

# Statistical Inference in a Zero-Inflated Bell Regression Model

Essoham Ali<sup>1,2\*</sup>, Mamadou Lamine Diop<sup>2,3\*\*</sup>, and Aliou Diop<sup>2\*\*\*</sup>

<sup>1</sup>*Université de Bretagne Sud, LMBA-CNRS UMR 6205, 56000 Vannes, France*

<sup>2</sup>*Université Gaston Berger, LERSTAD, BP: 234, Saint-Louis, Sénégal*

<sup>3</sup>*THEMA, CY Cergy Paris Université, 33 Boulevard du Port, 95011 Cergy-Pontoise Cedex, France*

Received June 6, 2022; revised August 1, 2022; accepted August 4, 2022

**Abstract**—In this paper, we study the asymptotic properties of the Maximum Likelihood Estimator (MLE) for a Zero-Inflated Bell regression model. Under some regularity conditions, we establish that the estimator is consistent and asymptotically normal. This lends a substantial support to the empirical findings that have already been obtained by some authors. Monte Carlo simulations are conducted to numerically illustrate the main results. The model is applied to a dataset of healthcare demand in USA.

**DOI:** 10.3103/S1066530722030012

**Keywords:** Zero-Inflated Bell regression model, Maximum Likelihood Estimator, regularity conditions, asymptotic properties

## 1. INTRODUCTION

Statisticians have long recognized the difficulties of working with count data. Poisson and binomial [26] regression models are the usual tools for analyzing count data. They are applied to many fields such as health economics, epidemiology or environmental sciences. Generalized linear models [25] provide a powerful framework for analysing such data. But real-life applications continuously raise new challenging problems and a huge amount of work has been done to extend the scope of these models. For example, count data often show excess of zeros (that is, a proportion of zeros that cannot be explained by models based on standard distributional assumptions). The variance is considered as a function of the mean. These relationships between variance and mean make sense in a theoretical context where, for example, all the processes leading to the presence/absence or abundance of the species are well modeled and measured without error [16, 20]. In practice, it is illusory to think that all the variables that can affect these processes can be incorporated into the model, either because they are unknown or because they are impossible to measure in the field, and this can lead to a mis-specification of the variance, with the mean remaining theoretically unaffected. The most frequent situation is overdispersion; i.e., when the variance of the observations is higher than the one expected under the model used. Many solutions have been proposed to take into account this overdispersion [20] for an overview among which the development of new models, the negative binomial regression and the zero-inflated models such as Zero-Inflated Poisson (ZIP) and Zero-Inflated Binomial (ZIB). Recently, [3] introduced an alternative distribution model called Bell regression model (BRM) to model count data with overdispersion. BRM is generally preferred over the Poisson regression model to overcome the restriction that the mean equals the variance. In the presence of excess of zeros in the count data (which is very frequent in many applications), the Bell regression model is limited. In order to remedy this situation, [22] have recently proposed a zero-inflation Bell regression. In their works, they have addressed the inference by using the maximum likelihood estimator (MLE). Some simulation results show that, the estimate obtained is approaching the true value of the parameter when the sample size increases; which suggests that the MLE is consistent for this model. However, it is important to note that the asymptotic behaviour of the

---

\*E-mail: ali.essoham@ugb.edu.sn

\*\*E-mail: mamadou-lamine.diop@cyu.fr

\*\*\*E-mail: aliou.diop@ugb.edu.sn

estimator has not yet been established. The increasing interest for Zero-Inflated Bell regression model in recent years (see [3, 24]) render necessary to establish such result for the parameter estimate. For example, the asymptotic properties are essential when we want to construct a statistic that can test the nullity of some parameter's components, which in particular, allows us to assess the relevance of the exogenous covariates.

In this new contribution, we provide conditions for the strong consistency and derive the asymptotic distribution of the MLE.

The current paper fills this gap and thus provide rigorous proofs of consistency and asymptotic normality of the maximum likelihood estimators. 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 Bell regression model.

The rest of this paper is organized as follows. We briefly review ZIBell regression model, basics notations and maximum likelihood estimation are introduced in Section 2. In Section 3 we investigate asymptotic properties of maximum likelihood estimates in this model. Results of a simulation study are reported in Section 4. An application of the proposed ZIBell regression model to the analysis of health-care demand data is described in Section 5. Discussion and perspectives are given in Section 6. Technical proofs are postponed to an Appendix.

## 2. ZERO-INFLATED BELL REGRESSION MODEL AND BASIC NOTATIONS

This section will briefly illustrate the Zero-Inflated Bell (ZIBell) regression model and introduces some basic notations.

### 2.1. The ZIBell Regression Model

Let  $y_i$  be a zero-inflated count response and  $\mathbf{S}_i$  a vector of explanatory variables. Let  $\mathbf{X}_i$  and  $\mathbf{Z}_i$  subsets of  $\mathbf{S}_i$ , which may overlap one another. The Zero-Inflated Bell regression model is defined by [22]

$$y_i \sim \text{ZIBell}(\mu_i, \pi_i), \quad \log(\mu_i) = \beta^\top \mathbf{X}_i, \quad \text{logit}(\pi_i) = \gamma^\top \mathbf{Z}_i, \quad 1 \leq i \leq n, \quad (2.1)$$

where  $\beta = (\beta_1, \dots, \beta_p)^\top \in \mathbb{R}^p$  and  $\gamma = (\gamma_1, \dots, \gamma_q)^\top \in \mathbb{R}^q$  are parameters of the regression models to be estimated,  $\top$  denotes the transpose operator, and  $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{ip})^\top$  and  $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, \dots, Z_{iq})^\top$  are the corresponding design matrices and  $\text{ZIBell}(\mu, \pi)$  denotes the ZIBell distribution with the probability mass function

$$\mathbb{P}(Y = y) = \begin{cases} \pi + (1 - \pi) \exp(1 - e^{W(\mu)}) & \text{if } y = 0 \\ (1 - \pi) \exp(1 - e^{W(\mu)}) \frac{W(\mu)^y B_y}{y!} & \text{if } y > 0, \end{cases} \quad (2.2)$$

where  $\mu > 0$ ,  $W(\cdot)$  is the Lambert function [4], and  $B_y$  is the Bell number defined by (see also [1])

$$B_y = \frac{1}{e} \sum_{d=1}^{\infty} \frac{d^y}{d!} \quad \text{for } y = 0, 1, \dots$$

Let  $\theta := (\beta^\top, \gamma^\top)^\top \in \Theta \subset \mathbb{R}^{p+q}$  denotes the unknown  $k$ -dimensional ( $k = p + q$ ) parameter in the model (2.1). Assume that the observations  $(Y_1, \mathbf{X}_1, \mathbf{Z}_1), \dots, (Y_n, \mathbf{X}_n, \mathbf{Z}_n)$  are generated from (2.1) according to the true parameter  $\theta_0 = (\beta_0^\top, \gamma_0^\top)^\top$  which is unknown. The log-likelihood function is given by

$$\begin{aligned} \ell_n(\theta) = & \sum_{i=1}^n \left\{ \delta_i \log \left[ e^{\gamma^\top \mathbf{Z}_i} + \exp(1 - e^{W(e^{\beta^\top \mathbf{X}_i})}) \right] - \log \left( 1 + e^{\gamma^\top \mathbf{Z}_i} \right) \right. \\ & \left. + (1 - \delta_i) \left[ Y_i \left( \beta^\top \mathbf{X}_i - W(e^{\beta^\top \mathbf{X}_i}) \right) - e^{W(e^{\beta^\top \mathbf{X}_i})} - \log(Y_i!) + \log B_{Y_i} + 1 \right] \right\}, \end{aligned}$$

where  $\delta_i = 1_{\{Y_i=0\}}$ . The MLE  $\hat{\theta}_n := (\hat{\beta}_n^\top, \hat{\gamma}_n^\top)^\top$  of  $\theta_0$  is

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\text{argmax}}(\ell_n(\theta)).$$

This is obtained by solving the score equation

$$\frac{\partial \ell_n(\theta)}{\partial \theta} = 0, \tag{2.3}$$

which can be achieved by nonlinear optimization. In this paper, all estimates are obtained using the R function `maxLik` [19], which implements Newton-type algorithms.

### 2.2. Some Further Notations

To obtain the first and second partial derivatives of the log-likelihood function, we make use of the following intermediate notations:

$$f_i(\theta) := e^{\gamma^\top \mathbf{Z}_i} + \exp(1 - e^{W(e^{\beta^\top \mathbf{X}_i})}),$$

$$g_i(\beta) := \exp(1 - e^{W(e^{\beta^\top \mathbf{X}_i})}), \quad \text{and} \quad k_i(\beta) := 1 + W(e^{\beta^\top \mathbf{X}_i}).$$

The first partial derivatives of the log-likelihood function for the ZIBell model are as follows:

$$\dot{\ell}_{\beta_\ell, i}(\theta) = \sum_{i=1}^n X_{i\ell} \left( -\delta_i \frac{g_i(\beta)}{k_i(\beta) f_i(\theta)} + (1 - \delta_i) \left[ Y_i \left( 1 - \frac{W(e^{\beta^\top \mathbf{X}_i})}{e^{\beta^\top \mathbf{X}_i} k_i(\beta)} \right) - \frac{e^{\beta^\top \mathbf{X}_i}}{k_i(\beta)} \right] \right), \quad \ell = 1, \dots, p,$$

$$\dot{\ell}_{\gamma_\ell, i}(\theta) = \sum_{i=1}^n Z_{i\ell} \left( \delta_i \frac{e^{\gamma^\top \mathbf{Z}_i}}{f_i(\theta)} - \frac{e^{\gamma^\top \mathbf{Z}_i}}{1 + e^{\gamma^\top \mathbf{Z}_i}} \right), \quad \ell = 1, \dots, q.$$

Then, some tedious albeit not difficult algebra shows that

$$\frac{\partial^2 \ell_n(\theta)}{\partial \beta_\ell \partial \beta_m} = \sum_{i=1}^n X_{i\ell} X_{im} \left( \delta_i \frac{f_i(\theta) g_i(\beta) [e^{\beta^\top \mathbf{X}_i} k_i(\beta) + W(e^{\beta^\top \mathbf{X}_i})] + g_i^2(\beta)}{e^{\beta^\top \mathbf{X}_i} f_i^2(\theta) k_i^3(\beta)} \right.$$

$$\left. - (1 - \delta_i) \left[ Y_i \frac{W(e^{\beta^\top \mathbf{X}_i}) - e^{\beta^\top \mathbf{X}_i} k_i^2(\beta)}{e^{2\beta^\top \mathbf{X}_i} k_i^2(\beta)} + \frac{e^{\beta^\top \mathbf{X}_i} k_i^2(\beta) - W(e^{\beta^\top \mathbf{X}_i})}{k_i^3(\beta)} \right] \right), \quad \ell, m = 1, \dots, p,$$

$$\frac{\partial^2 \ell_n(\theta)}{\partial \gamma_\ell \partial \gamma_m} = \sum_{i=1}^n Z_{i\ell} Z_{im} \left( \delta_i \frac{e^{\gamma^\top \mathbf{Z}_i} (f_i(\theta) - e^{\gamma^\top \mathbf{Z}_i})}{f_i^2(\theta)} - \frac{e^{\gamma^\top \mathbf{Z}_i}}{(1 + e^{\gamma^\top \mathbf{Z}_i})^2} \right), \quad \ell, m = 1, \dots, q,$$

$$\frac{\partial^2 \ell_n(\theta)}{\partial \beta_\ell \partial \gamma_m} = \sum_{i=1}^n X_{i\ell} Z_{im} \left( -\delta_i \frac{e^{\gamma^\top \mathbf{Z}_i} g_i(\beta)}{f_i^2(\theta)} \right), \quad \ell = 1, \dots, p, \quad \text{and} \quad m = 1, \dots, q.$$

We set  $S_n(\theta) = \partial \ell_n(\theta) / \partial \theta$ ,  $H_n(\theta) = -\partial^2 \ell_n(\theta) / \partial \theta \partial \theta^\top$ ,  $F_n(\theta) = \mathbb{E}(H_n(\theta))$  and let  $\mathbb{I}_r$  be the identity matrix of order  $r$ .  $H_n(\theta)$  is assumed to be positive definite.

### 3. LARGE-SAMPLE PROPERTIES FOR THE MLE

In this section, we start by giving some regularity conditions that are needed to ensure the large-sample properties of the MLE. Now denote by  $N_n(\delta) = \{\theta \in \mathcal{C} : \|F_n^{\top/2}(\theta - \theta_0)\| \leq \delta^2\}$  a neighborhood of the unknown true parameter  $\theta_0$  for  $\delta > 0$ . To study the consistency and the asymptotic normality of the MLE, we consider the following regularity assumptions.

(C1) The covariates are bounded that is, there exist compact sets  $\mathcal{X} \subset \mathbb{R}^p$  and  $\mathcal{Z} \subset \mathbb{R}^q$  such.

(C2)  $\beta_0$  and  $\gamma_0$  lie in the interior of known compact sets  $\mathcal{B} \subset \mathbb{R}^p$  and  $\mathcal{G} \subset \mathbb{R}^q$ , respectively.

(C3) There exists a positive constant  $c_1$  such that  $n / \mu_{\min}(F_n(\theta_0)) \leq c_1$  for every  $n = 1, 2, \dots$ , where  $c_1$  is a positive constant.

(C4) For all  $\delta > 0$ ,  $\max_{\theta \in N_n(\delta)} \|V_n(\theta) - \mathbb{I}_r\| \rightarrow 0$ , where  $V_n(\theta) := F_n^{-\frac{1}{2}} H_n(\theta) F_n^{\frac{1}{2}}$ .

Conditions C1–C3 are classical in generalized linear regression and zero-inflated regression models (see [5, 14]). Condition C4 is required in the convergence in probability of  $V_n(\theta)$  to  $\mathbb{I}_r$  uniformly in  $\theta \in N_n(\delta)$ , as  $n \rightarrow \infty$ .

In what follows, the space  $\mathbb{R}^r$  of  $r$ -dimensional (column) vectors is provided with the Euclidean norm  $\|\cdot\|_2$  and the space of  $(r \times r)$  real matrices is provided with the norm  $\|A\|_2 := \max_{\|x\|_2=1} \|Ax\|_2$  (for notations simplicity, we use  $\|\cdot\|$  for both norms). Recall that for a symmetric real  $(r \times r)$ -matrix  $A$  with eigenvalues  $\mu_1, \dots, \mu_r$ ,  $\|A\| = \max_i |\mu_i|$  (from now on,  $\mu_{\min}(A)$  and  $\mu_{\max}(A)$  will denote the smallest and largest eigenvalues of  $A$ , respectively). Our first result states that the solution of (2.3) exists, lies in the neighbourhood  $N_n(\delta)$  of  $\theta_0$  when  $n$  is sufficiently large and is consistent for  $\theta_0$ . The following theorem gives the consistency of  $\hat{\theta}_n$ .

**Theorem 3.1.** *Assume that C1–C4 hold. We have:*

- (i)  $\mathbb{P}(S_n(\hat{\theta}_n) = 0) \rightarrow 1$  (asymptotic existence),
- (ii)  $\hat{\theta}_n \xrightarrow{p} \theta_0 = (\beta_0^\top, \gamma_0^\top)^\top$  as  $n \rightarrow \infty$  (weak consistency).

The following theorem gives the asymptotic normality of  $\hat{\theta}_n$ .

**Theorem 3.2.** *Assume that conditions C1–C4 hold. Then,  $F_n^{\frac{T}{2}}(\hat{\theta}_n - \theta_0)$  converges in distribution to the Gaussian vector  $\mathcal{N}(0, \mathbb{I}_r)$ , as  $n \rightarrow \infty$ .*

To prove the asymptotic normality, we show weak convergence of  $F_n^{-\frac{1}{2}}(\hat{\theta}_n - \theta_0)$ . More precisely, a Taylor series expansion of  $S_n := S_n(\theta_0)$  about  $\hat{\theta}_n$  can be used. By Lemma 6.1 in Appendix, it follows that for every  $a \in \mathbb{R}^r$ ,  $a^\top F_n^{-1/2} S_n$  converges in distribution to  $\mathcal{N}(0, 1)$ . An application of Cramer–Wold device concludes the proof. All details are given in Appendix.

## 4. A SIMULATION STUDY

This section contains a brief review of the data generation with a discussion of the various factors that are considered in the simulation study, including the zero-inflation proportions and the sample size. To evaluate the performance of the MLE, we report some simulations results.

### 4.1. The Design of the Experiment

The data are generated from a ZIBell such that:

$$\log(\mu_i) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i}$$

and

$$\text{logit}(\pi_i) = \gamma_1 + \gamma_2 Z_{2i} + \gamma_3 Z_{3i} + \gamma_4 Z_{4i} + \gamma_5 Z_{5i} + \gamma_6 Z_{6i} + \gamma_7 Z_{7i}.$$

Here, we take  $\beta = (0.7, -0.3, 0.5, -0.8, 0.5)^\top$ . We consider successively three values for  $\gamma$ , namely :

- first setting:  $\gamma = (-0.4, 0.9, -0.3, 1, 0.5, -0.5, 0, 0)^\top$ ,
- second setting:  $\gamma = (-0.5, -0.5, 0.9, 1.3, -0.7, -0.3, 0.9, 0)^\top$ ,
- third setting:  $\gamma = (0.2, 0.5, 0.7, -1, 0.8, 1.5, 0, 0)^\top$ ,

with average proportions of zero-inflated 25, 50, and 90%, respectively. The covariates  $X_{i2}, \dots, X_{i5}$  and  $Z_{i3}, \dots, Z_{i7}$  are generated as follows:  $X_{i2} \sim \mathcal{N}(0, 1)$ ,  $X_{i3} \sim \mathcal{U}[1, 3]$ ,  $X_{i4} \sim \mathcal{B}(1, 1.5)$ ,  $X_{i5} \sim \mathcal{N}(-2, 1)$  and  $Z_{i3} \sim \mathcal{N}(-1, 1)$ ,  $Z_{i4} \sim \mathcal{E}(1)$ ,  $Z_{i5} \sim \mathcal{N}(0, 1)$ ,  $Z_{i6} \sim \mathcal{B}(0.8)$ ,  $Z_{i7} \sim \mathcal{U}[2, 5]$ . Linear predictors in  $\log(\mu_i(\beta))$  and  $\text{logit}(\pi_i(\gamma))$  are allowed to share common terms by letting  $X_{i2} = Z_{i2}$ . We consider the following sample sizes:  $n = 500, 1000$ , and  $1500$ . Finally, for each combination of the simulation design parameters (sample size, proportions of zero-inflation, we simulate  $N = 1000$  samples and we calculate the maximum likelihood estimate  $\hat{\theta}$ . Simulations are conducted by using the statistical software **R**. To solve the likelihood equation, we use the package **maxLik** [19] which implements various Newton–Raphson-like algorithms.

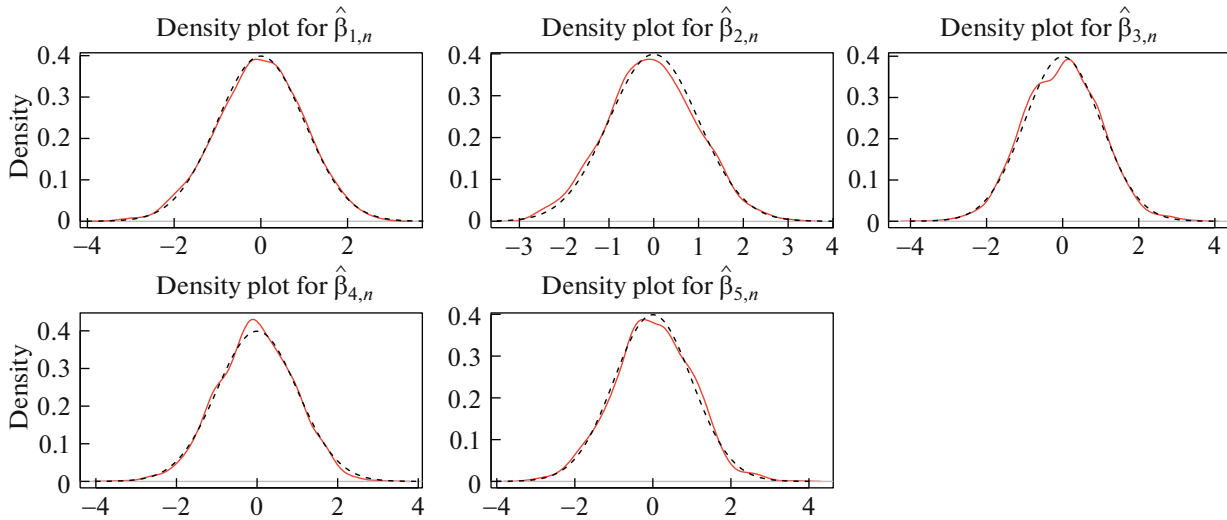


Fig. 1. Density estimates of the  $(\hat{\beta}_{j,n} - \beta_j)/\text{standard error}(\hat{\beta}_{j,n})$ ,  $j = 1, \dots, 5$  with  $n = 1500$  and 25% of zero-inflation.

### 4.2. Numerical Results

For each combination [sample size  $\times$  proportion of zero-inflation] of the simulation parameters, we calculate the average absolute relative bias (as a percentage) of the estimates  $\hat{\beta}_{j,n}$  and  $\hat{\gamma}_{j,n}$  over the  $N$  simulated samples (for example, the absolute relative bias of  $\hat{\beta}_{j,n}$  is obtained as  $N^{-1} \sum_{t=1}^N |(\hat{\beta}_{j,n}^t - \beta_j)|/\beta_j \times 100$ , where  $\hat{\beta}_{j,n}^t$  denotes the MLE estimate of  $\beta_j$  in the  $t$ th simulated sample). We also obtain the mean and root mean square error (RMSE) for the regression parameters. Finally, we provide the empirical coverage probability and average length of 95%-level Wald confidence intervals for the  $\beta_j$  and  $\gamma_k$ . The numerical results are given in Tables 1–3. As in [22], our study shows that the performance of the MLE is quite good, exhibiting small average absolute relative bias and respectable RMSE in all cases considered; that is, the MLE are quite stable and, more importantly, are very close to the true values. The average absolute relative bias and RMSE of all estimators decrease as the sample size increases and that the empirical coverage probabilities are close to the nominal confidence level, even for moderate sample size. Additional observations arise from our study. In particular, for fixed  $n$ , we observe that the performances of  $\hat{\beta}_{j,n}$  remain stable when the average proportion of zero-inflation varies from small to moderate values (here, from 25 to 50%) and deteriorate when zero-inflation achieves higher values, and the performances of  $\hat{\gamma}_{k,n}$ 's improve and then deteriorate as zero-inflation increases. To assess the normal approximation stated in Theorem 3.2, the estimated densities and normal Q–Q plots of  $(\hat{\beta}_{j,n} - \beta_j)/se(\hat{\beta}_{j,n})$  and  $(\hat{\gamma}_{k,n} - \gamma_k)/se(\hat{\gamma}_{k,n})$ ,  $j = 1, \dots, 5, k = 1, \dots, 8$  (with  $se$  denotes the standard error) are plotted in Figs. 3 and 4 for  $n = 1500$ . From these figures, we remark that the estimated density is very close to that of the normal distribution; which confirms the results of Theorem 3.2.

## 5. REAL DATA APPLICATION

In this section, we describe an application of the ZIBell regression model to the analysis a real data example. To this end, we consider data from the USA National Medical Expenditure Survey in 1987 and 1988 (NMES) are considered. NMES data have been widely used in the literature to model the demand for medical care [6]. As an illustration, we use the number of physician visit (**OPF**) as the response variable and the following covariates: number of hospital stays (**hosp**), sex (**gender**: 1 for female, 0 for male), **age** (in years, divided by 10), **marital status**, education level (**school**), perceived health level (“**health1**” if health is poor, 0 otherwise and “**health2**” if health is excellent, 0 otherwise), number of chronic conditions (**numchron**) and a binary variable indicating whether the individual is covered by **medicaid** or not (we code 1 if the individual is covered and 0 otherwise). Table 4 shows the mean and standard deviation of the selected variables.

First, we determine appropriate predictors for zero-inflation modelling. We fit a logistic regression model to the indicators  $1_{\{Y_i=0\}}$ ,  $i = 1, \dots, n$ , considered as the response variable. Note that this is not a model for zero-inflation since some of the 0 may arise from the count distribution. However, we may expect that this rough procedure will still identify a relevant subset of predictors, that will be used in a second step in the logistic model for  $\pi_i$ . Given the very large number of potential predictors, 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 AIC and BIC criterions decreases at each step of this procedure). For comparison, we run the similar analysis using the ZIP model. Results are provided in Table 5.

Secondly, we will check goodness-of-fit. This is an important aspect of regression modeling. We suggest some additional empirical tools and apply them to our model. Note that, in the ZIBell model, the variance of  $Y_i$  is:

$$\text{var}(Y_i|\mathbf{X}_i, \mathbf{Z}_i) = (1 - \pi_i)\mu_i [1 + W(\mu_i) + \mu_i\pi_i].$$

Based on this, we define the Pearson residual for the  $i$ th observation as:

$$r_i = \frac{Y_i - (1 - \hat{\pi}_i)\hat{\mu}_i}{\sqrt{(1 - \hat{\pi}_i)\hat{\mu}_i [1 + W(\hat{\mu}_i) + \hat{\mu}_i\hat{\pi}_i]}}$$

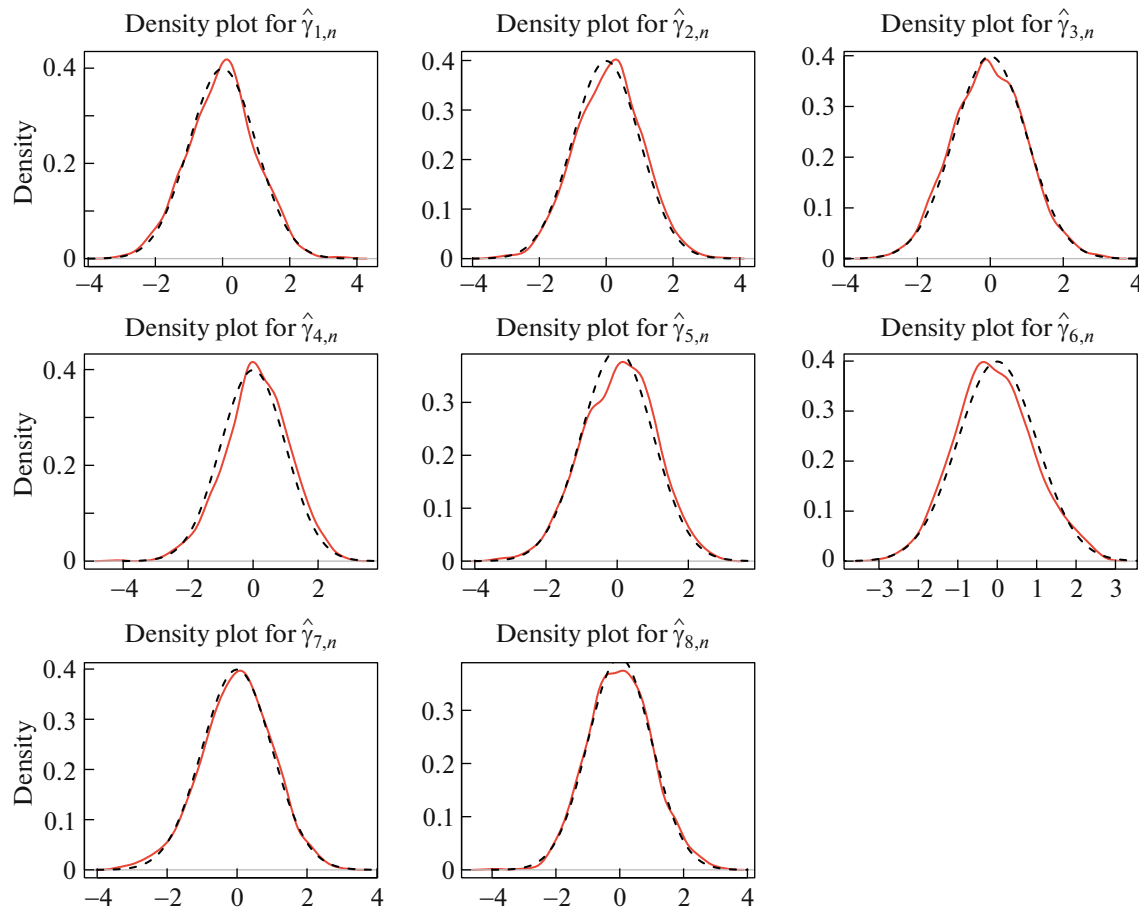
where  $\hat{\mu}_i$  and  $\hat{\pi}_i$  are obtained by replacing  $\beta$  and  $\gamma$  by their estimates in  $\mu_i$  and  $\pi_i$ . If the model is correct, one may expect these residuals to lie within a limited range (e.g., no more than 5% should be greater than 1.96 in absolute value). In our final model, the proportion of residuals greater than 1.96 is 7.6%, which is only slightly greater than 5% and suggests that our model fits reasonably well the data.

We also propose to calculate the Pearson chi-squared statistic

$$X^2 = \sum_{\ell=1}^N \frac{(\tilde{Y}_\ell - (1 - \tilde{\pi}_\ell)\tilde{\mu}_\ell)^2}{(1 - \tilde{\pi}_\ell)\tilde{\mu}_\ell [1 + W(\tilde{\mu}_\ell) + \tilde{\mu}_\ell\tilde{\pi}_\ell]}$$

**Table 1.** Simulation results: **first setting.** Average proportion of zero-inflation = 25% when  $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^\top = (0.7, -0.3, 0.5, -0.8, 0.5)^\top$  and  $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7, \gamma_8)^\top = (-0.4, 0.9, -0.3, 1, 0.5, -0.5, 0, 0)^\top$ . Mean: sample mean, rel. bias: absolute relative bias, RMSE: root of the mean square error, CP: empirical coverage probability of 95%-level confidence intervals

Parameter	$n = 500$				$n = 1000$				$n = 1500$			
	mean	rel. bias	RMSE	CP	mean	rel. bias	RMSE	CP	mean	rel. bias	RMSE	CP
$\hat{\beta}_{1,n}$	0.6825	0.0250	0.3812	0.9465	0.6964	0.0052	0.2601	0.9550	0.6952	0.0068	0.2121	0.9540
$\hat{\beta}_{2,n}$	-0.3026	0.0086	0.0910	0.9404	-0.3021	0.0070	0.0630	0.9440	-0.2988	0.0039	0.0505	0.9500
$\hat{\beta}_{3,n}$	0.5006	0.0013	0.1550	0.944	0.4995	0.0009	0.1062	0.9570	0.5014	0.0027	0.0859	0.9590
$\hat{\beta}_{4,n}$	-0.8037	0.0046	0.1671	0.9354	-0.8026	0.0032	0.1124	0.9670	-0.8028	0.0035	0.0950	0.9410
$\hat{\beta}_{5,n}$	0.4999	0.0002	0.0786	0.9434	0.5004	0.0007	0.0553	0.9540	0.4995	0.0011	0.0446	0.9540
$\hat{\gamma}_{1,n}$	-0.6084	0.5210	2.9055	0.9444	-0.4984	0.2461	1.4467	0.9410	-0.4423	0.1057	1.1230	0.9500
$\hat{\gamma}_{2,n}$	1.0842	0.2046	1.5605	0.9535	0.9433	0.0481	0.2878	0.9500	0.9236	0.0262	0.2240	0.9550
$\hat{\gamma}_{3,n}$	-0.2660	0.1135	1.9257	0.9596	-0.2951	0.0163	0.4273	0.9560	-0.3050	0.0166	0.3373	0.9550
$\hat{\gamma}_{4,n}$	1.3785	0.3785	5.6585	0.9535	1.0576	0.0576	0.3088	0.9450	1.0349	0.0349	0.2354	0.9480
$\hat{\gamma}_{5,n}$	0.5783	0.1567	0.4936	0.9687	0.5307	0.0614	0.2131	0.9480	0.5120	0.0241	0.1666	0.9520
$\hat{\gamma}_{6,n}$	-0.6579	0.3157	2.2398	0.9404	-0.5288	0.0576	0.2478	0.9470	-0.5175	0.0350	0.1905	0.9520
$\hat{\gamma}_{7,n}$	-0.1203	-	4.0704	0.9586	0.0146	-	0.5724	0.9610	0.0154	-	0.4493	0.9400
$\hat{\gamma}_{8,n}$	-0.0182	-	0.6457	0.9556	0.0035	-	0.2588	0.9490	0.0026	-	0.2042	0.9550



**Fig. 2.** Density estimates of the  $(\hat{\gamma}_{k,n} - \gamma_k)/\text{standard error}(\hat{\gamma}_{k,n})$ ,  $k = 1, \dots, 8$  with  $n = 1500$  and 25% of zero-inflation.

where  $N$  is the number of distinct covariate patterns (i.e., observations with the same values for all covariates) and  $\tilde{Y}_\ell$  is the sum of the  $Y_i$ 's over individuals  $i$  with the same covariate pattern. We compare  $X^2$  with the  $\alpha$ -quantile of the  $\chi^2(N - \ell)$ , where  $\ell$  is the number of parameters estimated. The result of the test for our final model is significant ( $p$ -value = 0.029), suggesting that the model does not fit well the data. However, this apparent bad fit is due to outliers with the largest residuals. Removing them from the analysis yields an  $X^2$   $p$ -value equal to 0.132, suggesting that our model fits the data well. We performed a similar analysis for the ZIP model and obtained similar results. However, according to the AIC and BIC criteria in Table 5, our model provides a slightly better fit than the ZIP model.

## 6. DISCUSSION

The ZIBell regression is often used to model count data thanks its simplicity. In this paper, we provide a rigorous basis for maximum likelihood inference in this model. Precisely, we establish consistency and asymptotic normality of the MLE in ZIBell regression. Moreover, our simulation study suggests that the estimator performs well under a wide range of conditions pertaining to sample size and proportion of zero-inflation.

Several other generalizations of ZIBell regression may be developed to account the increasing complexity of experimental data. 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. 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.

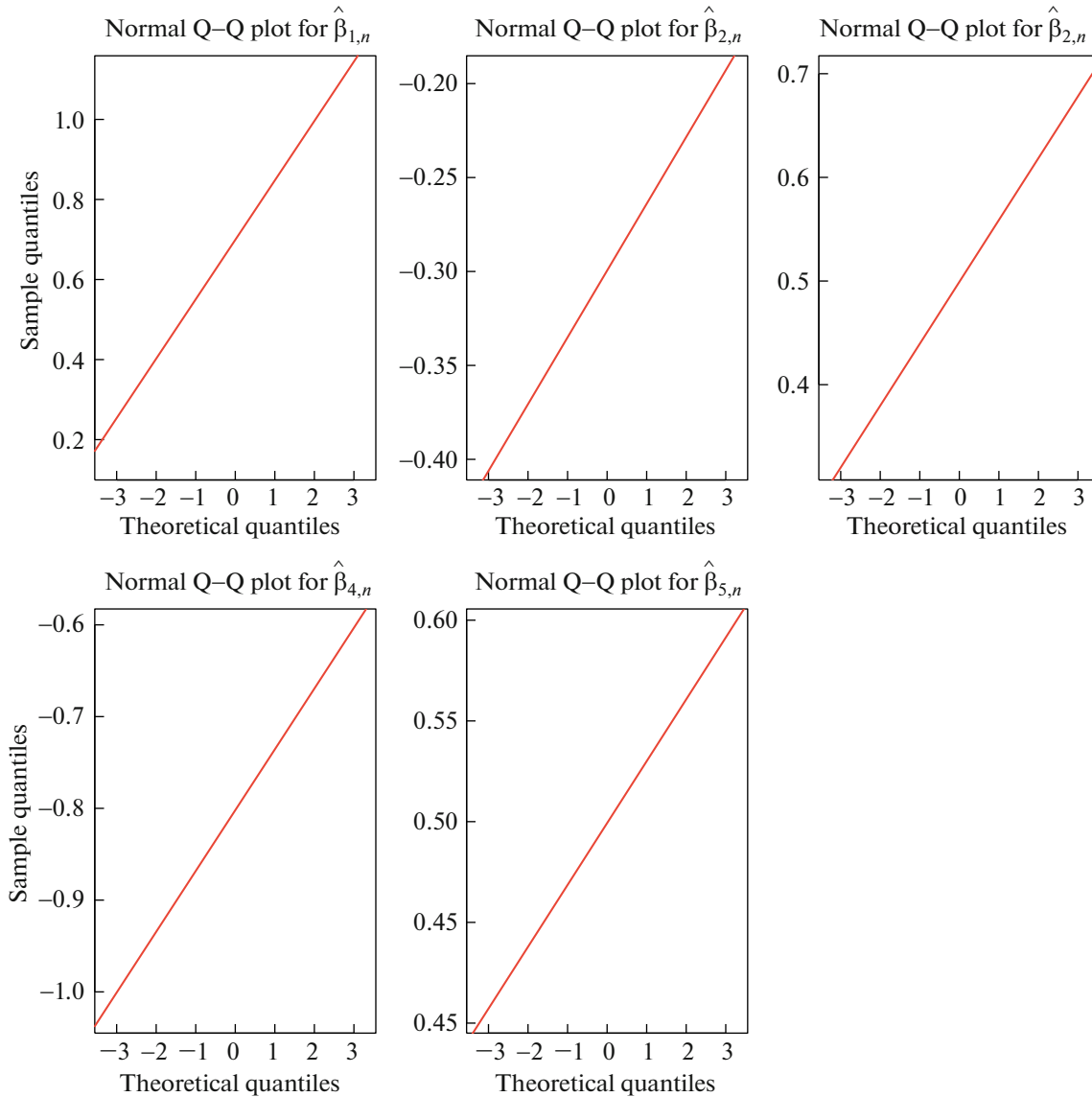


Fig. 3. Normal Q–Q plots for  $(\hat{\beta}_{j,n} - \beta_j)/\text{standard error}(\hat{\beta}_{j,n}), j = 1, \dots, 5$  with  $n = 1500$  and 25% of zero-inflation.

Appendix

PROOFS OF ASYMPTOTIC RESULTS

**Proof of Theorem 3.1.** This proof is inspired by the proof of consistency of the MLE of [29]

(i) Proof of the asymptotic existence of  $\hat{\theta}_n$ .

Let  $\mathcal{D}_n = \ell_n(\theta) - \ell_n(\theta_0)$ . For any  $n, \delta > 0$ , we have the event

$$\mathcal{D}_n < 0, \quad \forall \theta \in \partial N_n(\delta),$$

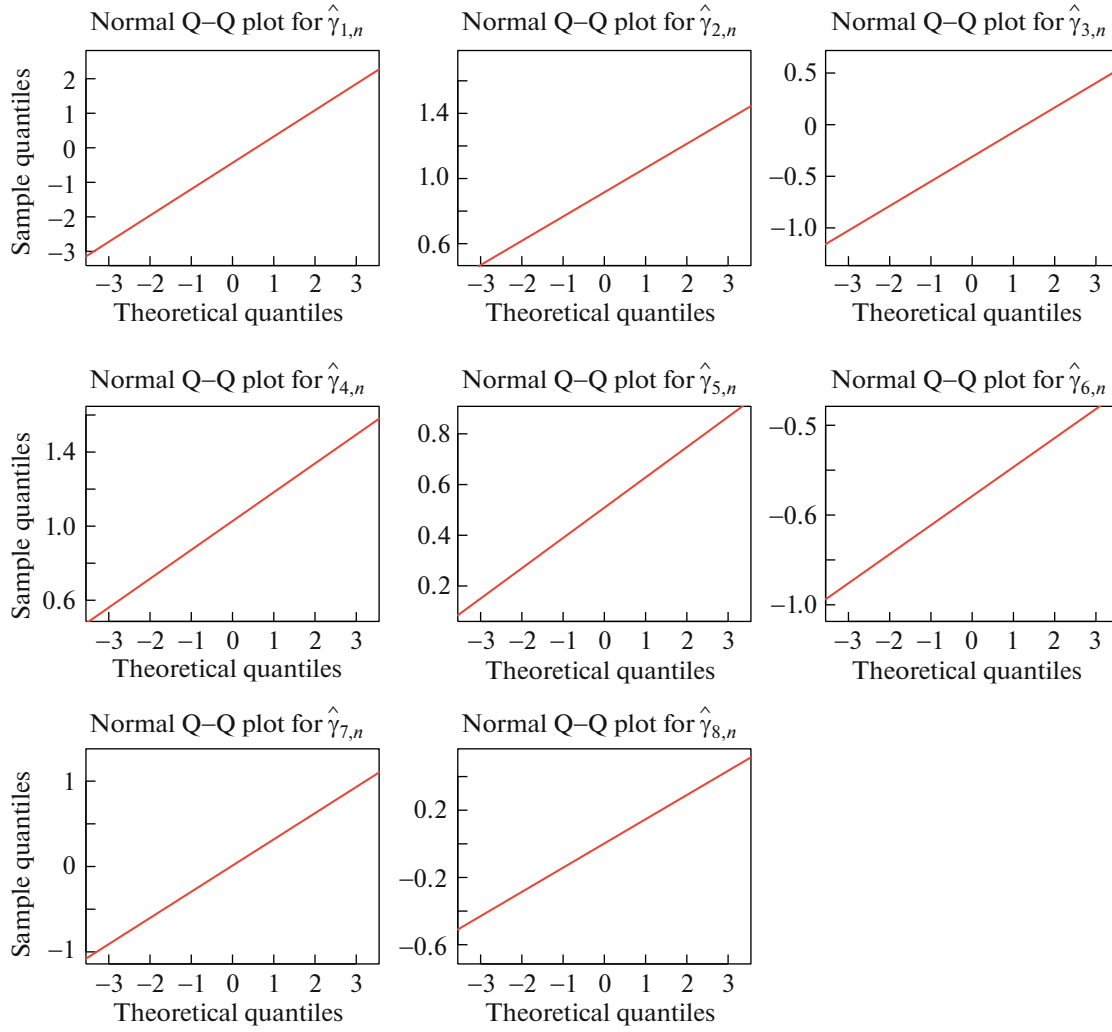
where  $\partial N_n(\delta)$  is the boundary  $\{\theta \in \mathcal{C} : (\theta - \theta_0)^\top F_n(\theta - \theta_0) \leq \delta^2\}$  of  $N_n(\delta)$ .

It is show below that for any  $\eta > 0$ , there exist  $\delta > 0$  and  $n_1 \in \mathbb{N}$  such that,

$$\mathbb{P}(\mathcal{D}_n < 0 \quad \forall \theta \in \partial N_n(\delta)) \geq 1 - \eta, \quad \text{for } n \geq n_1. \tag{6.1}$$

This will imply the existence of a local maximum of  $\ell_n$  in  $N_n(\delta)$ . Due to the positive definition of  $H_n$  and the convexity of  $\mathcal{C}$  there is a most one zero of the score function. This ensures that this maximum is





**Fig. 4.** Normal Q–Q plots for  $(\hat{\gamma}_{k,n} - \gamma_k)/\text{standard error}(\hat{\gamma}_{k,n})$ ,  $k = 1, \dots, 8$  with  $n = 1500$  and 25% of zero-inflation.

global and unique.

Similar to (6.1), we show that for every  $\eta > 0$  there exists  $\delta$  and  $n_1 \in \mathbb{N}$ , such that, for some  $\theta \in \partial N_n(\delta)$ , we have

$$\mathbb{P}(\mathcal{D}_n > 0) \leq \eta \quad \text{for } n \geq n_1.$$

More precisely, a Taylor series expansion at  $\theta_0$  yields

$$\mathcal{D}_n = (\theta - \theta_0)^\top S_n(\theta_0) - \frac{1}{2}(\theta - \theta_0)^\top H_n(\theta_0^*)(\theta - \theta_0),$$

where  $\theta_0^* = \tau\theta + (1 - \tau)\theta_0$  with  $0 \leq \tau \leq 1$  lies between  $\theta$  and  $\theta_0$ . Let  $0 < \lambda < \frac{1}{2}$  and for some  $\theta \in \partial N_n(\delta)$  we have:

$$\begin{aligned} \mathbb{P}(\mathcal{D}_n > 0) &= \mathbb{P}\left((\theta - \theta_0)^\top S_n(\theta) \geq U_n(\theta) \text{ and } U_n(\theta) \geq \lambda\delta^2\right) \\ &+ \mathbb{P}\left((\theta - \theta_0)^\top S_n(\theta) \geq U_n(\theta) \text{ and } U_n(\theta) \leq \lambda\delta^2\right) \leq \mathbb{P}(I) + \mathbb{P}(J), \end{aligned}$$

where  $I = \{(\theta - \theta_0)^\top S_n(\theta) \geq \lambda\delta^2, \text{ for some } \theta \in \partial N_n(\delta)\}$  and  $J = \{U_n(\theta) \leq \lambda\delta^2, \text{ for some } \theta \in \partial N_n(\delta)\}$ , respectively. Assuming  $w_n(\theta) = \frac{1}{\delta}F^{-\frac{1}{2}}(\theta - \theta_0)$ . Then

$$I = \{w_n(\theta)^\top F^{-\frac{1}{2}}S_n(\theta_0) > \lambda\delta\} \subseteq \{\max_{\theta \in N_n(\varepsilon)} |w_n(\theta)^\top F^{-\frac{1}{2}}S_n(\theta_0)| > \lambda\delta\}$$

**Table 2.** Simulation results: **second setting.** Average proportion of zero-inflation = 50% when  $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^\top = (0.7, -0.3, 0.5, -0.8, 0.5)^\top$  and  $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7, \gamma_8)^\top = (-0.5, -0.5, 0.9, 1.3, -0.7, -0.3, 0.9, 0)^\top$ . Mean: sample mean, rel. bias: absolute relative bias, RMSE: root of the mean square error, CP: empirical coverage probability of 95%-level confidence intervals

Parameter	n = 500				n = 1000				n = 1500			
	mean	rel. bias	RMSE	CP	mean	rel. bias	RMSE	CP	mean	rel. bias	RMSE	CP
$\hat{\beta}_{1,n}$	0.6740	0.0371	0.4729	0.9440	0.6922	0.0111	0.3288	0.9510	0.6943	0.0081	0.2685	0.9410
$\hat{\beta}_{2,n}$	-0.3030	0.0100	0.1186	0.9480	-0.2997	0.0011	0.0822	0.9470	-0.3004	0.0012	0.0683	0.9350
$\hat{\beta}_{3,n}$	0.5046	0.0092	0.2029	0.9500	0.4996	0.0007	0.1421	0.9420	0.4999	0.0002	0.1162	0.9430
$\hat{\beta}_{4,n}$	-0.8073	0.0091	0.2267	0.9460	-0.8009	0.0011	0.1560	0.9490	-0.7962	0.0047	0.1284	0.9530
$\hat{\beta}_{5,n}$	0.4983	0.0035	0.1083	0.9510	0.4992	0.0015	0.0746	0.9470	0.5006	0.0012	0.0615	0.9410
$\hat{\gamma}_{1,n}$	-0.5900	0.1799	1.6547	0.9540	-0.5576	0.1152	1.1091	0.9520	-0.5276	0.0552	0.8697	0.9580
$\hat{\gamma}_{2,n}$	-0.5465	0.0930	0.3101	0.9480	-0.5240	0.0479	0.2066	0.9380	-0.5112	0.0225	0.1662	0.9480
$\hat{\gamma}_{3,n}$	0.9630	0.0700	0.5389	0.9440	0.9439	0.0488	0.3618	0.9450	0.9204	0.0227	0.2751	0.9620
$\hat{\gamma}_{4,n}$	1.3901	0.0693	0.3725	0.9430	1.3522	0.0401	0.2505	0.9470	1.3240	0.0185	0.1937	0.9570
$\hat{\gamma}_{5,n}$	-0.7577	0.0824	0.3309	0.9580	-0.7307	0.0438	0.2255	0.9390	-0.7126	0.0180	0.1697	0.9710
$\hat{\gamma}_{6,n}$	-0.3258	0.0860	0.2602	0.9590	-0.3090	0.0301	0.1733	0.9500	-0.3062	0.0206	0.1399	0.9530
$\hat{\gamma}_{7,n}$	0.9704	0.0782	0.6850	0.9460	0.9354	0.0394	0.4479	0.9520	0.9108	0.0119	0.3552	0.9610
$\hat{\gamma}_{8,n}$	0.0036	—	0.2830	0.9560	0.0011	—	0.1971	0.9480	0.0021	—	0.1568	0.9450

$$\subseteq \{ \max_{\|w_n(\theta)\|=1} |w_n(\theta)^\top F^{-\frac{1}{2}} S_n(\theta_0)| > \lambda\delta \} = \{ \|F^{-\frac{1}{2}} S_n(\theta)\| > \lambda\delta \}.$$

As a result, it follows that  $\mathbb{P}(I) \leq \mathbb{P}(\|F^{-\frac{1}{2}} S_n(\theta)\| > \lambda\delta)$ . Based on Theorem 1.5 of Seber and Lee (2012), we have  $\mathbb{E}\|F^{-\frac{1}{2}} S_n(\theta)\|^2 = k$  and Chebyshev’s inequality implies

$$\mathbb{P}(I) \leq \frac{k}{(\lambda\delta)^2}.$$

Now, for event  $J$ , we have

$$\begin{aligned} J &= \left\{ \frac{1}{2}(\theta - \theta_0)^\top H_n(\theta^*)(\theta - \theta_0) \leq \lambda\delta^2 \right\} = \left\{ \frac{1}{2}w_n(\theta)^\top F^{-\frac{1}{2}} H_n(\theta^*) F^{-\frac{1}{2}} w_n(\theta) \leq \lambda \right\} \\ &\subseteq \left\{ \frac{1}{2}\mu_{\min} \left( F^{-\frac{1}{2}} H_n(\theta^*) F^{-\frac{1}{2}} \right) w_n(\theta)^\top w_n(\theta) \leq \lambda \right\} = \left\{ \frac{1}{2}\mu_{\min} \left( F^{-\frac{1}{2}} H_n(\theta^*) F^{-\frac{1}{2}} \right) \leq \lambda \right\}. \end{aligned}$$

Thus  $\mathbb{P}(J) \leq \mathbb{P}(\{\mu_{\min}(F^{-\frac{1}{2}} H_n(\theta^*) F^{-\frac{1}{2}}) \leq 2\lambda\})$ . By Condition C4,  $F_n^{-\frac{1}{2}} H_n(\theta) F_n^{-\frac{1}{2}}$  converges in probability to  $\mathbb{I}_r$  uniformly in  $\theta \in N_n(\delta)$ , as  $n \rightarrow \infty$ . Thus, by [23],  $\mu_{\min}(F^{-\frac{1}{2}} H_n(\theta^*) F^{-\frac{1}{2}})$  converges in probability to 1 uniformly in  $\theta \in N_n(\delta)$ , as  $n \rightarrow \infty$ .

If  $\theta_0^* = \tau\theta + (1 - \tau)\theta_0$  for some  $0 \leq \tau \leq 1$  and  $\theta \in N_n(\delta)$ , then

$$\begin{aligned} \|F_n^{\frac{1}{2}} \theta_n^* - \theta_0\| &= \|F_n^{\frac{1}{2}} (\tau\theta + (1 - \tau)\theta_0 - \theta_0)\| \\ &\leq \tau \|F_n^{\frac{1}{2}} (\theta - \theta_0)\| \leq \delta. \end{aligned}$$

From the above, it follows that  $\mu_{\min}(F^{-\frac{1}{2}} H_n(\theta^*) F^{-\frac{1}{2}})$  converges in probability to 1 as  $n \rightarrow \infty$ , since  $|\mu_{\min}(F^{-\frac{1}{2}} H_n(\theta^*) F^{-\frac{1}{2}}) - 1| \leq \max_{\theta \in N_n(\delta)} |\mu_{\min}(F^{-\frac{1}{2}} H_n(\theta^*) F^{-\frac{1}{2}}) - 1|$ . Therefore for  $n \geq n_1$ ,  $\mathbb{P}$

**Table 3.** Simulation results: **third setting.** Average proportion of zero-inflation = 90% when  $\beta = (\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^\top = (0.7, -0.3, 0.5, -0.8, 0.5)^\top$  and  $\gamma = (\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7, \gamma_8)^\top = (0.2, 0.5, 0.7, -1, 0.8, 1.5, 0, 0)^\top$ . Mean: sample mean, rel. bias: absolute relative bias, RMSE: root of the mean square error, CP: empirical coverage probability of 95%-level confidence intervals

Parameter	n = 500				n = 1000				n = 1500			
	mean	rel. bias	RMSE	CP	mean	rel. bias	RMSE	CP	mean	rel. bias	RMSE	CP
$\hat{\beta}_{1,n}$	0.6010	0.1414	1.1985	0.9400	0.6439	0.0802	0.7439	0.9500	0.6690	0.0443	0.5846	0.9550
$\hat{\beta}_{2,n}$	-0.3045	0.0149	0.3103	0.9430	-0.3084	0.0281	0.1977	0.9290	-0.3007	0.0023	0.1496	0.9450
$\hat{\beta}_{3,n}$	0.5166	0.0332	0.5156	0.9480	0.5086	0.0171	0.3241	0.9590	0.5062	0.0123	0.2517	0.9540
$\hat{\beta}_{4,n}$	-0.8383	0.0479	0.5451	0.9440	-0.8028	0.0035	0.3563	0.9430	-0.8107	0.0134	0.2864	0.9390
$\hat{\beta}_{5,n}$	0.5185	0.0369	0.2721	0.9350	0.5029	0.0058	0.1744	0.9400	0.4998	0.0004	0.1357	0.9530
$\hat{\gamma}_{1,n}$	-0.0342	-1.1712	2.6364	0.9590	0.1231	0.3847	1.5960	0.9530	0.1942	0.0291	1.2507	0.9620
$\hat{\gamma}_{2,n}$	0.5461	0.0922	0.4888	0.9480	0.5098	0.0196	0.2948	0.9400	0.4990	0.0020	0.2297	0.9500
$\hat{\gamma}_{3,n}$	0.8344	0.1921	0.8767	0.9400	0.7579	0.0827	0.4971	0.9540	0.7270	0.0385	0.3874	0.9660
$\hat{\gamma}_{4,n}$	-1.1431	0.1431	0.5226	0.9520	-1.0593	0.0593	0.3054	0.9430	-1.0312	0.0312	0.2330	0.9620
$\hat{\gamma}_{5,n}$	0.9965	0.2457	0.6998	0.9460	0.8774	0.0967	0.4058	0.9460	0.8441	0.0551	0.3057	0.9570
$\hat{\gamma}_{6,n}$	1.7152	0.1435	0.6530	0.9540	1.5786	0.0524	0.3556	0.9540	1.5477	0.0318	0.2707	0.9630
$\hat{\gamma}_{7,n}$	-0.0987	-	1.0670	0.9550	-0.0323	-	0.6445	0.9550	-0.0221	-	0.5022	0.9510
$\hat{\gamma}_{8,n}$	0.0075	-	0.4589	0.9550	-0.0013	-	0.2935	0.9490	-0.0076	-	0.2352	0.9390

**Table 4.** Descriptive summary (NMES data)

Variable	Measurement	Mean	Std. dev.
OFPP	Number of physician visit	5.774	6.759
HOSP	Number of hospitalizations	0.295	0.746
AGE	Age (in years, divided by 10)	7.402	0.633
NUMCHRON	Number of chronic conditions	1.542	1.349
SCHOOL	Educational level (number of years of education)	10.290	3.739
MEDICAID	Number of chronic conditions	0.091	0.288

(there exists  $\theta \in \partial N_n(\delta)$  such that  $\mu_{\min}(F^{-\frac{1}{2}}H_n(\theta^*)F^{-\frac{1}{2}}) \leq 2\lambda \leq \eta/2$ , since  $2\lambda < 1$ . This implies that  $\mathbb{P}(J) \leq \eta/2$ . Finally,

$$\mathbb{P}(\mathcal{D}_n \geq 0) \leq \mathbb{P}(I) + \mathbb{P}(J) \leq \eta, \quad \text{from some } \theta \in \partial N_n(\delta) \tag{6.2}$$

hence the proof of (6.1) and thus, the existence of a unique global maximum of  $\ell_n$  on  $N_n(\delta)$ , which coincides with  $\hat{\theta}_n$ .

(ii) Proof of the consistency of  $\hat{\theta}_n$

$$\begin{aligned} \mu_{\min}(F_n) \|\hat{\theta}_n - \theta_0\|^2 &= (\hat{\theta}_n - \theta_0)^\top \mu_{\min}(F_n) \mathbb{I}_r (\hat{\theta}_n - \theta_0) \\ &\leq (\hat{\theta}_n - \theta_0)^\top F_n (\hat{\theta}_n - \theta_0) = \|F_n^{\frac{1}{2}}(\hat{\theta}_n - \theta_0)\|^2 \leq \lambda^2 \end{aligned}$$

with probability tending to 1 as  $n \rightarrow \infty$ , by (i). By condition C3,  $\mu_{\min}(F_n)$  tend to  $\infty$  as  $n \rightarrow \infty$ .

**Table 5.** NMES data analysis (for each model, only significant covariates at the 5% level are reported)

Parameter	ZIBell model		ZIP model			
	estimate	standard error	<i>p</i> -value	estimate	standard error	<i>p</i> -value
Covariates of $\mu_i$						
intercept	2.1386	0.1304	< 2e-16	2.2591	0.07998	< 2e-16
health1	0.2800	0.0285	<2e-16	0.2768	0.0172	< 2e-16
health2	-0.3523	0.0476	1.41e-13	-0.3070	0.0315	< 2e-16
age	-0.0694	0.0171	4.89e-05	-0.0738	0.0105	2.38e-12
marital status	-0.0596	0.0212	0.0050	-0.0676	0.0131	2.25e-07
gender	0.1434	0.0077	<2e-16	0.1195	0.0046	< 2e-16
Covariates of $\pi_i$						
intercept	-0.6764	0.16789	5.60e-05	-0.6046	0.1331	5.60e-06
health1	-0.4198	0.1769	0.0177	-0.5897	0.1512	9.56e-05
gender	-0.4990	0.1189	2.72e-05	-0.3757	0.0879	1.93e-05
school	-0.1104	0.0156	1.63e-12	-0.0837	0.0116	6.49e-13
medicaid	-0.1521	0.2112	0.4713	-0.1085	0.1593	0.0496
AIC	15 482.94				44 825.78	
BIC	63 926.94				93 269.78	

Therefore  $\|\hat{\theta}_n - \theta_0\|$  converges to 0 with probability tending to 1 as  $n \rightarrow \infty$ , which concludes the proof.  $\square$

**Proof of Theorem 3.2.**

**Lemma 6.1.** *Assume conditions C1–C3 hold. Then  $F_n^{-\frac{1}{2}}S_n$  converges in distribution to  $\mathcal{N}(0, \mathbb{I}_r)$ , as  $n \rightarrow \infty$ , where  $\mathcal{N}(0, \mathbb{I}_r)$  is  $r$ -dimensional normal distribution with mean vector 0 and covariance matrix  $\mathbb{I}_r$ .*

**Proof of Lemma 6.1.** It is sufficient to show that a linear combination of  $F_n^{-\frac{1}{2}}S_n$  converges in distribution to  $\mathcal{N}(0, a^\top a)$  for any vector  $a \in \mathbb{R}^r (a \neq 0)$ . Without loss of generality, we fix  $\|a\| = 1$ . We notice that  $S_n$  can be written as the sum of independent random vectors, namely

$$S_n = \sum_{i=1}^n S_{n,i},$$

where  $S_{n,i} = (S_{n,i,1}, S_{n,i,2}, \dots, S_{n,i,r})$ .

Under conditions C1, C2, and C4, components of  $S_{n,i}$  are bounded by some finite positive constant  $c_2$  that is,  $|S_{n,i,\ell}| < c + 2, \ell = 1, \dots, r$ . Therefore,  $\|S_{n,i}\|^2 < c_3$ .

Further, define independent random variables  $\xi_{n,i}$  by  $\xi_{n,i} := a^\top F_n^{-\frac{1}{2}}S_{n,i}$ .

Since  $\mathbb{E}(\xi_{n,i}) = 0$  and  $\text{var}(\sum_{i=1}^n \xi_{n,i}) = 1$ , let's show that Lindeberg's condition is verified, that is

$$\forall \varepsilon > 0, \sum_{i=1}^n \mathbb{E} \left( \xi_{n,i}^2 1_{\{|\xi_{n,i}| > \varepsilon\}} \right) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Let  $\varepsilon > 0$ , we have:

$$\sum_{i=1}^n \mathbb{E} \left( \xi_{n,i}^2 1_{\{|\xi_{n,i}| > \varepsilon\}} \right) \leq \sum_{i=1}^n \mathbb{E} \left( \|a^\top\|^2 \|F_n^{-\frac{1}{2}}\|^2 \|S_{n,i}\|^2 1_{\{|\xi_{n,i}| > \varepsilon\}} \right)$$

$$\leq \frac{c_1 c_2}{n} \sum_{i=1}^n \mathbb{E} \left( 1_{\{|\xi_{n,i}| > \varepsilon\}} \right) \quad \text{by condition C3.}$$

Now,  $\{|\xi_{n,i}| > \varepsilon\}$  implies that  $\{\mu_{\min}(F_n) < c_3/\varepsilon^2\}$ , therefore,  $1_{\{|\xi_{n,i}| > \varepsilon\}} \leq 1_{\{\mu_{\min}(F_n) < c_3/\varepsilon^2\}}$ . Thus,

$$\sum_{i=1}^n \mathbb{E} \left( \xi_{n,i}^2 1_{\{|\xi_{n,i}| > \varepsilon\}} \right) \leq \frac{c_1 c_2}{n} \sum_{i=1}^n 1_{\{\mu_{\min}(F_n) < c_3/\varepsilon^2\}} = c_1 c_2 1_{\{\mu_{\min}(F_n) < c_3/\varepsilon^2\}}.$$

Under C3,  $\mu_{\min}(F_n) \rightarrow \infty$  as  $n \rightarrow \infty$ . Therefore,

$$\sum_{i=1}^n \mathbb{E} \left( \xi_{n,i}^2 1_{\{|\xi_{n,i}| > \varepsilon\}} \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Then it follows that,  $\forall a \neq \mathbb{R}^r, a^\top F_n^{-\frac{1}{2}} S_n$  converges in distribution to  $\mathcal{N}(0, 1)$  and by Cramer–Wold device  $F_n^{-\frac{1}{2}} S_n$  converges in distribution to  $\mathcal{N}(0, \mathbb{I}_r)$ .

To prove  $F_n^{-\frac{1}{2}}(\hat{\theta}_n - \theta_0) \xrightarrow{\ell} \mathcal{N}(0, \mathbb{I}_r)$ , the Taylor expansion of  $S_n$  at  $\hat{\theta}_n$  can be used. From the mean value theorem for vector valued functions (e.g., Heuser, 1981, p. 278), we have

$$S_n(\hat{\theta} - \theta_0) = F_n(\theta_0 + \tau(\hat{\theta} - \theta_0))(\hat{\theta} - \theta_0)$$

and

$$S_n = \left[ \int_0^1 F_n(\theta_0 + \tau(\hat{\theta} - \theta_0)) d\tau \right] (\hat{\theta} - \theta_0) \quad \text{for } 0 < \tau < 1.$$

By pre-multiplying  $F_n^{\frac{T}{2}}$  and integrating with respect to  $\tau$  on  $[0, 1]$ , we have

$$F_n^{-\frac{1}{2}} S_n = \left[ \int_0^1 V_n(\theta_0 + \tau(\hat{\theta} - \theta_0)) d\tau \right] F_n^{\frac{T}{2}} (\hat{\theta} - \theta_0). \tag{6.3}$$

Under Condition C4 we have

$$\int_0^1 V_n(\theta_0 + \tau(\hat{\theta} - \theta_0)) d\tau \xrightarrow{d} \mathbb{I}_r. \tag{6.4}$$

It follows from Lemma 6.1 and the continuous mapping theorem [2], that Theorem 3.2 holds. □

### ACKNOWLEDGMENTS

Authors are grateful to a referee and the Associate Editor for their comments and suggestions that led substantial improvements of this paper.

### CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

### REFERENCES

1. E. T. Bell, “Exponential polynomials,” *Ann. Math.* **35**, 258–277 (1934a).
2. P. Billingsley, *Convergence of Probability Measures* (John Wiley and Sons, 2013).
3. F. Castellares, S. L. Ferrari, and A. J. Lemonte, “On the Bell distribution and its associated regression model for count data,” *Applied Mathematical Modelling*.  
<https://doi.org/10.1016/j.apm.2017.12.014>
4. R. M. Corless, G. H. Gonnet, D. E. GHare, D. Jeffrey, and D. E. Knuth, “On the LambertW function,” *Adv. Comput. Math.* **5** 329–359 (1996).

5. C. Czado and A. Min, *Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in a Zero-Inflated Generalized Poisson Regression*, Collaborative Research Center 386, Discussion Paper 423 (Ludwig-Maximilians-Universität, München, 2005).
6. P. Deb and P. K. Trivedi, "Demand for medical care by the elderly: a finite mixture approach," *Journal of Applied Econometrics* **12** (3), 313–336 (1997).
7. D. Deng and Y. Zhang, "Score tests for both extra zeros and extra ones in binomial mixed regression models," *Communications in Statistics—Theory and Methods* **44**, 2881–2897 (2015).
8. A. Diop, A. Diop, and J.-F. Dupuy, "Maximum likelihood estimation in the logistic regression model with a cure fraction," *Electronic Journal of Statistics* **5**, 460–483 (2011).
9. A. Diop, A. Diop, and J.-F. Dupuy, "Simulation-based inference in a zero-inflated Bernoulli regression model," *Communications in Statistics—Simulation and Computation* **45** (10), 3597–3614 (2016).
10. E. Dietz and D. Böhning, "On estimation of the Poisson parameter in zero-modified Poisson models," *Computational Statistics and Data Analysis* **34** (4), 441–459 (2000).
11. J.-F. Dupuy, *Statistical Methods for Overdispersed Count Data* (ISTE Press—Elsevier, 2018).
12. Essoham Ali, A simulation-based study of ZIP regression with various zero-inflated submodels, *Communications in Statistics—Simulation and Computation*, <https://doi.org/10.1080/03610918.2022.2025840>
13. F. Eicker, "A multivariate central limit theorem for random linear vector forms," *The Annals of Mathematical Statistics* **37** (6), 1825–1828 (1966).
14. L. Fahrmeir and H. Kaufmann, "Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models," *The Annals of Statistics* **13** (1), 342–368 (1985)
15. J. Feng and Z. Zhu, "Semiparametric analysis of longitudinal zero-inflated count data," *Journal of Multivariate Analysis* **102**, 61–72 (2011).
16. R. A. Fisher, "The negative binomial distribution," *Annals of Eugenics* **11** (1), 182–187 (2011).
17. R. V. Foutz, "On the unique consistent solution to the likelihood equations," *Journal of the American Statistical Association* **72**, 147–148 (1977).
18. D. B. Hall, "Zero-inflated Poisson and binomial regression with random effects: a case study," *Biometrics* **56** (4), 1030–1039 (2000).
19. A. Henningsen and O. Toomet, "maxLik: A package for maximum likelihood estimation in R," *Computational Statistics* **26** (3), 443–458 (2011).
20. J. Hinde and C. G. B. Demetrio, "Overdispersion: models and estimation," *Computational Statistics and Data Analysis* **27**, 151–170 (1998).
21. D. Lambert, "Zero-inflated Poisson regression, with an application to defects in manufacturing," *Technometrics* **34**, 1–14 (2019).
22. A. J. Lemonte, G. Moreno-Arenas, and F. Castellares, "Zero-inflated Bell regression models for count data," *J. Appl. Stat.* **47** (2), 265–286 (2019).
23. R. A. Maller, "Asymptotics of regressions with stationary and nonstationary residuals," *Stochastic Processes and Their Applications* **105** (1), 33–67 (2003).
24. Muhammad Amin, Muhammad Nauman Akram and Abdul Majid, On the estimation of Bell regression model using ridge estimator, *Communications in Statistics—Simulation and Computation*, (2021). <https://doi.org/10.1080/03610918.2020.1870694>
25. P. McCullagh and J. A. Nelder, *Generalized Linear Models*, 2nd ed., Monographs on Statistics and Applied Probability (Chapman and Hall, London, 1989).
26. J.A. Nelder and R. W. M. Wedderburn, "Generalized linear models," *J. Roy. Statist. Soc. Ser. A* **135**, 370–384 (1972).
27. V. T. Nguyen and J.-F. Dupuy, "Asymptotic results in censored zero-inflated Poisson regression," *Communications in Statistics—Theory and Methods* **50** (12), 2759–2779 (2021).
28. R Core Team, *A Language and Environment for Statistical Computing. R Foundation for Statistical Computing* (Vienna, Austria, 2020). <https://www.R-project.org/>
29. M. Ridout, J. Hinde, and C. G. B. Demetrio, "A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives," *Biometrics* **57**(1), 219–223 (2001).
30. G. A. F. Seber and A. J. Lee, *Linear Regression Analysis. Wiley Series in Probability and Statistics* (Wiley, 2012).