)} + \mathbb{E} [Z_i | \mathbf{X}_i] \left( \frac{e^{\gamma_0^\top \mathbf{X}_i}}{f_i(\theta_0)} - 1 \right) \\ &\quad - \frac{e^{\gamma_0^\top \mathbf{X}_i}}{f_i(\theta_0)} + \frac{m_i e^{\beta_0^\top \mathbf{X}_i} + e^{\gamma_0^\top \mathbf{X}_i}}{g_i(\theta_0)}. \end{aligned} \quad (6.14)$$

Now,

$$\begin{aligned} \mathbb{E} [J_i | \mathbf{X}_i] &= \mathbb{P}(Z_i = 0 | \mathbf{X}_i) \\ &= \frac{k_i(\theta_0)}{f_i(\theta_0)}, \end{aligned} \quad (6.15)$$

and

$$\begin{aligned} \mathbb{E} [Z_i | \mathbf{X}_i] &= m_i p_i (1 - \pi_i) \\ &= m_i \frac{e^{\beta_0^\top \mathbf{X}_i}}{1 + e^{\beta_0^\top \mathbf{X}_i}}. \end{aligned} \quad (6.16)$$

We plug (6.15) and (6.16) in (6.14) and we obtain :

$$\begin{aligned} \mathbb{E} \left[ \dot{\ell}_{\gamma,i}(\theta_0) | \mathbf{X}_i \right] &= -\frac{m_i e^{\beta_0^\top \mathbf{X}_i} + e^{\gamma_0^\top \mathbf{X}_i}}{g_i(\theta_0) k_i(\theta_0)} \frac{k_i(\theta_0)}{f_i(\theta_0)} + m_i \frac{e^{\beta_0^\top \mathbf{X}_i}}{1 + e^{\beta_0^\top \mathbf{X}_i}} \left( \frac{e^{\gamma_0^\top \mathbf{X}_i}}{f_i(\theta_0)} - 1 \right) \\ &\quad - \frac{e^{\gamma_0^\top \mathbf{X}_i}}{f_i(\theta_0)} + \frac{m_i e^{\beta_0^\top \mathbf{X}_i} + e^{\gamma_0^\top \mathbf{X}_i}}{g_i(\theta_0)} \frac{f_i(\theta_0)}{f_i(\theta_0)} \\ &= \frac{m_i e^{\beta_0^\top \mathbf{X}_i} + e^{\gamma_0^\top \mathbf{X}_i}}{g_i(\theta_0)} \frac{g_i(\theta_0)}{f_i(\theta_0)} - m_i \frac{e^{\beta_0^\top \mathbf{X}_i}}{1 + e^{\beta_0^\top \mathbf{X}_i}} \left( \frac{1 + e^{\beta_0^\top \mathbf{X}_i}}{f_i(\theta_0)} \right) - \frac{e^{\gamma_0^\top \mathbf{X}_i}}{f_i(\theta_0)} \\ &= 0, \end{aligned}$$

which in turn implies :

$$\mathbb{E} \left[ X_{ij} \dot{\ell}_{\gamma,i}(\theta_0) \right] = 0.$$

Now, for every  $i = 1, \dots, n$  and  $j = 1, \dots, p$ , we have :

$$\begin{aligned} \text{var} \left( X_{ij} \dot{\ell}_{\gamma,i}(\theta_0) \right) &= \mathbb{E} \left[ \text{var} \left( X_{ij} \dot{\ell}_{\gamma,i}(\theta_0) | \mathbf{X}_i \right) \right] + \text{var} \left( \mathbb{E} \left[ X_{ij} \dot{\ell}_{\gamma,i}(\theta_0) | \mathbf{X}_i \right] \right) \\ &= \mathbb{E} \left[ X_{ij}^2 \text{var} \left( \dot{\ell}_{\gamma,i}(\theta_0) | \mathbf{X}_i \right) \right] \\ &\leq c_1^2 \mathbb{E} \left[ \text{var} \left( \dot{\ell}_{\gamma,i}(\theta_0) | \mathbf{X}_i \right) \right] \quad (\text{by condition 1}). \end{aligned}$$

It is not difficult to check that  $\mathbb{E}[\text{var}(\dot{\ell}_{\gamma,i}(\theta_0)|\mathbf{X}_i)]$  is bounded. To see this, note that :

$$\begin{aligned} \text{var} \left( \dot{\ell}_{\gamma,i}(\theta_0) | \mathbf{X}_i \right) &= \left( \frac{m_i e^{\beta_0^\top \mathbf{X}_i} + e^{\gamma_0^\top \mathbf{X}_i}}{g_i(\theta_0) k_i(\theta_0)} \right)^2 \text{var} (J_i | \mathbf{X}_i) + \left( \frac{e^{\gamma_0^\top \mathbf{X}_i}}{f_i(\theta_0)} - 1 \right)^2 \text{var} (Z_i | \mathbf{X}_i) \\ &\quad + 2 \left( \frac{m_i e^{\beta_0^\top \mathbf{X}_i} + e^{\gamma_0^\top \mathbf{X}_i}}{g_i(\theta_0) k_i(\theta_0)} \right) \left( \frac{e^{\gamma_0^\top \mathbf{X}_i}}{f_i(\theta_0)} - 1 \right) \mathbb{E} [J_i | \mathbf{X}_i] \mathbb{E} [Z_i | \mathbf{X}_i] \\ &\leq \left( \frac{m_i e^{\beta_0^\top \mathbf{X}_i} + e^{\gamma_0^\top \mathbf{X}_i}}{g_i(\theta_0) k_i(\theta_0)} \right)^2 + \left( \frac{e^{\gamma_0^\top \mathbf{X}_i}}{f_i(\theta_0)} - 1 \right)^2 M^2, \end{aligned} \quad (6.17)$$

since  $\text{var}(J_i | \mathbf{X}_i) \leq 1$ ,  $\text{var}(Z_i | \mathbf{X}_i) \leq M^2$  and  $(e^{\gamma_0^\top \mathbf{X}_i}/f_i(\theta_0) - 1)$  is negative. Let  $M_{\mathbf{X}} = \max_{u \in \mathcal{C}, \mathbf{X} \in \mathcal{X}} e^{u^\top \mathbf{X}}$  and  $m_{\mathbf{X}} = \min_{u \in \mathcal{C}, \mathbf{X} \in \mathcal{X}} e^{u^\top \mathbf{X}}$ , where  $\mathcal{X} = [-c_1, c_1]^p$ . We deduce from (6.17) that :

$$\text{var} \left( \dot{\ell}_{\gamma,i}(\theta_0) | \mathbf{X}_i \right) \leq c_2 := \left( \frac{(M+1)M_{\mathbf{X}}}{m_{\mathbf{X}}} \right)^2 + M^2 < \infty,$$

and thus,

$$\text{var} \left( X_{ij} \dot{\ell}_{\gamma,i}(\theta_0) \right) \leq c_1^2 c_2.$$

It follows that

$$\sum_{i=1}^{\infty} \frac{\text{var}(X_{ij} \dot{\ell}_{\gamma,i}(\theta_0))}{i^2} \leq c_1^2 c_2 \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty.$$

By the SLLN (Jiang, 2010, Theorem 6.7), we conclude that

$$\frac{1}{n} \sum_{i=1}^n \left\{ X_{ij} \dot{\ell}_{\gamma,i}(\theta_0) - \mathbb{E} \left[ X_{ij} \dot{\ell}_{\gamma,i}(\theta_0) \right] \right\} = \frac{1}{n} \sum_{i=1}^n X_{ij} \dot{\ell}_{\gamma,i}(\theta_0), \quad j = 1, \dots, p$$

converge almost surely to 0 as  $n \rightarrow \infty$ . Using similar arguments, we can prove that  $\frac{1}{n} \sum_{i=1}^n X_{ij} \dot{\ell}_{\beta,i}(\theta_0)$ ,  $j = 1, \dots, p$  converge almost surely to 0. Therefore,  $n^{-1/2} \ell_n(\theta_0)$  converges almost surely to 0, which concludes the proof.  $\square$

**Lemma 6.3.** *As  $n \rightarrow \infty$ ,  $n^{-1/2} \partial \dot{\ell}_n(\theta) / \partial \theta^\top$  converges almost surely to  $L(\theta)$ , uniformly in an open neighborhood of  $\theta_0$ . Moreover,  $L$  is continuous.*

**Proof of Lemma 6.3.** Let  $\tilde{\ell}_n(\theta) = n^{-1/2} \partial \dot{\ell}_n(\theta) / \partial \theta^\top$  and  $\mathcal{V}_{\theta_0}$  be an open neighborhood of  $\theta_0$ . Let  $\theta \in \mathcal{V}_{\theta_0}$  and consider the  $(\ell, j)$ -th element of  $\tilde{\ell}_n(\theta)$ , namely:

$$\left( \tilde{\ell}_n(\theta) \right)_{(\ell, j)} = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ell_i(\theta)}{\partial \theta_\ell \partial \theta_j}.$$

We have:

$$\left( \tilde{\ell}_n(\theta) \right)_{(\ell, j)} = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\partial^2 \ell_i(\theta)}{\partial \theta_\ell \partial \theta_j} - \mathbb{E} \left[ \frac{\partial^2 \ell_i(\theta)}{\partial \theta_\ell \partial \theta_j} \right] \right\} + \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \frac{\partial^2 \ell_i(\theta)}{\partial \theta_\ell \partial \theta_j} \right].$$

Now,

$$\text{var} \left( \frac{\partial^2 \ell_i(\theta)}{\partial \theta_\ell \partial \theta_j} \right) \leq \mathbb{E} \left( \left\{ \frac{\partial^2 \ell_i(\theta)}{\partial \theta_\ell \partial \theta_j} \right\}^2 \right).$$

We prove that  $\text{var}(\partial^2 \ell_i(\theta) / \partial \theta_\ell \partial \theta_j)$  is bounded. Some tedious (although uncomplicated) algebra shows that  $\partial^2 \ell_i(\theta) / \partial \theta_\ell \partial \theta_j$  is the  $(\ell, j)$ -th element of the  $(2p \times 2p)$  matrix  $-\mathbf{V}_i \mathbf{D}_i(\theta) \mathbf{V}_i^\top$ , where  $\mathbf{V}_i$  is the  $(2p \times 2)$  matrix defined as

$$\mathbf{V}_i = \begin{pmatrix} \mathbf{X}_i & 0_{p,1} \\ 0_{p,1} & \mathbf{X}_i \end{pmatrix}$$

and

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

is the  $(2 \times 2)$  symmetric matrix defined by

$$\begin{aligned} \mathbf{D}_{1,i}(\theta) &= J_i \frac{e^{\beta^\top \mathbf{X}_i} U_i(\theta)}{[(1 + e^{\beta^\top \mathbf{X}_i}) g_i(\theta) k_i(\theta)]^2} + (Z_i - 1) \frac{e^{\beta^\top \mathbf{X}_i} (1 + e^{\gamma^\top \mathbf{X}_i})}{f_i^2(\theta)} + \frac{e^{\beta^\top \mathbf{X}_i} V_i(\theta)}{[(1 + e^{\beta^\top \mathbf{X}_i}) g_i(\theta)]^2}, \\ \mathbf{D}_{2,i}(\theta) &= J_i \frac{W_i(\theta)}{[g_i(\theta) k_i(\theta)]^2} + (Z_i - 1) \frac{e^{\gamma^\top \mathbf{X}_i} (1 + e^{\beta^\top \mathbf{X}_i})}{f_i^2(\theta)} + \frac{(1 - m_i) e^{(\beta + \gamma)^\top \mathbf{X}_i}}{g_i^2(\theta)}, \\ \mathbf{D}_{3,i}(\theta) &= -J_i \frac{e^{\beta^\top \mathbf{X}_i} Y_i(\theta)}{(1 + e^{\beta^\top \mathbf{X}_i}) [g_i(\theta) k_i(\theta)]^2} + (1 - Z_i) \frac{e^{(\beta + \gamma)^\top \mathbf{X}_i}}{f_i^2(\theta)} + \frac{(m_i - 1) e^{(\beta + \gamma)^\top \mathbf{X}_i}}{g_i^2(\theta)}, \end{aligned}$$

with

$$\begin{aligned}
U_i(\theta) &= e^{\beta^\top \mathbf{x}_i} Q_i(\theta) \left[ g_i(\theta) k_i(\theta) + \left(1 + e^{\beta^\top \mathbf{x}_i}\right) k_i(\theta) + e^{\beta^\top \mathbf{x}_i} Q_i(\theta) g_i(\theta) h_i^{m_i}(\theta) \right] \\
&\quad - \left( Q_i(\theta) + e^{\beta^\top \mathbf{x}_i} (1 - 2m_i) \right) \left(1 + e^{\beta^\top \mathbf{x}_i}\right) g_i(\theta) k_i(\theta), \\
V_i(\theta) &= \left[ Q_i(\theta) + e^{\beta^\top \mathbf{x}_i} (1 - 2m_i) \right] \left(1 + e^{\beta^\top \mathbf{x}_i}\right) g_i(\theta) - e^{\beta^\top \mathbf{x}_i} Q_i(\theta) \left( f_i(\theta) + e^{\beta^\top \mathbf{x}_i} \right), \\
W_i(\theta) &= e^{(\beta+\gamma)^\top \mathbf{x}_i} k_i(\theta) (m_i - 1) + g_i(\theta) h_i^{m_i}(\theta) \left( m_i e^{\beta^\top \mathbf{x}_i} + e^{\gamma^\top \mathbf{x}_i} \right)^2, \\
Y_i(\theta) &= m_i e^{\gamma^\top \mathbf{x}_i} g_i(\theta) k_i(\theta) + \left[ e^{\gamma^\top \mathbf{x}_i} k_i(\theta) + g_i(\theta) h_i^{m_i}(\theta) \left( m_i e^{\beta^\top \mathbf{x}_i} + e^{\gamma^\top \mathbf{x}_i} \right) \right] Q_i(\theta).
\end{aligned}$$

With these notations, it is easy to see that

$$\begin{aligned}
\frac{\partial^2 \ell_i(\theta)}{\partial \theta_\ell \partial \theta_j} &= - \left( \mathbf{V}_{i,(\ell,1)} \mathbf{D}_{1,i}(\theta) + \mathbf{V}_{i,(\ell,2)} \mathbf{D}_{3,i}(\theta) \right) \mathbf{V}_{i,(j,1)} \\
&\quad - \left( \mathbf{V}_{i,(\ell,1)} \mathbf{D}_{3,i}(\theta) + \mathbf{V}_{i,(\ell,2)} \mathbf{D}_{2,i}(\theta) \right) \mathbf{V}_{i,(j,2)}, \tag{6.18}
\end{aligned}$$

where for example,  $\mathbf{V}_{i,(\ell,1)}$  denotes the element on the  $\ell$ -th row and first column of  $\mathbf{V}_i$ . For every  $\ell$  ( $\ell = 1, \dots, 2p$ ), at least one of  $\mathbf{V}_{i,(\ell,1)}$  and  $\mathbf{V}_{i,(\ell,2)}$  must be equal to 0 (this is straightforward from the expression of  $\mathbf{V}_i$ ). Suppose for example that  $\mathbf{V}_{i,(\ell,1)} = 0$  and  $\mathbf{V}_{i,(j,2)} = 0$ . Then :

$$\frac{\partial^2 \ell_i(\theta)}{\partial \theta_\ell \partial \theta_j} = - \mathbf{V}_{i,(\ell,2)} \mathbf{D}_{3,i}(\theta) \mathbf{V}_{i,(j,1)}.$$

Any other combination of null and non-null values among  $(\mathbf{V}_{i,(\ell,1)}, \mathbf{V}_{i,(\ell,2)})$  and  $(\mathbf{V}_{i,(j,1)}, \mathbf{V}_{i,(j,2)})$  yields a similar expression, of the form  $-\mathbf{V}_{i,(\ell,a)} \mathbf{D}_{c,i}(\theta) \mathbf{V}_{i,(j,b)}$ , with  $a, b \in \{1, 2\}$  and  $c \in \{1, 2, 3\}$ . It is not difficult to see that under conditions 1, 2 and 4, the terms  $\mathbf{D}_{c,i}(\theta)$  are bounded by some finite constant  $c_3$  (proof is omitted), and thus,

$$\text{var} \left( \frac{\partial^2 \ell_i(\theta)}{\partial \theta_\ell \partial \theta_j} \right) \leq c_1^4 c_3^2.$$

It follows that

$$\sum_{i=1}^{\infty} \frac{\text{var} \left( \frac{\partial^2 \ell_i(\theta)}{\partial \theta_\ell \partial \theta_j} \right)}{i^2} \leq c_1^4 c_3^2 \sum_{i=1}^{\infty} \frac{1}{i^2} < \infty.$$

Therefore, the SLLN implies that

$$\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\partial^2 \ell_i(\theta)}{\partial \theta_\ell \partial \theta_j} - \mathbb{E} \left[ \frac{\partial^2 \ell_i(\theta)}{\partial \theta_\ell \partial \theta_j} \right] \right\}$$

converges almost surely to 0 as  $n \rightarrow \infty$  and by condition 3,  $(\tilde{\ell}_n(\theta))_{(\ell,j)}$  converges almost surely to the  $(\ell, j)$ -th element of the matrix  $L(\theta)$ . Under conditions 1, 2 and 4, the derivative of  $\tilde{\ell}_n(\theta)$  with respect to  $\theta$  is bounded, for every  $n$ . Therefore, the sequence  $(\tilde{\ell}_n(\theta))_n$  is equicontinuous. It follows that the convergence of  $(\tilde{\ell}_n(\theta))_n$  to  $L(\theta)$  is uniform on  $\mathcal{V}_{\theta_0}$ , and that  $L$  must be continuous.  $\square$

Having verified the conditions of Foutz (1977) inverse function theorem, we conclude that  $\hat{\theta}_n$  converges almost surely to  $\theta_0$ .  $\square$

**Proof of Theorem 3.3.** A Taylor series expansion of  $\dot{\ell}_n(\hat{\theta}_n)$  at  $\theta_0$  yields :

$$0 = \dot{\ell}_n(\hat{\theta}_n) = \dot{\ell}_n(\theta_0) + \frac{1}{\sqrt{n}} \frac{\partial \dot{\ell}_n(\theta_0)}{\partial \theta^\top} \sqrt{n}(\hat{\theta}_n - \theta_0) + o_{\mathbb{P}}(1),$$

and thus

$$\begin{aligned} \sqrt{n}(\hat{\theta}_n - \theta_0) &= - \left( \frac{1}{\sqrt{n}} \frac{\partial \dot{\ell}_n(\theta_0)}{\partial \theta^\top} \right)^{-1} \dot{\ell}_n(\theta_0) + o_{\mathbb{P}}(1) \\ &= - \left( \frac{1}{\sqrt{n}} \frac{\partial \dot{\ell}_n(\theta_0)}{\partial \theta^\top} \right)^{-1} \sum_{j=1}^{2n} \mathbb{V}_{\bullet j} S_{j,n}(\theta_0) + o_{\mathbb{P}}(1), \end{aligned}$$

where  $S_{j,n}(\theta) = \frac{1}{\sqrt{n}} S_j(\theta)$ . First, we prove that  $\sum_{j=1}^{2n} \mathbb{V}_{\bullet j} S_{j,n}(\theta_0)$  is asymptotically normal. To see this, we apply the multivariate central limit theorem for linear combinations of random vectors of Eicker (1966).

Consider the random linear combination  $T_n = \mathbb{S}_n^{-1} \sum_{j=1}^{2n} \mathbb{V}_{\bullet j} S_{j,n}(\theta_0)$ , where  $\mathbb{S}_n^2 = \text{var}(\dot{\ell}_n(\theta_0))$ . By Eicker (1966),  $T_n$  is asymptotically distributed as a multivariate standard normal if the following conditions hold:

- a)  $\max_{1 \leq j \leq 2n} \mathbb{V}_{\bullet j}^\top (\mathbb{V}\mathbb{V}^\top)^{-1} \mathbb{V}_{\bullet j} \rightarrow 0$  as  $n \rightarrow \infty$ ,
- b)  $\sup_{1 \leq j \leq 2n} \mathbb{E}[S_{j,n}^2(\theta_0) 1_{\{|S_{j,n}(\theta_0)| > c\}}] \rightarrow 0$  as  $c \rightarrow \infty$ ,
- c)  $\inf_{1 \leq j \leq 2n} \mathbb{E}[S_{j,n}^2(\theta_0)] > 0$ .

Note first that

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

Since  $\|\mathbb{V}_{\bullet j}\|$  is bounded, condition 3 implies that a) is satisfied. Condition b) follows by noting that the  $S_{j,n}(\theta_0)$ ,  $j = 1, \dots, 2n$  are bounded under conditions 1, 2 and 4. Finally, under conditions 1, 2 and 4, we have  $\mathbb{E}[S_{j,n}^2(\theta_0)] > 0$  for every  $j = 1, \dots, 2n$ . Now,  $\mathbb{S}_n^2 = \text{var}(\dot{\ell}_n(\theta_0)) = n^{-1} \sum_{i=1}^n \text{var}(\partial \ell_i(\theta_0) / \partial \theta) = -n^{-1} \sum_{i=1}^n \mathbb{E}[\partial^2 \ell_i(\theta) / \partial \theta \partial \theta^\top]$  converges to

$-L(\theta_0)$  by condition 3. It follows that  $\sum_{j=1}^{2n} \mathbb{V}_{\bullet j} S_{j,n}(\theta_0)$  is asymptotically distributed as a multivariate normal with mean zero and variance  $L(\theta_0)$ .

Finally, by Lemma 6.3 and Slutsky's theorem, we conclude that  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  is asymptotically distributed as a multivariate normal with mean zero and variance  $L(\theta_0)^{-1}$ .

Now, as  $n \rightarrow \infty$ , we have :

$$\left\| n^{-1/2} \partial \dot{\ell}_n(\hat{\theta}_n) / \partial \theta^\top - L(\theta_0) \right\| \leq \sup_{\theta \in \mathcal{V}_{\theta_0}} \left\| \tilde{\ell}_n(\theta) - L(\theta) \right\| + \left\| L(\hat{\theta}_n) - L(\theta_0) \right\| \quad (6.19)$$

The first term in the right-hand-side of (6.19) converges to 0 by Lemma 6.3. Consistency of  $\hat{\theta}_n$  and continuity of  $L$  imply that the second term also converges to 0. It follows that  $n^{-1/2} \partial \dot{\ell}_n(\hat{\theta}_n) / \partial \theta^\top$  converges to  $L(\theta_0)$  and finally, that  $(n^{-1/2} \partial \dot{\ell}_n(\hat{\theta}_n) / \partial \theta^\top)^{-1}$  is a consistent estimate of the asymptotic variance  $L(\theta_0)^{-1}$ .  $\square$

## Acknowledgements

Authors acknowledge supports from CEA-MITIC (an African Center of Excellence in Mathematics, Informatics) and ICT, implemented by Gaston Berger University (Senegal) and from Université de Rennes 1 ("Action internationale 2019").

## References

- Albert, J. M., Wang, W., Nelson, S., 2014. Estimating overall exposure effects for zero-inflated regression models with application to dental caries. *Statistical Methods in Medical Research* 23(3), 257-278.
- 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.
- Diallo, A. O., Diop, A., Dupuy, J.-F., 2018. Analysis of multinomial counts with joint zero-inflation, with an application to health economics. *Journal of Statistical Planning and Inference* 194, 85-105.
- Diallo, A. O., Diop, A., Dupuy, J.-F., 2019. Estimation in zero-inflated binomial regression with missing covariates. *Statistics* 53(4), 839-865.
- Dupuy, J.-F., 2018. *Statistical Methods for Overdispersed Count Data*. ISTE Press - Elsevier.
- Eicker, F., 1966. A multivariate central limit theorem for random linear vector forms. *The Annals of Mathematical Statistics* 37(6), 1825-1828.
- Feng, J., Zhu, Z., 2011. Semiparametric analysis of longitudinal zero-inflated count data. *Journal of Multivariate Analysis* 102, 61-72.
- Foutz, R. V., 1977. On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association* 72, 147-148.

- Hall, D. B., 2000. Zero-inflated Poisson and binomial regression with random effects: a case study. *Biometrics* 56(4), 1030-1039.
- 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.
- Long, D. L., Preisser, J. S., Herring, A. H., Golin, C. E., 2014. A marginalized zero-inflated Poisson regression model with overall exposure effects. *Statistics in medicine* 33(29), 5151-5165.
- Long, D. L., Preisser, J. S., Herring, A. H., Golin, C. E., 2015. A Marginalized Zero-inflated Poisson Regression Model with Random Effects. *Journal of the Royal Statistical Society. Series C, Applied statistics* 64(5), 815-830.
- Martin, J., Hall, D. B., 2017. Marginal zero-inflated regression models for count data. *Journal of Applied Statistics* 44(10), 1807-1826.
- 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.
- Preisser, J. S., Stamm, J. W., Long, D. L., Kincade, M. E., 2012. Review and recommendations for zero-inflated count regression modeling of dental caries indices in epidemiological studies. *Caries research* 46(4), 413-423.
- Preisser, J. S., Das, K., Long, D. L., and Divaris, K., 2016. Marginalized zero-inflated negative binomial regression with application to dental caries. *Statistics in Medicine* 35(10), 1722-1735.
- R Core Team, 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://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.

Todem, D., Kim, K., Hsu, W. W., 2016. Marginal mean models for zero-inflated count data. *Biometrics* 72(3), 986-994.

$c$	$\hat{\beta}_n$								$\hat{\gamma}_n$								
	$\hat{\beta}_{1,n}$	$\hat{\beta}_{2,n}$	$\hat{\beta}_{3,n}$	$\hat{\beta}_{4,n}$	$\hat{\beta}_{5,n}$	$\hat{\beta}_{6,n}$	$\hat{\beta}_{7,n}$	$\hat{\beta}_{8,n}$	$\hat{\gamma}_{1,n}$	$\hat{\gamma}_{2,n}$	$\hat{\gamma}_{3,n}$	$\hat{\gamma}_{4,n}$	$\hat{\gamma}_{5,n}$	$\hat{\gamma}_{6,n}$	$\hat{\gamma}_{7,n}$	$\hat{\gamma}_{8,n}$	
0.25	bias	-0.0308	-0.0116	0.0015	-0.0073	0.0117	0.0062	0.0094	0.0052	-0.1443	0.0959	-0.0982	0.2043	-0.2137	0.1716	0.0007	0.0235
	SD	0.4602	0.1117	0.0935	0.0843	0.1163	0.2170	0.1774	0.2533	1.3297	0.4019	0.3135	0.4312	0.5275	0.7809	0.5245	0.7620
	SE	0.4648	0.1105	0.0883	0.0825	0.1156	0.2131	0.1727	0.2520	1.2953	0.3744	0.2950	0.3885	0.4697	0.7365	0.4851	0.7235
	RMSE	0.6546	0.1575	0.1286	0.1182	0.1644	0.3041	0.2477	0.3573	1.8615	0.5574	0.4414	0.6151	0.7378	1.0867	0.7142	1.0507
	CP	0.9530	0.9490	0.9450	0.9460	0.9580	0.9450	0.9380	0.9430	0.9610	0.9470	0.9570	0.9530	0.9490	0.9540	0.9450	0.9490
	$\ell(\text{CI})$	1.8180	0.4321	0.3457	0.3228	0.4522	0.8338	0.6757	0.9837	4.9720	1.4017	1.1135	1.4151	1.7266	2.6520	1.8743	2.7536
0.50	bias	-0.0214	-0.0103	0.0048	-0.0081	0.0159	-0.0045	0.0031	0.0232	-0.0787	0.0850	0.0898	-0.0114	-0.1679	-0.0767	0.0023	0.0139
	SD	0.4710	0.1156	0.0890	0.0776	0.1171	0.2252	0.1726	0.2539	1.2360	0.3615	0.2649	0.2359	0.4394	0.5723	0.4864	0.7114
	SE	0.4610	0.1105	0.0894	0.0772	0.1167	0.2229	0.1714	0.2499	1.1857	0.3334	0.2514	0.2142	0.4125	0.5549	0.4492	0.6714
	RMSE	0.6593	0.1602	0.1262	0.1097	0.1661	0.3168	0.2432	0.3569	1.7141	0.4989	0.3759	0.3188	0.6255	0.8006	0.6619	0.9780
	CP	0.9510	0.9349	0.9570	0.9520	0.9499	0.9469	0.9469	0.9469	0.9429	0.9550	0.9550	0.9309	0.9499	0.9520	0.9439	0.9489
	$\ell(\text{CI})$	1.8031	0.4322	0.3498	0.3018	0.4567	0.8715	0.6707	0.9757	4.5859	1.2776	0.9634	0.8272	1.5634	2.1471	1.7401	2.5600
0.75	bias	0.0081	-0.0293	0.0119	-0.0224	0.0189	-0.0080	-0.0109	0.0079	0.0176	0.0163	0.0540	-0.0304	-0.0693	-0.0690	-0.0555	-0.0529
	SD	0.5719	0.1434	0.1154	0.1000	0.1456	0.2818	0.2241	0.2986	1.0455	0.2647	0.2177	0.1875	0.3201	0.5067	0.3946	0.5990
	SE	0.5646	0.1377	0.1135	0.0977	0.1390	0.2852	0.2184	0.2996	1.0152	0.2685	0.2073	0.1797	0.3126	0.4964	0.3921	0.5616
	RMSE	0.8035	0.2009	0.1622	0.1415	0.2022	0.4009	0.3130	0.4229	1.4570	0.3773	0.3054	0.2614	0.4527	0.7125	0.5589	0.8226
	CP	0.9500	0.9350	0.9470	0.9410	0.9420	0.9550	0.9480	0.9600	0.9470	0.9610	0.9380	0.9450	0.9460	0.9470	0.9580	0.9360
	$\ell(\text{CI})$	2.2071	0.5383	0.4439	0.3818	0.5435	1.1146	0.8543	1.1691	3.9598	1.0440	0.8073	0.7007	1.2118	1.9369	1.5307	2.1860
0.90	bias	-0.0233	-0.0299	0.0115	-0.0297	0.0338	-0.0078	0.0020	0.0130	-0.1404	0.0133	0.0615	-0.0229	0.0132	-0.1967	-0.0558	0.0226
	SD	0.7438	0.1869	0.1462	0.1337	0.1789	0.3626	0.2797	0.4134	1.5928	0.3651	0.3118	0.2656	0.3670	1.1673	0.5991	0.9384
	SE	0.7129	0.1742	0.1422	0.1264	0.1802	0.3544	0.2762	0.3921	1.4259	0.3644	0.3053	0.2506	0.3616	0.9626	0.5584	0.7911
	RMSE	1.0303	0.2571	0.2042	0.1863	0.2561	0.5070	0.3930	0.5698	2.1419	0.5159	0.4406	0.3658	0.5152	1.5252	0.8207	1.2272
	CP	0.9440	0.9390	0.9490	0.9390	0.9500	0.9490	0.9490	0.9380	0.9490	0.9560	0.9530	0.9500	0.9520	0.9620	0.9480	0.9470
	$\ell(\text{CI})$	2.7847	0.6797	0.5559	0.4932	0.7034	1.3849	1.0799	1.5289	5.4261	1.4063	1.1754	0.9683	1.3958	3.1572	2.1576	2.8570

Table 1: Simulation results for  $n = 500$ .  $c$ : average proportion of zero-inflation. SD: empirical standard deviation. SE: average standard error. CP: empirical coverage probability of 95%-level confidence intervals.  $\ell(\text{CI})$ : average length of confidence intervals.

$c$	$\hat{\beta}_n$								$\hat{\gamma}_n$							
	$\hat{\beta}_{1,n}$	$\hat{\beta}_{2,n}$	$\hat{\beta}_{3,n}$	$\hat{\beta}_{4,n}$	$\hat{\beta}_{5,n}$	$\hat{\beta}_{6,n}$	$\hat{\beta}_{7,n}$	$\hat{\beta}_{8,n}$	$\hat{\gamma}_{1,n}$	$\hat{\gamma}_{2,n}$	$\hat{\gamma}_{3,n}$	$\hat{\gamma}_{4,n}$	$\hat{\gamma}_{5,n}$	$\hat{\gamma}_{6,n}$	$\hat{\gamma}_{7,n}$	$\hat{\gamma}_{8,n}$
0.25																
bias	-0.0042	-0.0027	-0.0001	-0.0033	0.0011	-0.0010	-0.0021	0.0052	-0.0034	0.0104	-0.0172	0.0243	-0.0278	0.0260	-0.0057	-0.0039
SD	0.2273	0.0557	0.0411	0.0405	0.0583	0.1058	0.0861	0.1244	0.5636	0.1546	0.1219	0.1451	0.1797	0.2835	0.2190	0.3114
SE	0.2281	0.0542	0.0434	0.0407	0.0567	0.1047	0.0851	0.1230	0.5519	0.1526	0.1211	0.1460	0.1805	0.2844	0.2096	0.3035
RMSE	0.3220	0.0777	0.0598	0.0575	0.0813	0.1488	0.1210	0.1749	0.7886	0.2174	0.1727	0.2072	0.2562	0.4023	0.3031	0.4348
CP	0.9580	0.9380	0.9620	0.9560	0.9390	0.9500	0.9480	0.9500	0.9510	0.9580	0.9590	0.9560	0.9560	0.9520	0.9430	0.9450
$\ell(\text{CI})$	0.8937	0.2125	0.1702	0.1594	0.2222	0.4104	0.3334	0.4816	2.1595	0.5968	0.4734	0.5697	0.7050	1.1123	0.8207	1.1873
0.50																
bias	-0.0039	-0.0008	0.0010	-0.0037	0.0031	0.0024	-0.0022	0.0085	0.0031	0.0189	0.0156	-0.0071	-0.0343	-0.0168	-0.0031	-0.0037
SD	0.2245	0.0551	0.0442	0.0385	0.0587	0.1124	0.0856	0.1246	0.5032	0.1464	0.1122	0.0941	0.1710	0.2488	0.1956	0.3059
SE	0.2251	0.0541	0.0438	0.0377	0.0571	0.1090	0.0839	0.1214	0.5210	0.1434	0.1075	0.0944	0.1704	0.2436	0.1976	0.2864
RMSE	0.3179	0.0772	0.0622	0.0540	0.0819	0.1566	0.1199	0.1741	0.7242	0.2057	0.1561	0.1335	0.2438	0.3485	0.2780	0.4189
CP	0.9510	0.9430	0.9550	0.9510	0.9380	0.9370	0.9400	0.9560	0.9620	0.9470	0.9450	0.9520	0.9600	0.9410	0.9540	0.9320
$\ell(\text{CI})$	0.8820	0.2119	0.1716	0.1476	0.2237	0.4271	0.3286	0.4753	2.0388	0.5606	0.4204	0.3694	0.6658	0.9539	0.7737	1.1204
0.75																
bias	-0.0256	-0.0001	0.0033	-0.0024	0.0014	-0.0005	0.0075	0.0047	-0.0233	0.0074	0.0134	-0.0035	-0.0158	-0.0062	-0.0055	-0.0006
SD	0.2759	0.0652	0.0580	0.0479	0.0662	0.1380	0.1049	0.1508	0.4898	0.1247	0.0953	0.0833	0.1390	0.2241	0.1894	0.2590
SE	0.2745	0.0668	0.0554	0.0473	0.0675	0.1388	0.1063	0.1455	0.4704	0.1233	0.0954	0.0829	0.1418	0.2303	0.1819	0.2575
RMSE	0.3899	0.0933	0.0803	0.0674	0.0945	0.1957	0.1495	0.2095	0.6794	0.1755	0.1355	0.1175	0.1991	0.3213	0.2626	0.3652
CP	0.9460	0.9600	0.9380	0.9560	0.9530	0.9510	0.9530	0.9510	0.9400	0.9430	0.9380	0.9460	0.9540	0.9590	0.9510	0.9510
$\ell(\text{CI})$	1.0752	0.2616	0.2170	0.1852	0.2646	0.5437	0.4166	0.5698	1.8425	0.4826	0.3736	0.3245	0.5548	0.9021	0.7127	1.0085
0.90																
bias	0.0219	-0.0076	0.0037	-0.0080	0.0118	-0.0007	-0.0088	-0.0028	-0.0007	0.0026	0.0091	-0.0055	0.0130	-0.0291	-0.0186	0.0073
SD	0.3475	0.0835	0.0704	0.0608	0.0888	0.1710	0.1333	0.1923	0.6263	0.1609	0.1355	0.1088	0.1584	0.3300	0.2464	0.3290
SE	0.3438	0.0839	0.0688	0.0608	0.0869	0.1715	0.1340	0.1882	0.6126	0.1583	0.1336	0.1095	0.1572	0.3356	0.2464	0.3173
RMSE	0.4892	0.1186	0.0985	0.0863	0.1247	0.2421	0.1892	0.2690	0.8758	0.2257	0.1904	0.1545	0.2235	0.4715	0.3489	0.4570
CP	0.9520	0.9500	0.9450	0.9510	0.9380	0.9640	0.9530	0.9520	0.9470	0.9530	0.9450	0.9490	0.9470	0.9610	0.9530	0.9420
$\ell(\text{CI})$	1.3465	0.3286	0.2695	0.2380	0.3402	0.6719	0.5249	0.7368	2.3972	0.6192	0.5224	0.4284	0.6149	1.3102	0.9645	1.2413

Table 2: Simulation results for  $n = 2000$ .  $c$ : average proportion of zero-inflation. SD: empirical standard deviation. SE: average standard error. CP: empirical coverage probability of 95%-level confidence intervals.  $\ell(\text{CI})$ : average length of confidence intervals.

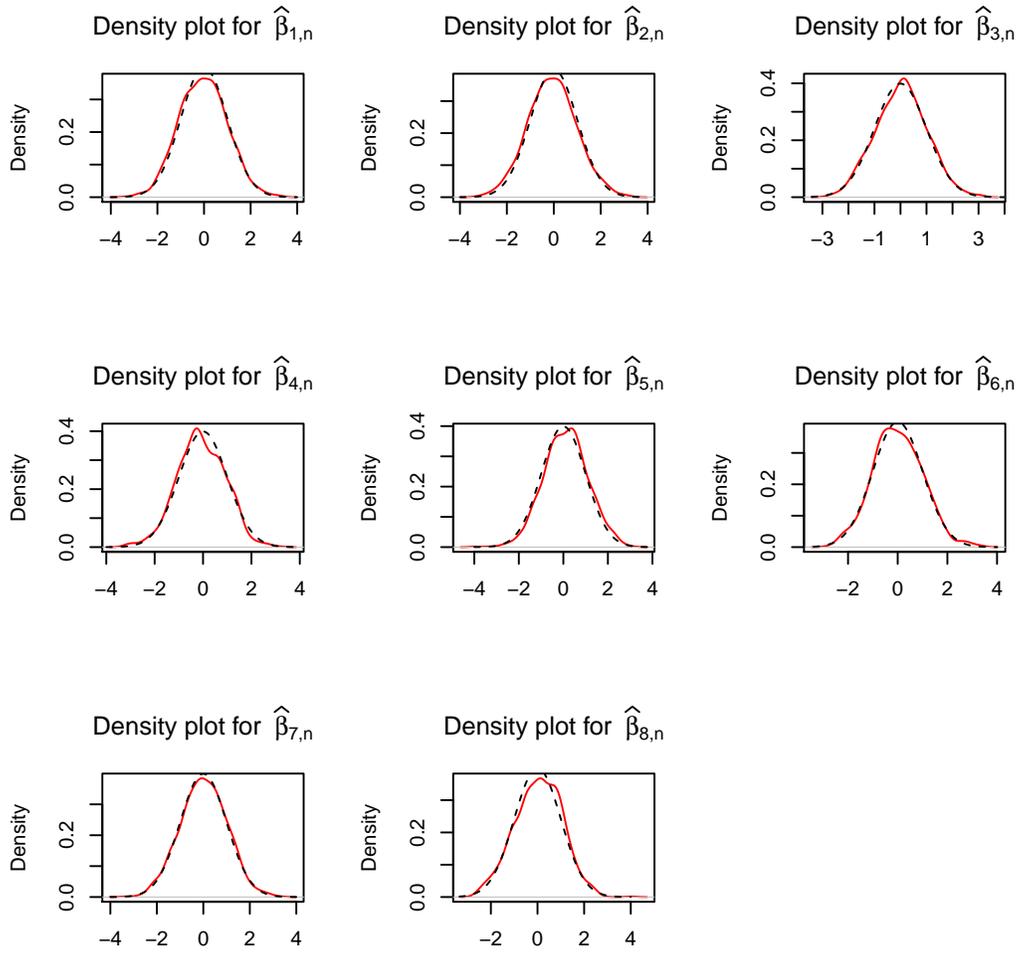


Figure 1: Density estimates of the  $(\hat{\beta}_{j,n} - \beta_j)/\text{standard error}(\hat{\beta}_{j,n}), j = 1, \dots, 8$  with  $n = 500$  and 50% of zero-inflation.

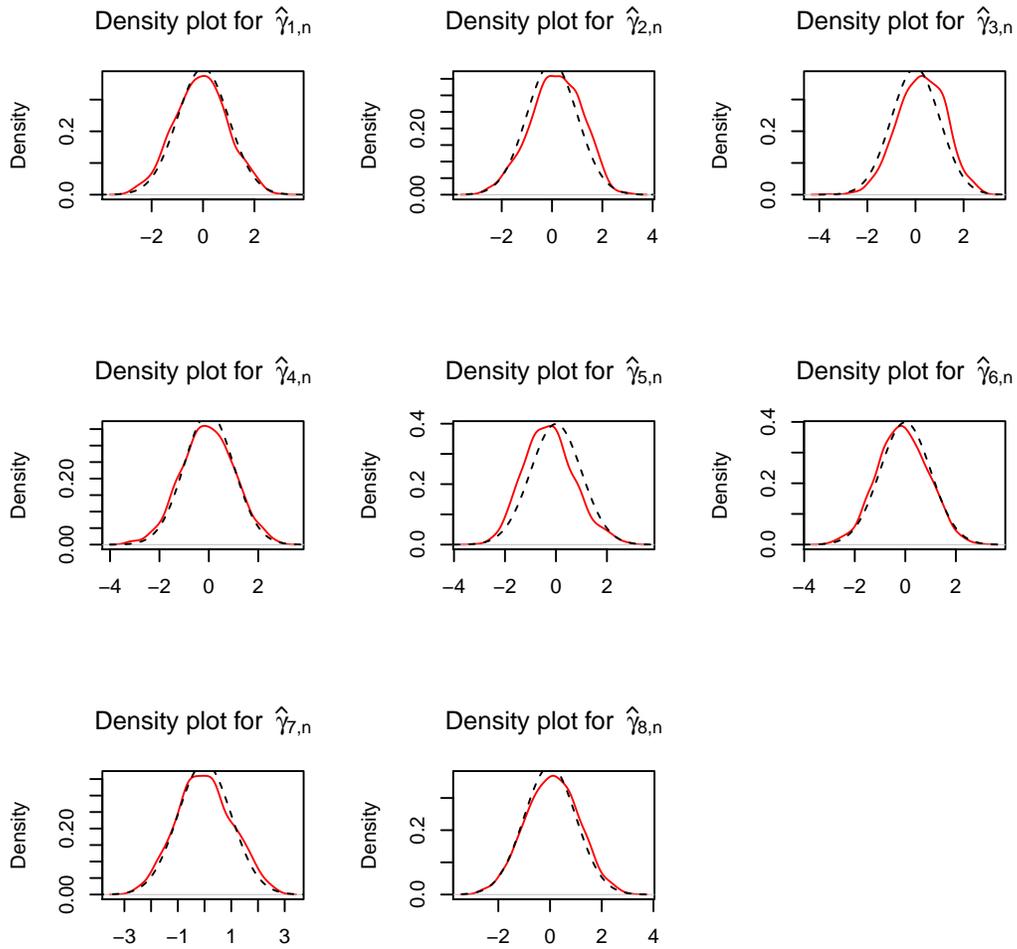


Figure 2: Density estimates of the  $(\hat{\gamma}_{k,n} - \gamma_k)/\text{standard error}(\hat{\gamma}_{k,n}), k = 1, \dots, 8$  with  $n = 500$  and 50% of zero-inflation.

	Parameter	Parameter estimate	Standard error
intercept	$\beta_1$	-0.8181	0.2130
health1	$\beta_2$	-0.4284	0.0532
health2	$\beta_3$	0.2547	0.0599
chronic	$\beta_4$	-0.1049	0.0116
age	$\beta_5$	-0.1315	0.0253
gender	$\beta_6$	0.1714	0.0567
education	$\beta_7$	0.0326	0.0076
medicaid	$\beta_8$	-0.2532	0.1215
intercept	$\gamma_1$	-1.4925	0.1480
gender	$\gamma_2$	0.3573	0.0909
education	$\gamma_3$	0.0742	0.0119
medicaid	$\gamma_4$	-0.3991	0.1766

Table 3: Health-care data analysis with the marginal ZIB regression model.