



HAL
open science

A novel model for count data: Zero-Inflated Probit Bell model with application to ecological data

Essoham Ali, Kim-Hung Pho

► **To cite this version:**

Essoham Ali, Kim-Hung Pho. A novel model for count data: Zero-Inflated Probit Bell model with application to ecological data. 2024. hal-04448694

HAL Id: hal-04448694

<https://hal.science/hal-04448694>

Preprint submitted on 9 Feb 2024

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 novel model for count data: Zero-Inflated Probit Bell model with application to ecological data

Essoham Ali

Université Bretagne Sud, UMR CNRS 6205, LMBA, F-56000 Vannes, France

E-mail: essoham.ali@univ-ubs.fr

Kim-Hung Pho

Fractional Calculus, Optimization and Algebra Research Group, Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam

Corresponding author: phokimhung@tdtu.edu.vn

Abstract

A novel model for count data is proposed in this work. This new model is briefly called the Zero-Inflated Probit Bell (ZIPBell) model. The ZIPBell model can be used to simulate and analyze traditional count data and can be used well in the simulation and analysis of count data with the phenomenon that the frequency for zero occupies a very high percentage. In the scope of this study, we first provide the general formula, functions or related equations for the ZIPBell model. The results of asymptotic inferences for the ZIPBell model are also given in this work. Maximum likelihood estimator (MLE) is used to check its performance of the proposed model. According to the numerical illustrations via several simulation researches and the analysis of a practical data set, it has been seen that the MLE method has offered correct and reliable inferences. Finally, the inferences and conclusions of this study are mentioned.

Keywords: Estimation, Simulation, Probit, Zero-inflated Bell, Ecological data.

1 Introduction

In scientific research, the topic of regression models is always interested and noticed by scientists and researchers. As we all know that, regression models can be applied in almost all areas of life and has many interesting practical applications. It can be practically applied in applied statistics, applied mathematics, economics, finance, medicine, education, engineering, social sciences and so on. Because the application of the regression model is so great, it is used a lot in the researches of scientists. Therefore in this study we are also interested in regression models.

When it comes to regression models, the first thing we are concerned in is the data for the outcome variable. Depending on the type of data, there are suitable regression models to choose for scientific research. Specifically, in this work we are interested in count data. Count data can be understood simply that it will receive values that are natural numbers for the response variable, and this data is very common and exceedingly often encountered a lot in real data. Applications on using regression models for analysis related to count data can be found in Truong et al. (2021).

For count data, conventional regression models comprise Poisson, Negative Binomial (NB) and Binomial are regularly performed as the first attention in regression analyses. If in that situation the count data occupies a very large number of zeros in the data set, it is often referred to as zero-inflated (ZI) count data. For the ZI count data set, the regression models as mentioned above often do not fit well enough in the analysis (see Lukusa et al., 2017). For such data sets, the ZI count models commonly activate better than conventional regression models. The widespread and crucial ZI count models such as: ZI Poisson (ZIP), ZI Negative Binomial (ZINB), and ZI Binomial (ZIB). These models were first introduced and proposed in the researches of Lambert (1992), Ridout et al. (2001), and Hall (2000), respectively. Moreover, a list of ubiquitous ZI models can be referred in Lukusa et al. (2017).

Recently as we know a novel model that can be employed and executed to ZI count data is ZI Bell (ZIBell) model, which was first introduced by Lemonte et al. (2019). Some interesting studies on the ZIBell model can be seen in: Abolhassani et al. (2020) provided other perspectives on how to construct a ZIBell model. In addition, Algamal et al. (2022) conducted a comparison of the working performance of the Liu and Ridge estimators in this model. Recently, Ali et al. (2022) presented the asymptotic properties of the ZIBell model. The multivariate ZIBell model as well as its inferences and applications can be found in

Lemonte (2022), and so on.

Indeed in some cases when doing regression analysis for real data sets, the ZIBell model may be a better fit for the data set than the more widespread ZI counting models such as ZIP, ZINB and ZIB. No need to say where far away, in the study of Lemonte et al. (2019) have showed that the count of the infected blood cell (IBC) data set would fit the ZIBell model better than some other ZI count models. Therefore we would like to study and learn the problems related to the ZIBell model in this work.

In this paper, we are interested in proposing a new model in the family of the ZI count models, which is shortly called the ZI Probit Bell (ZIPBell) model. This proposed model can be executed to simulate count data or ZI count data. The ZIPBell as well as the ZIBell model has numerous important meanings and applications in practice. Therefore, it is interesting and makes a lot of sense to propose this new regression model in the family of ZI count models. Everything from formulating the general formula for the proposed model, the necessary equations for the estimation, the asymptotic properties, numerical analyzes for the ZIPBell model are presented very fully in this work.

The layout of this job is set up as follows. Section 2 briefly presents the general form the ZIBell model and its estimating equation (EE). The general formual and related function of the proposed model are introduced in Section 3. The results of asymptotic inferences for the ZIPBell model are also given in Section 4. Numerous simulation studies is executed in Section 5. Section 6 performs the IBC data set to the ZIPBell model. Some discussions and conclusions of this work are offered in the last section.

2 The ZIBell Model and Estimating Equations

The probability mass function (PMF) of the ZIBell distribution is offered by (see Ali et al., 2022):

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

where $\mu > 0$ and $\pi \in (0, 1)$. Note also that the original for Eqn. (1) mentioned in the study of Lemonte et al. (2019). If Y follows a ZIBell distribution with the two parameters μ and π , then we desinagate that $Y \sim \text{ZIBell}(\mu, \pi)$. In addition, we should know that $W(\cdot)$ is the

Lambert function, and B_y are the Bell numbers with

$$B_n = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}.$$

We let

$$\log(\mu_i) = \boldsymbol{\beta}^\top x_i \quad \text{and} \quad \text{logit}(\pi_i) = \boldsymbol{\tau}^\top s_i,$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$, and $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_q)^\top$ are parameters need to be evaluated, $x_i^\top = [x_1, x_2, \dots, x_n]^\top$, and $s_i^\top = [s_1, s_2, \dots, s_n]^\top$. The log-likelihood of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top)^\top$ of the ZIBell model is given by (see, Lemonte et al. (2019))

$$\sum_{i=1}^n \left\{ -\log \left(1 + e^{\boldsymbol{\tau}^\top s_i} \right) + J_i \log \left[e^{\boldsymbol{\tau}^\top s_i} + \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right] \right. \\ \left. + (1 - J_i) \left[Y_i \log \left(W(e^{\boldsymbol{\beta}^\top x_i}) \right) - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right] \right\},$$

where $J_i = 1_{\{Z_i=0\}}$. We should note that the MLE technique can be used optimize $\ell(\boldsymbol{\theta})$ to get $\hat{\boldsymbol{\theta}}$ where the estimate of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top)^\top$.

3 The Proposed Model: ZI Probit Bell (ZIPBell) model

3.1 Research motivation

Note that the ZIBell model is created by performing two processes to generate the zeros. A process is usually modeled using the Bell distribution to produce a count (counting zeros), some of which may be zero. The remaining process is frequently modeled with a binary distribution in order to generate zeros. The logit model is often viewed as the easiest of the family of binary models because it is built through exponential functions. It is therefore not too difficult to understand that the logit model is frequently used in processes that use the binary model.

Besides, as we all know that the probit model is established and built on the basis of the normal distribution and this distribution is always considered an important distribution with many good and beautiful properties in statistics. For that reason, of the two steps mentioned above, one uses a binary model. We aim to use the probit model (see Pho and Truong, 2023) to create a new model, which is called for short: the ZI Probit Bell (ZIPBell) model. The proposed and the ZIBell models have numerous important meanings and applications in

practice. Therefore, it is interesting and makes a lot of sense to propose this new regression model in the family of ZI count models.

Besides, the proposal of a new regression model helps researchers, readers or science lovers have more tools and more choices to analyze data. Because data is more and more available, both in quantity and in different forms. Hence, this proposal is very important and makes a lot of sense in the current situation and reality. Everything from formulating the general formula for the proposed model, the necessary equations for the estimation, the asymptotic properties, numerical analyzes for the proposed model are fully provided in the next subsection.

3.2 The general formula

Notice that the variables and two parameters π and μ in this section are similarly defined as in the previous part. The general formula of the ZIPBell model has the following form:

$$P(Y = y|x, s) = \pi I(y = 0) + (1 - \pi) \exp(1 - e^{W(\mu)}) \frac{W(\mu)^y B_y}{y!}, \quad (2)$$

for $y = 0, 1, 2, \dots$, $W(\cdot)$ is the Lambert function, and B_y are the Bell numbers.

3.3 The log-likelihood function

According to (1) then the ZIPBell model has the general likelihood function form:

$$L_n(\boldsymbol{\theta}) = \prod_{i=1}^n [\pi_i + (1 - \pi_i) \exp(1 - e^{W(\mu)})]^{I(Y_i=0)} \left[(1 - \pi_i) \exp(1 - e^{W(\mu)}) \frac{W(\mu)^{Y_i} \cdot B_{Y_i}}{Y_i!} \right]^{I(Y_i>0)} \quad (3)$$

Performing the natural logarithm $L(\boldsymbol{\theta})$ in (3), one has the associated log-likelihood function of the ZIPBell model can then be expressed as follows:

$$\begin{aligned} \log L_n(\boldsymbol{\theta}) &= \sum_{i=1}^n I(Y_i = 0) \log [\pi + (1 - \pi) \exp(1 - e^{W(\mu)})] \\ &\quad + \sum_{i=1}^n I(Y_i > 0) [Y_i \log(W(\mu)) + \log(1 - \pi) - e^{W(\mu)} + 1 + \log(B_{Y_i}) - \log(Y_i!)] \end{aligned} \quad (4)$$

For the sake of simplicity as in the study by Lemonte et al. (2019), formula (5) can be rewritten as follows:

$$\begin{aligned} \log L_n(\boldsymbol{\theta}) &= \sum_{i=1}^n I(Y_i = 0) \log [\pi + (1 - \pi) \exp(1 - e^{W(\mu)})] \\ &\quad + \sum_{i=1}^n I(Y_i > 0) [Y_i \log(W(\mu)) + \log(1 - \pi) - e^{W(\mu)}] \end{aligned} \quad (5)$$

Note that the formula (5) is the general form for the ZIPBell model. In the case of the probability of ZI is modeled by the probit model as follows:

$$\log(\mu_i) = \boldsymbol{\beta}^\top x_i \quad \text{and} \quad \text{probit}(\pi_i) = F(\boldsymbol{\tau}^\top s_i)$$

where F is the C.D.F of $\mathcal{N}(0,1)$, and f is the first derivative of F . Basing on (5) the associated log-likelihood of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top)^\top$ is given by

$$\begin{aligned} \ell(\boldsymbol{\theta}) = & \sum_{i=1}^n \left\{ J_i \log \left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right] \right. \\ & \left. + (1 - J_i) \left[Y_i \log[W(e^{\boldsymbol{\beta}^\top x_i})] + \log[1 - F(\boldsymbol{\tau}^\top s_i)] - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right] \right\}. \end{aligned} \quad (6)$$

and

$$\begin{aligned} P(Y = 0|x, s) &= \pi + (1 - \pi) \exp(1 - e^{W(\mu)}) \\ &= F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \end{aligned}$$

3.4 Estimating Score Function (ESF)

Following to $\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \ell_i(\boldsymbol{\theta})$, where $\ell(\boldsymbol{\theta})$ is given in (6), the ESF is offered as follows:

$$U_{F,n}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{\sqrt{n}} \begin{pmatrix} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}} \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^n S_i(\boldsymbol{\theta}). \quad (7)$$

$$S_i(\boldsymbol{\theta}) = \partial \ell_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} = (S_{i1}^T(\boldsymbol{\theta}), S_{i2}^T(\boldsymbol{\theta}))^T, \quad S_{i1}(\boldsymbol{\theta}) = \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}}, \quad S_{i2}(\boldsymbol{\theta}) = \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \quad i = 1, \dots, n.$$

$$S_{i1}(\boldsymbol{\theta}) = \frac{f(\boldsymbol{\tau}^\top x_i)}{1 - F(\boldsymbol{\tau}^\top s_i)} \cdot \frac{I(Y_i = 0) - \left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right]}{\left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right]}, \quad (8)$$

$$\begin{aligned} S_{i2}(\boldsymbol{\theta}) &= D \cdot e^{W(e^{\boldsymbol{\beta}^\top x_i})} \cdot F(\boldsymbol{\tau}^\top s_i) \cdot \frac{I(Y_i = 0) - \left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right]}{F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right)} \\ &+ \frac{D}{W(e^{\boldsymbol{\beta}^\top x_i})} \cdot \left\{ Y_i - \left(1 - F(\boldsymbol{\tau}^\top s_i) \right) W(e^{\boldsymbol{\beta}^\top x_i}) \cdot e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right\} \end{aligned} \quad (9)$$

with $D = \frac{\partial}{\partial \boldsymbol{\beta}} [W(e^{\boldsymbol{\beta}^\top x_i})]$. The detailed transformations of $S_i(\boldsymbol{\theta})$ is offered in **Appendix**. We need to know that, the way of expressing $S_{i1}(\boldsymbol{\theta})$ in (8) and $S_{i2}(\boldsymbol{\theta})$ in (9) can help us in inferring the asymptotic properties of the estimator $\boldsymbol{\theta}$. Besides, we can easily see that $E[U_{F,n}(\boldsymbol{\theta})] = \mathbf{0}$, $U_{F,n}(\boldsymbol{\theta})$ is an unbiased ESF, and thus the MLE $\widehat{\boldsymbol{\theta}}_F$ of $\boldsymbol{\theta}$ can be obtained by solving $U_{F,n}(\boldsymbol{\theta}) = \mathbf{0}$.

4 Asymptotic Properties

Before proceeding to study the asymptotic properties of $\widehat{\boldsymbol{\theta}}_F$, some necessary conditions are needed to carry out this work. These conditions are similar to the study of Pho (2023), they are briefly expressed as (K1) and (K2). Recall that:

$$U_{F,n}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n S_i(\boldsymbol{\theta}),$$

$S_i(\boldsymbol{\theta}) = \partial \ell_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} = (S_{i1}^T(\boldsymbol{\theta}), S_{i2}^T(\boldsymbol{\theta}))^T$, $S_{i1}(\boldsymbol{\theta}) = \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}}$, $S_{i2}(\boldsymbol{\theta}) = \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}$ $i = 1, \dots, n$. where $S_{i1}(\boldsymbol{\theta})$ and $S_{i2}(\boldsymbol{\theta})$ are given in (8) and (9), respectively.

(K1) For the sake of simplicity, with any column vector \mathbf{a} , we denote $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$, and $E[(S_1(\boldsymbol{\theta}))^{\otimes 2}]$ is positive definite in a neighborhood of the true $\boldsymbol{\theta}$.

(K2) The first derivatives of $U_{F,n}(\boldsymbol{\theta})$ w.r.t $\boldsymbol{\theta}$ exist almost surely in a neighborhood of true $\boldsymbol{\theta}$. Besides, note that in such a neighborhood, the first derivatives are uniformly bounded above by a function of (Y, X, Z) , in the case of the expectations exist.

For the convenience of the following inferences, we call:

$$G_{F,n}(\boldsymbol{\theta}) = -\frac{1}{\sqrt{n}} \frac{\partial U_{F,n}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} = -\frac{1}{n} \sum_{i=1}^n \frac{\partial S_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}.$$

Because every component of $G_{F,n}(\boldsymbol{\theta})$ is a mean of i.i.d random variables, we have:

$$\begin{aligned} E[G_{F,n}(\boldsymbol{\theta})] &= \frac{1}{n} \sum_{i=1}^n E \left\{ E \left[\left(\frac{-\partial S_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right) \middle| Y_i, X_i, Z_i \right] \right\} \\ &= E \left\{ E \left[\left(\frac{-\partial S_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right) \middle| Y_1, X_1, Z_1 \right] \right\} = E \left[\frac{-\partial S_1(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right] = G_F(\boldsymbol{\theta}). \end{aligned}$$

By performing the weak law of large numbers (WLLN), it's not too hard to see that $G_{F,n}(\boldsymbol{\theta}) \xrightarrow{p} G_F(\boldsymbol{\theta})$.

We represent $\widehat{\boldsymbol{\theta}}_F$ as the root of $U_{F,n}(\boldsymbol{\theta}) = 0$. Theorem 1 provides the asymptotic properties of $\widehat{\boldsymbol{\theta}}_F$ and the proof for this theorem is discussed in detail in **Appendix**.

Theorem 1 *Under (K1) and (K2)*

(a) $\widehat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$ as $n \rightarrow \infty$.

(b) $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(0, \Delta_F)$, where $\Delta_F = G_F^{-1}(\boldsymbol{\theta}) Q_F(\boldsymbol{\theta}) [G_F^{-1}(\boldsymbol{\theta})]^T$ and $Q_F(\boldsymbol{\theta}) = E[(S_1(\boldsymbol{\theta}))^{\otimes 2}]$.

Because it's a fact that the variance of the score function is the inverse of the Fisher information matrix, thus $G_F(\boldsymbol{\theta}) = Q_F(\boldsymbol{\theta})$ hence a simple formula for Δ_F can be rewritten as: $\Delta_F = G_F^{-1}(\boldsymbol{\theta})$. The transformations to obtain expressions $G_F(\boldsymbol{\theta})$ and $Q_F(\boldsymbol{\theta})$ are similar in the study of Pho (2023). Due to the simplicity of scientific research and this research objective focusing on simulation studies and analyzing real data sets, the transformations to obtain $G_F(\boldsymbol{\theta})$ and $Q_F(\boldsymbol{\theta})$ are omitted in this work.

5 Numerical study

In this section, we assess finite-sample properties of the MLE with the first case that is both \mathbf{X} and \mathbf{S} are a univariate and the second case is both \mathbf{X} and \mathbf{S} are vectors of covariates.

Case (i) : Both \mathbf{X} and \mathbf{S} are a univariate. We simulate data from a ZIPBell model defined by :

$$\mu(\mathbf{X}) = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2) \quad \text{and} \quad \text{probit}(\pi_i) = \tau_0 + \tau_1 X_1 + \tau_2 S_2$$

where X_1 , X_2 and S_2 are independently drawn from normal $\mathcal{N}(0, 1)$, uniform $\mathcal{U}(1, 3)$ and exponential $\mathcal{E}(1)$ distributions respectively. Parameters β and τ are chosen as $\beta = (-0.25, -0.9, 0.3)^\top$ and $\tau = (-0.55, 0.7, 0)^\top$.

Case (ii) : Both \mathbf{X} and \mathbf{S} are vectors of covariates. Simulation design is essentially similar.

$$\mu(\mathbf{X}) = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5)$$

and

$$\text{probit}(\pi_i) = \tau_0 + \tau_1 S_1 + \tau_2 S_2 + \tau_3 S_3 + \tau_4 S_4 + \tau_5 S_5.$$

X_1, \dots, X_5 are simulated independently from normal $\mathcal{N}(0, 1)$, uniform $\mathcal{U}[1, 3]$, binomial $\mathcal{B}(1, 0.5)$, normal $\mathcal{N}(-2, 1)$, and binomial $\mathcal{B}(2, 0.5)$. Covariates S_2, \dots, S_5 are drawn independently from a normal $\mathcal{N}(0, 1)$, normal $\mathcal{N}(-1, 1)$, binomial $\mathcal{B}(1, 0.5)$ distributions respectively. Linear predictors are allowed to share some common terms by letting $S_1 = X_2$ and $S_2 = X_5$. Parameters β and τ are taken as $\beta = (0.5, -0.9, 0.3, 0.8, 0.1, -0.4)^\top$ and $\tau = (1.5, -0.9, -.2, 0.7, 0, 0)^\top$. Note that the parameter τ is chosen in order to obtain the desired zero rate which is 63%.

For the two cases, our simulation results are based on $N = 1000$ simulated samples. We consider the following sample sizes: $n = 200, 500, 1000$ and 2000 . For each estimator, we

report the average bias, empirical standard deviation (SD), average standard error (SE), empirical root mean square error (RMSE) and empirical coverage probability (CP) of 95%-level confidence intervals. Simulations are conducted using the statistical software R (R Core Team 2018), and all estimates are obtained by using the *maxLik* function via the Newton-Raphson algorithm (see Pho 2022).

Results. Tables 1 and 2 present the simulation results for the two simulated cases. In general, the biases and RMSE are closer to zero and CPs are closer to 0.95 as the sample size increases. In most cases, the biases of $\widehat{\beta}_{0,n}$, $\widehat{\beta}_{1,n}$, $\widehat{\beta}_{2,n}$ are closer to zero even for small sample sizes. Similarly, the CP of $\widehat{\beta}_{0,n}$ and $\widehat{\beta}_{1,n}$ in case (i) needs a bigger sample size (n=500) to achieve the nominal coverage. Overall, the bias is generally slightly larger for the $\widehat{\tau}$ than for the $\widehat{\beta}$. The empirical coverage probabilities are close to the nominal confidence level. In addition, from these tables we can see that, the bias, SE SD and RMSE of all estimators decrease as sample size increases.

We assess empirically the Gaussian approximation stated in Theorem (1) by plotting normal Q-Q plots of the estimates. Figures (1) and (2) provide plots for case (ii) with n = 2000 and an average sample proportion of the ZI equal to 63% (plots for the other simulation scenarios are similar and are thus omitted). From these plots, the distribution of the MLE in the ZIPBell model is reasonable approximated by the Gaussian distribution.

Table 1: Simulation results (case 1). Note that all results are based on $N = 1000$ simulated samples.

		$\hat{\beta}_n$			$\hat{\tau}_n$		
		$\hat{\beta}_{0,n}$	$\hat{\beta}_{1,n}$	$\hat{\beta}_{2,n}$	$\hat{\tau}_{0,n}$	$\hat{\tau}_{1,n}$	$\hat{\tau}_{2,n}$
200							
	bias	-0.0182	0.0009	0.0053	-0.0558	0.0633	-0.0017
	SD	0.3602	0.0938	0.1626	0.3080	0.2572	0.1994
	SE	0.3557	0.0946	0.1597	0.3143	0.2518	0.1885
	RMSE	0.5065	0.1332	0.2279	0.4435	0.3653	0.2743
	CP	0.9530	0.9480	0.9420	0.9580	0.9550	0.9660
500							
	bias	0.0140	0.0028	-0.0086	-0.0105	0.0230	-0.0118
	SD	0.2195	0.0576	0.0973	0.1674	0.1326	0.1099
	SE	0.2209	0.0580	0.0992	0.1678	0.1332	0.1063
	RMSE	0.3116	0.0818	0.1392	0.2372	0.1893	0.1533
	CP	0.9500	0.9500	0.9470	0.9600	0.9550	0.9640
1000							
	bias	0.0053	0.0007	-0.0034	-0.0136	0.0136	-0.0002
	SD	0.1485	0.0398	0.0668	0.1187	0.0941	0.0714
	SE	0.1551	0.0405	0.0695	0.1161	0.0921	0.0718
	RMSE	0.2148	0.0568	0.0964	0.1666	0.1324	0.1012
	CP	0.9570	0.9520	0.9590	0.9480	0.9500	0.9590
2000							
	bias	-0.0030	0.0001	0.0005	-0.0019	0.0051	-0.0025
	SD	0.1105	0.0292	0.0488	0.0781	0.0639	0.0483
	SE	0.1096	0.0285	0.0491	0.0806	0.0637	0.0499
	RMSE	0.1556	0.0408	0.0692	0.1122	0.0904	0.0695
	CP	0.9470	0.9390	0.9560	0.9560	0.9470	0.9570

Table 2: Simulation results (case 2). Note that all results are based on $N = 1000$ simulated samples.

		$\hat{\beta}_n$						$\hat{\tau}_n$							
		$\hat{\beta}_{1,n}$	$\hat{\beta}_{2,n}$	$\hat{\beta}_{3,n}$	$\hat{\beta}_{4,n}$	$\hat{\beta}_{5,n}$	$\hat{\beta}_{6,n}$	$\hat{\tau}_{1,n}$	$\hat{\tau}_{2,n}$	$\hat{\tau}_{3,n}$	$\hat{\tau}_{4,n}$	$\hat{\tau}_{5,n}$	$\hat{\tau}_{6,n}$		
II	200	bias	-0.0444	-0.0094	0.0013	0.0183	0.0015	-0.0011	0.1061	-0.0548	-0.0164	0.0621	0.0047	-0.0176	
		SD	0.4511	0.0909	0.1564	0.2337	0.0887	0.1356	0.5041	0.3153	0.2066	0.1894	0.1497	0.3470	
		SE	0.4359	0.0935	0.1524	0.2288	0.0894	0.1289	0.4752	0.2935	0.2036	0.1765	0.1418	0.3431	
		RMSE	0.6287	0.1307	0.2183	0.3275	0.1259	0.1871	0.7006	0.4341	0.2904	0.2662	0.2062	0.4882	
		CP	0.9439	0.9550	0.9389	0.9389	0.9540	0.9399	0.9600	0.9409	0.9590	0.9530	0.9540	0.9580	
		500	bias	-0.0199	-0.0007	0.0019	0.0017	-0.0019	0.0038	0.0522	-0.0322	-0.0080	0.0207	0.0048	-0.0178
			SD	0.2674	0.0555	0.0915	0.1377	0.0538	0.0770	0.2908	0.1789	0.1259	0.1010	0.0838	0.2123
			SE	0.2634	0.0546	0.0923	0.1366	0.0529	0.0771	0.2797	0.1737	0.1205	0.1012	0.0823	0.2036
			RMSE	0.3757	0.0779	0.1300	0.1939	0.0755	0.1090	0.4068	0.2514	0.1744	0.1444	0.1175	0.2946
			CP	0.9479	0.9529	0.9519	0.9519	0.9519	0.9509	0.9439	0.9459	0.9499	0.9549	0.9469	0.9379
		1000	bias	-0.0131	-0.0020	0.0015	0.0054	-0.0011	-0.0009	0.0129	-0.0038	-0.0036	0.0113	0.0006	0.0033
			SD	0.1798	0.0390	0.0643	0.0942	0.0376	0.0543	0.1901	0.1216	0.0821	0.0684	0.0571	0.1435
			SE	0.1831	0.0378	0.0642	0.0948	0.0368	0.0536	0.1937	0.1206	0.0838	0.0699	0.0572	0.1415
			RMSE	0.2570	0.0544	0.0908	0.1337	0.0526	0.0763	0.2716	0.1713	0.1173	0.0984	0.0807	0.2016
			CP	0.9548	0.9478	0.9578	0.9378	0.9398	0.9498	0.9568	0.9528	0.9498	0.9578	0.9578	0.9478
	2000	bias	0.0015	-0.0001	-0.0016	0.0009	0.0003	-0.0003	0.0176	-0.0087	-0.0056	0.0080	-0.0021	0.0011	
		SD	0.1306	0.0264	0.0452	0.0657	0.0254	0.0370	0.1312	0.0877	0.0578	0.0490	0.0410	0.0936	
		SE	0.1281	0.0263	0.0449	0.0666	0.0258	0.0375	0.1359	0.0848	0.0589	0.0489	0.0399	0.0994	
		RMSE	0.1829	0.0372	0.0637	0.0935	0.0361	0.0527	0.1897	0.1222	0.0827	0.0697	0.0573	0.1365	
		CP	0.9520	0.9350	0.9510	0.9540	0.9530	0.9480	0.9540	0.9390	0.9470	0.9570	0.9490	0.9680	

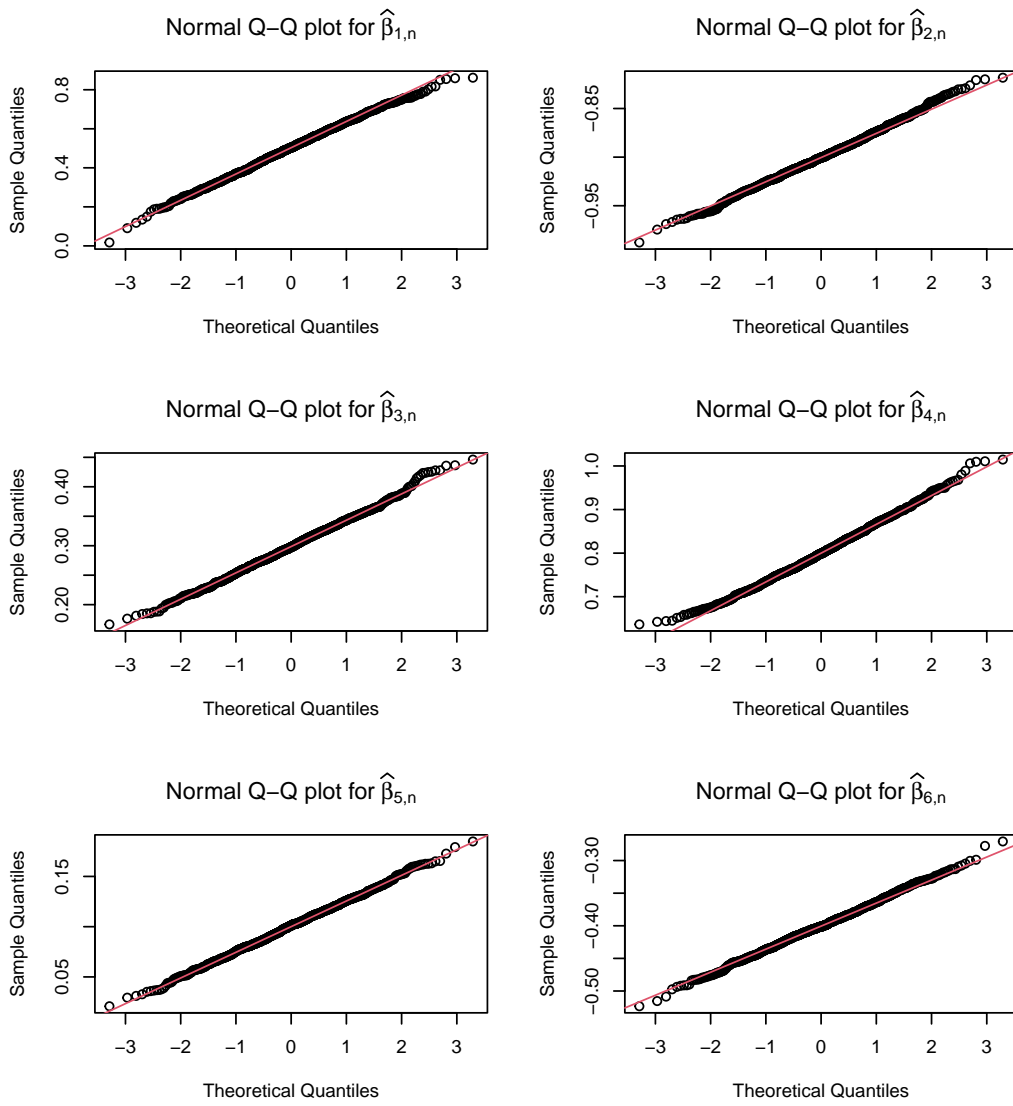


Figure 1: Normal Q-Q plots of $\hat{\beta}_{1,n}, \dots, \hat{\beta}_{6,n}$ for case (ii) with $n = 2000$ and average sample proportion of the ZI equal to 0.63.

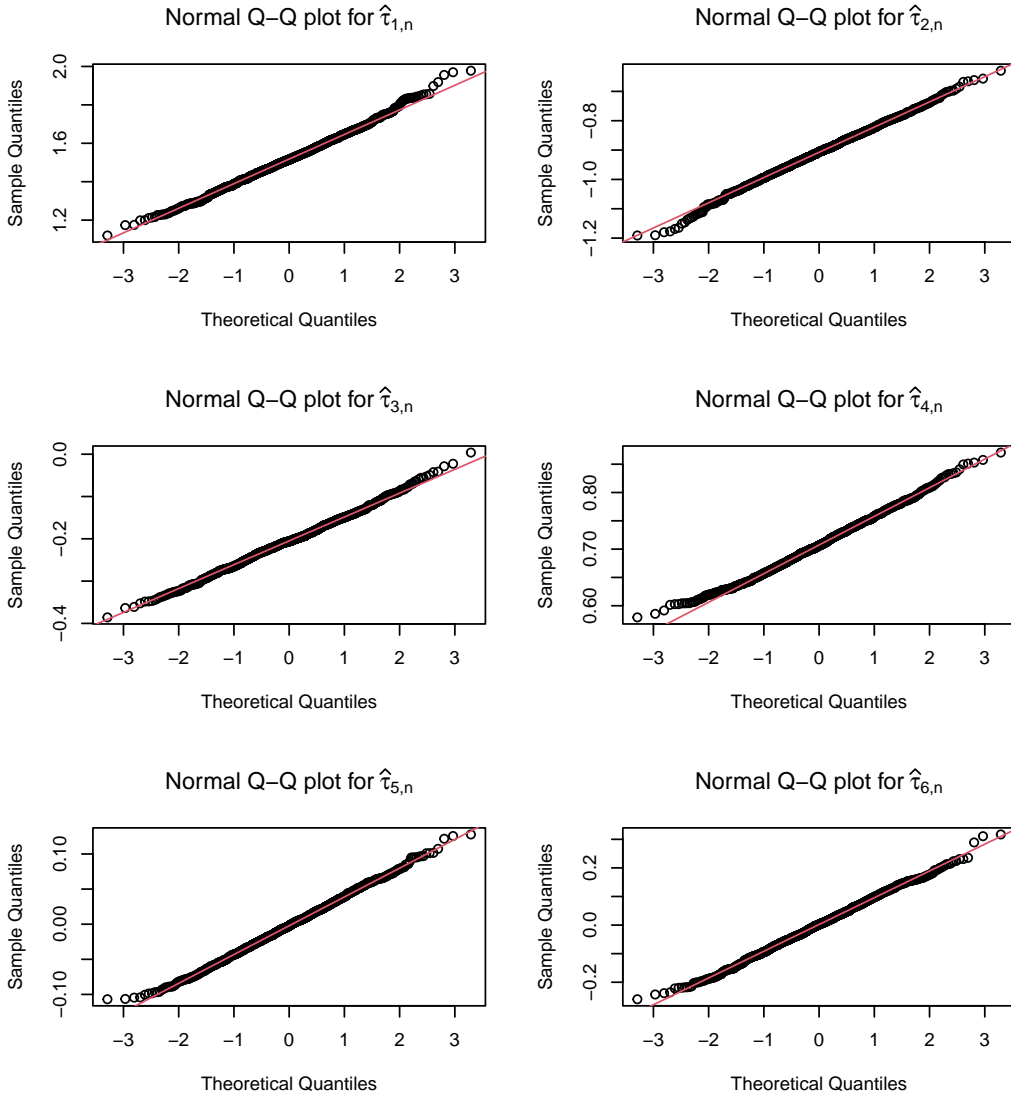


Figure 2: Normal Q-Q plots of $\hat{\tau}_{1,n}, \dots, \hat{\tau}_{6,n}$ for case (ii) with $n = 2000$ and average sample proportion of the ZI equal to 0.63.

6 Real Data Analysis

In this section, we describe an application of the ZIPBell model to the analysis of fish species occurrence records (FSOR) for surveys that were collected at different times in most of the aquatic ecosystems of Uganda. The dataset used here comes from the "Fish Species Occurrence Records for Uganda Mobilized Observation Archives", from the unpublished Mobilized Uganda Archives, available at <https://www.gbif.org>. The experiment is conducted on a single species, *Lates niloticus*. *Lates niloticus*, commonly known as "freshwater captain" or "Nile perch", is a fish whose flesh quality and rapid growth make it a highly prized species. Therefore, this species is very unique and has many practical meanings.

Indeed in recent years, fish farmers have turned their attention to this species, which also

allows natural control of fish density. The fish is also world-famous, having been introduced to the Ugandan part of Lake Victoria in the 1950 to stimulate a declining fishery and thus contribute to the region economic development Lowe (1997). Stocks of *Lates niloticus* increased rapidly during the 1970, leading to the decline and even, in some areas, the complete disappearance of certain native species, including many endemic cichlids Witte et al. (1992). Hence it is interesting to study or learn about this issue in this study.

Next we discuss a little about how to collect the data under investigation. Note that, data were collected in the main river basins, in particular lakes Kwanja, Kyoga, Nabugabo, Nakuwa, Nawampasa, Victoria and the Nile. This type of species is found in many types of freshwater, preferring warm tropical waters ($27^{\circ}\text{N} - 7^{\circ}\text{S}$) where they reach large sizes and are present in high densities. They are silvery in color with a blue tinge. They have a distinctive dark black eye with a bright yellow outer ring and generally weigh between 2 and 4 kg. *Lates niloticus* average 85 to 100 cm in length, but can reach 193 cm. Furthermore, females are generally larger than males.

Several measures were reported in this study, such as the number of *Lates niloticus* as the target variable. Informations useful for developing occurrences is also reported such as GPS coordinates (latitude and longitude) on the actual locations or sites where the fish were captured. The fish were captured using fishing gears such as gillnets, hooks and beach seines at specific sites of different waterbodies. A detailed description of the data can be found in Sidumo et al. (2023). Here we address the question of the influence of environmental factors on fish reproduction. Let Y_i be the number of *Lates niloticus* species caught and we consider this as the response variable in this regression analysis. However, we can observe that most of the count data lies between 0 and 50 suggests that that a ZIPBell regression model should be used instead.

The results are fully presented in Table 3. For this example, two models: ZIPBell and ZIBell, were used to study the influence of environmental variables on the abundance and distribution of the species *Lates niloticus* and to perform the MLE approach to estimate the parameters of these models. Besides, AIC and BIC are introduced to compare the fitting effects of these two models. In general, as we know the lower the AIC and BIC values, the better the fitting effect of the model. It is also important to note that the estimation results obtained for the ZIBell and ZIPBell models are not very different. To be more specific, then they are only slightly superior to each other.

Table 3: Ecological data analysis results.

	Proposed ZIPBell			ZIBell		
	Estimates	S.E.	<i>p</i> -values	Estimates	S.E.	<i>p</i> -values
β						
intercept	2.1487	0.0741	< 0.0001	2.182	0.0744	< 0.0001
waterBody	0.4444	0.0738	< 0.0001			
Latitude	-0.0087	0.0425	0.8370	0.1566	0.032	< 0.0001
Longitude	0.0265	0.0025	< 0.0001	0.0239	0.0024	< 0.0001
τ						
intercept	-1.4064	0.1284	< 0.0001	-2.7474	0.5583	< 0.0001
waterBody	1.0779	0.2882	0.0002	2.4628	0.7335	0.0008
Latitude	-0.6281	0.1924	0.0011	-1.4462	0.4904	0.0031
Longitude				0.0096	0.0205	0.6401
AIC		-5439.436			-5406.259	
BIC		-2485.436			-2452.259	

If we focus our comments on the models under consideration, based on the AIC and BIC results provided in the last two rows of Table 3. (Indeed, the lower the AIC and BIC values, the better the corresponding model). The effects of spatial factors (latitude, longitude) are used as a measure of spatial location. The environmental factors favoring fish reproduction in Uganda are waterBody, Latitude, and Longitude. Besides, in rivers and shallow lakes, water temperature is generally little different from air temperature. In deep lakes, its distribution in the water mass is less homogeneous.

In addition, Latitude and longitude are the two main factors determining water temperature, along with the local influence of atmospheric circulation. The latter explains the abundant production of the species in the lakes of Uganda. The ZIPBell model identifies almost the same set of significant covariates (waterBody and Longitude are significant, while Latitude is not). The magnitude and sign of significant parameters are similar in both models, leading to similar interpretations. Among the environmental parameters studied, temperature is one of the main parameters explaining spatial variations in *Lates niloticus*. It should be noted that the ZIBell model and the proposed ZIPbell model are based on the same specifications for the susceptibility probability π_i , when it comes to interpreting the τ parameter values in these models.

7 Discussions and Conclusions

This study is significant in many ways, first in terms of contributing to the literary treasure, we have proposed and introduced an additional tool for data analysis. Specifically, we first proposed the ZIPBell model. This can be seen as a new, meaningful and interesting regression model, which can work well for simulation and regression analysis for ZI count data sets. In this day and age, almost any profession or field has data, and therefore the addition of a new regression model is proposed, which means more tools to analyze the data in the most scientific way. From there we will provide the most accurate conclusions. In terms of practical applications, the ZIPBell model has shown great potential to be applied to real data sets. In this study, we have also demonstrated that the proposed model: ZIPBell is more suitable for the FSOR data set than the traditional ZIBell model.

In general, this work proposed a novel model for count data. This new model is briefly called the ZIPBell model. The ZIPBell model can be used to simulate and analyze traditional count data and can be used well in the simulation and analysis of count data with the phenomenon that the frequency for zero occupies a very high percentage. We provided the general formula, functions or related equations for the ZIPBell model. The results of asymptotic inferences for the ZIPBell model were also given in this work. MLE was used to check its performance of the proposed model. According to the numerical illustrations via several simulation researches and the analysis of a practical data set, it had been seen that the MLE method has offered correct and reliable inferences.

References

- Abolhassani, A., Prates, M. O., Castellares, F., & Mahmoodi, S. (2020). Zero-inflated Bell scan: A more flexible spatial scan statistic. *Spatial Statistics*, 36, 100433.
- Algamal, Z. Y., Lukman, A. F., Abonazel, M. R., & Awwad, F. A. (2022). Performance of the Ridge and Liu Estimators in the zero-inflated Bell Regression Model. *Journal of Mathematics*, 2022, 9503460.
- Ali, E., Diop, M. L., & Diop, A. (2022). Statistical Inference in a Zero-Inflated Bell Regression Model. *Mathematical Methods of Statistics*, 31(3), 91-104.
- Foutz, R.V. (1977). On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association*, 72(357), 147-148.
- Hall, D.B. (2000), Zero-inflated Poisson and binomial regression with random effects: a case study, *Biometrics*, 56(4), 1030-1039.
- Hassan, A., Ahmad, I. S., & Ahmad, P. B. (2022). Zero-inflated modified Borel-Tanner regression model for count data. *Austrian Journal of Statistics*, 51(2), 28-39.
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1), 1-14.
- Lemonte, A. J. (2022). Multivariate zero-inflated Bell distribution and its inference and applications. *Applied Mathematical Modelling*, 103, 543-556.
- Lemonte, A. J., Moreno-Arenas, G., & Castellares, F. (2019). Zero-inflated Bell regression models for count data. *Journal of Applied Statistics*, 47(2), 265-286.
- Lowe-McConnell, R. (1997). EAFRO and after: a guide to key events affecting fish communities in Lake Victoria(East Africa). *South African Journal of Science*, 93(11), 570-573.
- Lukusa, T. M., Lee, S. M., & Li, C. S. (2017). Review of zero-inflated models with missing data. *Current Research in Biostatistics*, 7(1), 1-12.
- Pho, K. H., Ly, S., Ly, S., and Lukusa, T. M. (2019), Comparison among Akaike information criterion, Bayesian information criterion and Vuong's test in model selection: A case study of violated speed regulation in Taiwan, *Journal of Advanced Engineering and Computation*, 3(1), 293-303.
- Pho, K. H. (2022). Improvements of the Newton–Raphson method. *Journal of Computa-*

- Pho, K. H. (2023). Zero-inflated probit Bernoulli model: a new model for binary data. *Communications in Statistics-Simulation and Computation*, 1-21.
- Pho, K. H., & Truong, B. C. (2023). Pearson chi-squared and unweighted residual sum of square tests of fit for a probit model. *Communications in Statistics-Simulation and Computation*, 1-16.
- Ridout, M., Hinde, J., & Demétrio, C. G. (2001). A score test for testing a zero-inflated Poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*, 57(1), 219-223.
- Sidumo, B., Sonono, E., & Takaidza, I. (2023). Count Regression and Machine Learning Techniques for Zero-Inflated Overdispersed Count Data: Application to Ecological Data. *Annals of Data Science*, 1-15.
- Truong, B. C., Pho, K. H., Dinh, C. C., & McAleer, M. (2021). Zero-inflated poisson regression models: Applications in the sciences and social sciences. *Annals of Financial Economics*, 16(02), 2150006.
- Witte, F., Goldschmidt, T., Goudswaard, P. C., Ligtoet, W., Van Oijen, M. J. P., & Wanink, J. (1992). Species extinction and concomitant ecological changes in Lake Victoria. *Netherlands Journal of Zoology*, 42(2), 214-232.

Appendix

The log-likelihood

$$\begin{aligned}
\log L_n(\theta) &= \sum_{i=1}^n I(Y_i = 0) \log [\pi + (1 - \pi) \exp(1 - e^{W(\mu)})] \\
&\quad + \sum_{i=1}^n I(Y_i > 0) \log \left[(1 - \pi) \exp(1 - e^{W(\mu)}) \frac{W(\mu)^{Y_i} \cdot B_y}{Y_i!} \right] \\
&= \sum_{i=1}^n I(Y_i = 0) \log [\pi + (1 - \pi) \exp(1 - e^{W(\mu)})] \\
&\quad + \sum_{i=1}^n I(Y_i > 0) \log \left[\exp(1 - e^{W(\mu)}) \frac{W(\mu)^{Y_i} \cdot B_y}{Y_i!} \right] + \sum_{i=1}^n I(Y_i > 0) \log (1 - \pi) \\
&= \sum_{i=1}^n I(Y_i = 0) \log [\pi + (1 - \pi) \exp(1 - e^{W(\mu)})] \\
&\quad + \sum_{i=1}^n I(Y_i > 0) [\log \exp(1 - e^{W(\mu)}) + Y_i \log(W(\mu)) + \log(B_y) - \log(Y_i!)] \\
&\quad + \sum_{i=1}^n I(Y_i > 0) \log (1 - \pi) \\
&= \sum_{i=1}^n I(Y_i = 0) \log [\pi + (1 - \pi) \exp(1 - e^{W(\mu)})] \\
&\quad + \sum_{i=1}^n I(Y_i > 0) [Y_i \log(W(\mu)) + 1 - e^{W(\mu)} + \log(B_y) - \log(Y_i!) + \log(1 - \pi)] \\
&= \sum_{i=1}^n I(Y_i = 0) \log [\pi + (1 - \pi) \exp(1 - e^{W(\mu)})] \\
&\quad + \sum_{i=1}^n I(Y_i > 0) [Y_i \log(W(\mu)) + \log(1 - \pi) - e^{W(\mu)} + 1 + \log(B_y) - \log(Y_i!)]
\end{aligned}$$

For the sake of simplicity and similarity in the study by Lemonte et al. (2019), the above can be rewritten as follows:

$$\begin{aligned}
\log L_n(\theta) &= \sum_{i=1}^n I(Y_i = 0) \log [\pi + (1 - \pi) \exp(1 - e^{W(\mu)})] \\
&\quad + \sum_{i=1}^n I(Y_i > 0) [Y_i \log(W(\mu)) + \log(1 - \pi) - e^{W(\mu)}]
\end{aligned}$$

In the case of the probability of ZI is modeled by the probit model as follows:

$$\log(\mu_i) = \boldsymbol{\beta}^\top x_i \quad \text{and} \quad \text{probit}(\pi_i) = F(\boldsymbol{\tau}^\top s_i)$$

where F is the C.D.F of $\mathcal{N}(0, 1)$. The log-likelihood of $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\tau}^\top)^\top$ is

$$\begin{aligned} \ell(\theta) = & \sum_{i=1}^n \left\{ J_i \log \left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right] \right. \\ & \left. + (1 - J_i) \left[Y_i \log[W(e^{\boldsymbol{\beta}^\top x_i})] + \log[1 - F(\boldsymbol{\tau}^\top s_i)] - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right] \right\}. \end{aligned}$$

The score function of the ZIPB regression model

Note that $I(Y_i > 0) = 1 - I(Y_i = 0)$ and $Y_i I(Y_i = 0) = 0$, $i = 1, \dots, n$.

We have

$$\begin{aligned}
S_{i1}(\boldsymbol{\theta}) &= \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\tau}} = \frac{\partial}{\partial \boldsymbol{\tau}} \left\{ I(Y_i = 0) \log \left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right] \right\} \\
&\quad + \frac{\partial}{\partial \boldsymbol{\tau}} \left\{ I(Y_i > 0) \left[Y_i \log[W(e^{\boldsymbol{\beta}^\top x_i})] + \log[1 - F(\boldsymbol{\tau}^\top s_i)] - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right] \right\} \\
&= I(Y_i = 0) \left[\frac{f(\boldsymbol{\tau}^T \mathcal{X}_i) - f(\boldsymbol{\tau}^T \mathcal{X}_i) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right)}{F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right)} \right] \\
&\quad + (1 - I(Y_i = 0)) \left[\frac{-f(\boldsymbol{\tau}^T \mathcal{X}_i)}{1 - F(\boldsymbol{\tau}^\top s_i)} \right] \\
&= I(Y_i = 0) \left[\frac{\left(1 - \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right) f(\boldsymbol{\tau}^T \mathcal{X}_i)}{F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right)} \right] \\
&\quad - \frac{f(\boldsymbol{\tau}^T \mathcal{X}_i)}{1 - F(\boldsymbol{\tau}^\top s_i)} + I(Y_i = 0) \left[\frac{f(\boldsymbol{\tau}^T \mathcal{X}_i)}{1 - F(\boldsymbol{\tau}^\top s_i)} \right] \\
&= I(Y_i = 0) \left[\frac{\left(1 - \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right) f(\boldsymbol{\tau}^T \mathcal{X}_i)}{F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right)} + \frac{f(\boldsymbol{\tau}^T \mathcal{X}_i)}{1 - F(\boldsymbol{\tau}^\top s_i)} \right] \\
&\quad - \frac{f(\boldsymbol{\tau}^T \mathcal{X}_i)}{1 - F(\boldsymbol{\tau}^\top s_i)} \\
&= I(Y_i = 0) \left\{ \frac{\left(1 - \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right) f(\boldsymbol{\tau}^T \mathcal{X}_i) (1 - F(\boldsymbol{\tau}^\top s_i))}{\left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right] (1 - F(\boldsymbol{\tau}^\top s_i))} \right\} \\
&\quad + I(Y_i = 0) \left\{ \frac{f(\boldsymbol{\tau}^T \mathcal{X}_i) \left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right]}{\left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right] (1 - F(\boldsymbol{\tau}^\top s_i))} \right\} \\
&\quad - \frac{f(\boldsymbol{\tau}^T \mathcal{X}_i)}{1 - F(\boldsymbol{\tau}^\top s_i)} \\
&= I(Y_i = 0) \left\{ \frac{f(\boldsymbol{\tau}^T \mathcal{X}_i)}{\left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right] (1 - F(\boldsymbol{\tau}^\top s_i))} \right\} - \frac{f(\boldsymbol{\tau}^T \mathcal{X}_i)}{1 - F(\boldsymbol{\tau}^\top s_i)} \\
&= I(Y_i = 0) \cdot \frac{1}{\left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right]} \cdot \frac{f(\boldsymbol{\tau}^T \mathcal{X}_i)}{1 - F(\boldsymbol{\tau}^\top s_i)} - \frac{f(\boldsymbol{\tau}^T \mathcal{X}_i)}{1 - F(\boldsymbol{\tau}^\top s_i)} \\
&= \frac{f(\boldsymbol{\tau}^T \mathcal{X}_i)}{1 - F(\boldsymbol{\tau}^\top s_i)} \cdot \frac{I(Y_i = 0)}{\left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right]} - \frac{f(\boldsymbol{\tau}^T \mathcal{X}_i)}{1 - F(\boldsymbol{\tau}^\top s_i)} \\
&= \frac{f(\boldsymbol{\tau}^T \mathcal{X}_i)}{1 - F(\boldsymbol{\tau}^\top s_i)} \cdot \frac{I(Y_i = 0) - \left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right]}{\left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp \left(1 - e^{W(e^{\boldsymbol{\beta}^\top x_i})} \right) \right]}
\end{aligned}$$

Note that

$$\frac{e^{W(e^{\beta^\top x_i})} \cdot (1 - F(\boldsymbol{\tau}^\top s_i)) \cdot \exp\left(1 - e^{W(e^{\beta^\top x_i})}\right)}{F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp\left(1 - e^{W(e^{\beta^\top x_i})}\right)} - e^{W(e^{\beta^\top x_i})} = \frac{-e^{W(e^{\beta^\top x_i})} \cdot F(\boldsymbol{\tau}^\top s_i)}{F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp\left(1 - e^{W(e^{\beta^\top x_i})}\right)}$$

Thus

$$\begin{aligned} S_{i2}(\boldsymbol{\theta}) &= \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \\ &= -I(Y_i = 0) \frac{\partial}{\partial \boldsymbol{\beta}} [W(e^{\beta^\top x_i})] \cdot \left[\frac{-e^{W(e^{\beta^\top x_i})} \cdot F(\boldsymbol{\tau}^\top s_i)}{F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp\left(1 - e^{W(e^{\beta^\top x_i})}\right)} \right] \\ &\quad + \frac{\partial}{\partial \boldsymbol{\beta}} [W(e^{\beta^\top x_i})] \cdot \left[\frac{Y_i}{W(e^{\beta^\top x_i})} - e^{W(e^{\beta^\top x_i})} \right] \\ &= I(Y_i = 0) \frac{\partial}{\partial \boldsymbol{\beta}} [W(e^{\beta^\top x_i})] \cdot \left[\frac{e^{W(e^{\beta^\top x_i})} \cdot F(\boldsymbol{\tau}^\top s_i)}{F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp\left(1 - e^{W(e^{\beta^\top x_i})}\right)} \right] \\ &\quad + \frac{\partial}{\partial \boldsymbol{\beta}} [W(e^{\beta^\top x_i})] \cdot \left[Y_i - W(e^{\beta^\top x_i}) \cdot e^{W(e^{\beta^\top x_i})} \right] \\ &= I(Y_i = 0) \frac{\partial}{\partial \boldsymbol{\beta}} [W(e^{\beta^\top x_i})] \cdot \left[\frac{e^{W(e^{\beta^\top x_i})} \cdot F(\boldsymbol{\tau}^\top s_i)}{F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp\left(1 - e^{W(e^{\beta^\top x_i})}\right)} \right] \\ &\quad + \frac{\partial}{\partial \boldsymbol{\beta}} [W(e^{\beta^\top x_i})] \cdot \left\{ Y_i - \left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \right] W(e^{\beta^\top x_i}) \cdot e^{W(e^{\beta^\top x_i})} \right\} \\ &= I(Y_i = 0) \frac{\partial}{\partial \boldsymbol{\beta}} [W(e^{\beta^\top x_i})] \cdot \left[\frac{e^{W(e^{\beta^\top x_i})} \cdot F(\boldsymbol{\tau}^\top s_i)}{F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp\left(1 - e^{W(e^{\beta^\top x_i})}\right)} \right] \\ &\quad + \frac{\partial}{\partial \boldsymbol{\beta}} [W(e^{\beta^\top x_i})] \cdot \left\{ Y_i - (1 - F(\boldsymbol{\tau}^\top s_i)) W(e^{\beta^\top x_i}) \cdot e^{W(e^{\beta^\top x_i})} \right\} \\ &\quad - \frac{\partial}{\partial \boldsymbol{\beta}} [W(e^{\beta^\top x_i})] \cdot \left\{ F(\boldsymbol{\tau}^\top s_i) \cdot W(e^{\beta^\top x_i}) \cdot e^{W(e^{\beta^\top x_i})} \right\} \\ &= \frac{\partial}{\partial \boldsymbol{\beta}} [W(e^{\beta^\top x_i})] \cdot e^{W(e^{\beta^\top x_i})} \cdot F(\boldsymbol{\tau}^\top s_i) \cdot \left[\frac{I(Y_i = 0)}{F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp\left(1 - e^{W(e^{\beta^\top x_i})}\right)} - 1 \right] \\ &\quad + \frac{\partial}{\partial \boldsymbol{\beta}} [W(e^{\beta^\top x_i})] \cdot \left\{ Y_i - (1 - F(\boldsymbol{\tau}^\top s_i)) W(e^{\beta^\top x_i}) \cdot e^{W(e^{\beta^\top x_i})} \right\} \\ &= \frac{\partial}{\partial \boldsymbol{\beta}} [W(e^{\beta^\top x_i})] \cdot e^{W(e^{\beta^\top x_i})} \cdot F(\boldsymbol{\tau}^\top s_i) \cdot \frac{I(Y_i = 0) - \left[F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp\left(1 - e^{W(e^{\beta^\top x_i})}\right) \right]}{F(\boldsymbol{\tau}^\top s_i) + (1 - F(\boldsymbol{\tau}^\top s_i)) \exp\left(1 - e^{W(e^{\beta^\top x_i})}\right)} \\ &\quad + \frac{\partial}{\partial \boldsymbol{\beta}} [W(e^{\beta^\top x_i})] \cdot \left\{ Y_i - (1 - F(\boldsymbol{\tau}^\top s_i)) W(e^{\beta^\top x_i}) \cdot e^{W(e^{\beta^\top x_i})} \right\} \end{aligned}$$

Let $S_i(\boldsymbol{\theta}) = \partial \ell_i(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} = (S_{i1}^T(\boldsymbol{\theta}), S_{i2}^T(\boldsymbol{\theta}))^T$, ($i = 1, \dots, n$) be the score function of the

ZIPBell model. Then $S_{i1}(\boldsymbol{\theta}) = \partial \ell_i(\boldsymbol{\theta}) / \partial \boldsymbol{\tau}$ and $S_{i2}(\boldsymbol{\theta}) = \partial \ell_i(\boldsymbol{\theta}) / \partial \boldsymbol{\beta}$ are respectively given by

$$S_{i1}(\boldsymbol{\theta}) = \frac{f(\boldsymbol{\tau}^T \boldsymbol{x}_i)}{1 - F(\boldsymbol{\tau}^T \boldsymbol{s}_i)} \cdot \frac{I(Y_i = 0) - \left[F(\boldsymbol{\tau}^T \boldsymbol{s}_i) + (1 - F(\boldsymbol{\tau}^T \boldsymbol{s}_i)) \exp\left(1 - e^{W(e^{\boldsymbol{\beta}^T \boldsymbol{x}_i)}\right)} \right]}{\left[F(\boldsymbol{\tau}^T \boldsymbol{s}_i) + (1 - F(\boldsymbol{\tau}^T \boldsymbol{s}_i)) \exp\left(1 - e^{W(e^{\boldsymbol{\beta}^T \boldsymbol{x}_i)}\right)} \right]},$$

$$S_{i2}(\boldsymbol{\theta}) = D \cdot e^{W(e^{\boldsymbol{\beta}^T \boldsymbol{x}_i})} \cdot F(\boldsymbol{\tau}^T \boldsymbol{s}_i) \cdot \frac{I(Y_i = 0) - \left[F(\boldsymbol{\tau}^T \boldsymbol{s}_i) + (1 - F(\boldsymbol{\tau}^T \boldsymbol{s}_i)) \exp\left(1 - e^{W(e^{\boldsymbol{\beta}^T \boldsymbol{x}_i)}\right)} \right]}{F(\boldsymbol{\tau}^T \boldsymbol{s}_i) + (1 - F(\boldsymbol{\tau}^T \boldsymbol{s}_i)) \exp\left(1 - e^{W(e^{\boldsymbol{\beta}^T \boldsymbol{x}_i)}\right)}$$

$$+ \frac{D}{W(e^{\boldsymbol{\beta}^T \boldsymbol{x}_i})} \cdot \left\{ Y_i - \left(1 - F(\boldsymbol{\tau}^T \boldsymbol{s}_i)\right) W(e^{\boldsymbol{\beta}^T \boldsymbol{x}_i}) \cdot e^{W(e^{\boldsymbol{\beta}^T \boldsymbol{x}_i})} \right\}$$

with $D = \frac{\partial}{\partial \boldsymbol{\beta}} [W(e^{\boldsymbol{\beta}^T \boldsymbol{x}_i})]$.

Proof of Theorem 1

To demonstrate that $\widehat{\boldsymbol{\theta}}_F$ is a consistent estimator of $\boldsymbol{\theta}$, note that

$$U_{F,n}(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n S_i(\boldsymbol{\theta}),$$

$$G_{F,n}(\boldsymbol{\theta}) = -\frac{1}{\sqrt{n}} \frac{\partial U_{F,n}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} = -\frac{1}{n} \sum_{i=1}^n \frac{\partial S_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}.$$

Basing on the WLLN, we have: $G_{F,n}(\boldsymbol{\theta}) \xrightarrow{p} G_F(\boldsymbol{\theta})$ and (K2), the convergence of $G_{F,n}(\boldsymbol{\theta})$ to $G_F(\boldsymbol{\theta})$ is uniform in a neighborhood of the true $\boldsymbol{\theta}$.

Depending on the inverse function theorem of Foutz (1977), it can be inferred that a unique consistent solution of $U_{F,n}(\boldsymbol{\theta}) = 0$ exists in a neighborhood of the true.

Consequently, $\widehat{\boldsymbol{\theta}}_F$ is a consistent estimator of $\boldsymbol{\theta}$.

The asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\theta}}_F - \boldsymbol{\theta})$ is inferred right here.

As we know $\widehat{\boldsymbol{\theta}}_F$ is the unique root of $U_{F,n}(\boldsymbol{\theta}) = 0$, and perform the Taylor's expansion of $U_{F,n}(\widehat{\boldsymbol{\theta}}_F)$ at $\boldsymbol{\theta}$ to have the main concern:

$$0 = U_{F,n}(\widehat{\boldsymbol{\theta}}_F) = U_{F,n}(\boldsymbol{\theta}) + \left(\frac{1}{\sqrt{n}} \frac{\partial U_{F,n}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right) \sqrt{n}(\widehat{\boldsymbol{\theta}}_F - \boldsymbol{\theta}) + o_p(1).$$

This can lead to an inference:

$$\begin{aligned} \sqrt{n}(\widehat{\boldsymbol{\theta}}_F - \boldsymbol{\theta}) &= G_{F,n}^{-1}(\boldsymbol{\theta}) U_{F,n}(\boldsymbol{\theta}) + o_p(1) \\ &= G_F^{-1}(\boldsymbol{\theta}) U_{F,n}(\boldsymbol{\theta}) + [G_{F,n}^{-1}(\boldsymbol{\theta}) - G_F^{-1}(\boldsymbol{\theta})] U_{F,n}(\boldsymbol{\theta}) + o_p(1). \end{aligned}$$

For $i = 1, 2, \dots, n$, $Var[S_i(\boldsymbol{\theta})] = E[[S_i(\boldsymbol{\theta})]^{\otimes 2}]$ and $E[S_i(\boldsymbol{\theta})] = 0$.

Applying the central limit theorem and (K2), we can demonstrate that:

$$U_{F,n}(\boldsymbol{\theta}) \xrightarrow{d} N(0, Q_F(\boldsymbol{\theta})),$$

with $Q_F(\boldsymbol{\theta}) = \text{Var}[U_{F,n}(\boldsymbol{\theta})]$. Besides, because $G_{F,n}(\boldsymbol{\theta}) - G_F(\boldsymbol{\theta}) \xrightarrow{p} 0$ by the WLLN and $G_F(\boldsymbol{\theta})$ is nonsingular by (K2), we execute the Slutsky's theorem to show that

$$[G_{F,n}^{-1}(\boldsymbol{\theta}) - G_F^{-1}(\boldsymbol{\theta})] U_{F,n}(\boldsymbol{\theta}) \xrightarrow{d} 0,$$

In consequence $[G_{F,n}^{-1}(\boldsymbol{\theta}) - G_F^{-1}(\boldsymbol{\theta})] U_{F,n}(\boldsymbol{\theta}) \xrightarrow{p} 0$.

Using the Slutsky's theorem we obtain $\sqrt{n}(\widehat{\boldsymbol{\theta}}_F - \boldsymbol{\theta}) \xrightarrow{d} N(0, \Delta_F)$,

with $\Delta_F = G_F^{-1}(\boldsymbol{\theta})Q_F(\boldsymbol{\theta})[G_F^{-1}(\boldsymbol{\theta})]^T = G_F^{-1}(\boldsymbol{\theta})$ since $G_F(\boldsymbol{\theta}) = Q_F(\boldsymbol{\theta})$.